A Nonparametric Test for Comparing Valuation Distributions in First-Price Auctions *

Nianqing Liu[†] [‡]

Yao Luo[§]

This version: May, 2016
First version: November, 2013

Abstract

This paper proposes a nonparametric test for comparing valuation distributions in first-price auctions. Our test is motivated by the fact that two valuation distributions are the same if and only if their integrated quantile functions are the same. Our method avoids estimating unobserved valuations and does not require smooth estimation of bid density. We show that our test is consistent against all fixed alternatives and has non-trivial power against root-N local alternatives. Monte Carlo experiments show that our test performs well in finite samples. We implement our method on data from U.S. Forest Service timber auctions. We also discuss how our test can be adapted to other testing problems in auctions.

Keywords: First-Price Auctions, Exogenous Participation, Integrated Quantile Function, Nonparametric, Hypothesis Test

JEL: D44, D82, C12, C14

*We are deeply indebted to Quang Vuong and Isabelle Perrigne for their advice and encouragement when we were studying at Penn State. We would like to thank the editor Hanming Fang and three anonymous referees for comments that have greatly improved the article. We also thank Chunrong Ai, Yanqin Fan, Jiti Gao, LungFei Lee, Arthur Lewbel, Yong Li, Jun Ma, Vadim Marmer, Joris Pinkse, Yuanyuan Wan, Adam Chi Leung Wong, Ping Yu, and seminar/conference participants at the 2014 North American Summer Meeting of the Econometric Society, the 9th GNYEC, the 31st CESG, the 2015 World Congress of Econometric Society, the 2015 Shanghai Econometrics Workshop, Fudan University, Shandong University and Renmin University of China for helpful comments. Zhe Yuan provided excellent research assistance. The authors respectively acknowledge financial support from the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education of China (201309KF02) and the National Natural Science Foundation of China (NSFC-71373283 and NSFC-71463019). The usual disclaimer applies.

[†]Shanghai University of Finance and Economics. e-mail: nliu@mail.shufe.edu.cn

[‡]Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, China

[§]University of Toronto. e-mail: yao.luo@utoronto.ca

1 Introduction

We propose a nonparametric test for comparing valuation distributions in first-price auctions. One important application is to justify the exogenous participation assumption which has been adopted to identify various auction models, such as first-price auctions with risk aversion (Guerre, Perrigne, and Vuong, 2009), ascending auctions (Aradillas-López, Gandhi, and Quint, 2013), and first-price auctions under ambiguity (Aryal, Grundl, Kim, and Zhu, 2015).[1] Many testing problems in auctions also reduce to the standard form of comparing valuation distributions, such as detecting collusion (Aryal and Gabrielli, 2012), distinguishing private value and common value auctions (Haile, Hong, and Shum, 2003), and testing different models of entry (Marmer, Shneyerov, and Xu, 2013).

Testing for the equality of distributions is a standard problem in statistics. Classic examples of such tests are the Kolmogorov-Smirnov test, the Cramer-von Mises test, and the Anderson-Darling test.[2] In first-price auctions, complications arise from the fact that valuations are estimated rather than observed directly. In a seminal paper, Guerre, Perrigne, and Vuong (2000) transformed the First-Order Conditions (FOC) for optimal bids into an expression of bidder's value as an explicit function of the submitted bid, the bid Probability Density Function (PDF) and the bid Cumulative Distribution Function (CDF). The authors then used this function to estimate each bidder's valuation by recovering the PDF and CDF of the bids. In principle, by treating these estimated values as a pseudo sample, one can adapt the existing tests of distributional equality mentioned above (e.g., Kolmogorov-Smirnov test) to obtain a two-step testing procedure in first-price auctions. In the first step, pseudo values are estimated; in the second step, existing tests of distributional equality are applied to the sample of pseudo values.

However, such a two-step testing procedure has at least two complications. First, it introduces (finite sample) dependence among pseudo values by estimating the bid CDF and PDF with the same sample of bids in the first step. Such a dependence brings technical difficulty in establishing the asymptotic validity of existing tests of distributional equality. Second, besides adding estimation error, the first-step recovery of pseudo values usually involves the complications of choosing a bandwidth and trimming in smooth estimation of bid density.[3] See also discussion in Haile, Hong, and Shum (2003).

The nonparametric test for distributional equality we propose here avoids the construction of pseudo valuations in first-price auctions. Our approach is motivated by a simple yet profoundly useful idea: two valuation distributions are the same if and only if their integrated

[1] Exogenous participation means that the valuation distribution does not depend on the number of bidders (see Athey and Haile, 2002), namely $F(v|I_1) = \cdots = F(v|I_K)$ for all $v \in [\underline{v}, \overline{v}]$.

[2] One recent related example is Barrett, Donald, and Bhattacharya (2014) who compared the mean-standardized Integrated-Quantile Functions (IQFs) (i.e., IQFs divided by their corresponding means) of two directly observed samples. Notice that equality of distributions implies equality of mean-standardized IQFs, but the reverse is not true.

[3] Trimming might be avoided in the smooth estimation of bid density if boundary (bias) correction is implemented (see, e.g., Hickman and Hubbard, 2014; Li and Liu, 2015).

quantile functions are the same. We show that the bidders' FOC allows us to express the integrated quantile function of a valuation distribution as a simple linear functional of the quantile function of the bids. In light of this observation, we propose a test statistic measuring the square of the L^2-distance between the sample analogues of this linear functional for two bid samples. Consequently, our test statistic only involves the two empirical quantile functions of the bids.

Our test has two attractive features. First, the test statistic is calculated in one step, which allows us to characterize its asymptotic properties conveniently under regularity conditions. In particular, we show that the test statistic converges to the square of the L^2-norm of a Gaussian process with mean zero at a parametric rate under the null hypothesis. We also show that our test is consistent against any fixed alternative and can detect local alternatives converging to the null hypothesis at a rate of root-N. Second, our test statistic is easy to calculate, as it involves no density estimation. Moreover, since the empirical quantile function of bids is a step function, the empirical counterpart of the integrated valuation quantile function is piecewise linear. Therefore, the test statistic, i.e., the square of the L^2-distance between two empirical counterparts of integrated valuation quantile functions, is simply the total area below a piecewise quadratic curve which has an explicit expression in the ordered bids. This feature makes our test easy to implement in practice.

In the auction literature, there has been increasing interest in the development of statistical tests. Examples include tests for affiliation, such as the ones proposed by Li and Zhang (2010) and Jun, Pinkse, and Wan (2010); tests for monotonicity of inverse bidding strategy such as that proposed by Liu and Vuong (2013); tests for discriminating entry models such as the one developed by Marmer, Shneyerov, and Xu (2013); and tests for risk aversion such as that developed by Fang and Tang (2014).

Although our testing approach is new, we are not the first to use quantile-based approaches in auctions. Marmer and Shneyerov (2012) and Marmer, Shneyerov, and Xu (2013) proposed a quantile-based estimator in first-price auctions and a quantile-based test for distinguishing different entry models, respectively. Liu and Vuong (2013) first introduced the idea of using the integrated valuation quantile function and greatest convex minorant (or least concave majorant) in auctions. They developed an integrated-quantile-based test of monotonicity of inverse bidding strategy in first-price auctions. Based on the equivalence between monotonicity of inverse bidding strategy and convexity of the integrated value quantile function of a bidder's strongest competitor, they proposed a test statistic comparing the difference between an estimator of such an integrated value quantile function and its greatest convex minorant. Their test is shown to have non-trivial power against root-N local alternatives. For nonlinear pricing models, Luo, Perrigne, and Vuong (2014) proposed a quantile-based estimator which achieves root-N consistency.

The reminder of the paper is organized as follows. In the next section, we describe our testing problem and introduce the test statistic. We then derive the asymptotic properties (i.e., asymptotic distribution under the null hypothesis, size, and power) of the test in Section

3. In Section 4, we report the results of a Monte Carlo study for moderate sample sizes. In section 5, we discuss applications of our test to auctions with an arbitrary number of samples, auctions with endogenous entry, auctions where only winning bids are recorded; and we accommodate the test to cases where assumptions regarding reserve price, risk aversion, unobserved heterogeneity, and asymmetric bidders are relaxed. Section 6 applies our method to data from U.S. Forest Service timber auctions. Section 7 concludes the whole paper. Supplementary results are presented in Appendix A, while proofs are gathered in Appendix B.

2 Comparing Value Distributions in First-Price Auctions

We briefly present the first-price sealed-bid auction model with independent private values. A single and indivisible object is auctioned. I potential bidders are symmetric and risk neutral. Their private values are i.i.d. draws from a common distribution $F(\cdot)$, which is absolutely continuous with density $f(\cdot)$ and support $[\underline{v}, \overline{v}]$. With a nonbinding reserve price, the equilibrium bid function takes the form of:

$$b = s(v|F, I) \equiv v - \frac{1}{F(v)^{I-1}} \int_0^v F(x)^{I-1} dx.$$

2.1 Motivation

The seminal paper by Guerre, Perrigne, and Vuong (2000) shows that a bidder's value (v) can be expressed as an explicit function of the submitted bid (b), the bid PDF ($g(\cdot)$), and the bid CDF ($G(\cdot)$)

$$v = \xi(b) \equiv b + \frac{1}{I-1} \frac{G(b)}{g(b)}. \tag{1}$$

This equation can be rewritten in terms of the quantile functions of values and bids as

$$v(\alpha) = b(\alpha) + \frac{1}{I-1} \frac{\alpha}{g(b(\alpha))}, \tag{2}$$

where $v(\alpha)$ and $b(\alpha)$ are the α quantiles of the valuations and bids, respectively. Marmer and Shneyerov (2012) used the well-known formula $f(v) = 1/v'(F(v))$ to estimate $f(\cdot)$. Haile, Hong, and Shum (2003) and Marmer, Shneyerov, and Xu (2013) exploited this quantile relationship to construct tests for detecting common value and to distinguish entry models, respectively. Note that the bid density $g(\cdot)$ is involved in both Equations (1) and (2). Thus, estimating the valuation distribution (quantile) function requires estimating both the bid distribution (quantile) function and the density function. Recovering the latter could be troublesome, as we need to choose a tuning parameter, namely, some bandwidth h, in density estimation.

To motivate our test, we focus on the integral of the quantile function of valuations. In particular, we rewrite the quantile relationship as follows:

$$\begin{aligned} v(\alpha) &= \frac{I-2}{I-1}b(\alpha) + \frac{1}{I-1}[b(\alpha) + \frac{\alpha}{g(b(\alpha))}] \\ &= \frac{I-2}{I-1}b(\alpha) + \frac{1}{I-1}\frac{d(b(\alpha)\cdot\alpha)}{d\alpha}. \end{aligned}$$

Integrating both sides from 0 to β leads to our basic formula:

$$V(\beta) \equiv \int_0^\beta v(\alpha)d\alpha = \frac{I-2}{I-1}\int_0^\beta b(\alpha)d\alpha + \frac{1}{I-1}b(\beta)\beta. \tag{3}$$

Note that the right-hand side involves only the bid quantile function $b(\cdot)$, which can be nonparametrically estimated at a root-N rate.

2.2 Test Statistic

We now introduce our testing problem formally. To clarify ideas, we consider first the case of two samples. We observe two independent i.i.d. samples $\{B_{1,1},\ldots,B_{1,N_1}\}$ and $\{B_{2,1},\ldots,B_{2,N_2}\}$, where $N_k = I_k \times L_k$ is the number of bids generated from L_k homogeneous first-price auctions with I_k bidders and valuation distribution $F(\cdot|I_k)$ for $k = 1,2$. We want to test whether the two valuation distributions $F(\cdot|I_1)$ and $F(\cdot|I_2)$ are the same. Specifically, our hypothesis of interest is:

$$H_0 : F(v|I_1) = F(v|I_2), \forall v \in [\underline{v}, \overline{v}] \quad \text{v.s.} \quad H_1 : F(v|I_1) \neq F(v|I_2), \text{for some } v \in [\underline{v}, \overline{v}].$$

Denote $v(\cdot|I_k)$ as the corresponding quantile function of $F(\cdot|I_k)$, where $k = 1,2$. Define $V(\cdot|I_k) \equiv \int_0^{\cdot} v(\alpha|I_k)d\alpha$. Our testing problem is equivalent to

$$H_0 : V(\beta|I_1) = V(\beta|I_2), \forall \beta \in [0,1] \quad \text{v.s.} \quad H_1 : V(\beta|I_1) \neq V(\beta|I_2), \text{for some } \beta \in [0,1]$$

where functions $V(\cdot|I_1)$ and $V(\cdot|I_2)$ can be expressed in terms of bid quantile functions

$$V(\beta|I_k) \equiv \int_0^\beta v(\alpha|I_k)d\alpha = \frac{I_k-2}{I_k-1}\int_0^\beta b(\alpha|I_k)d\alpha + \frac{1}{I_k-1}b(\beta|I_k)\beta, \tag{4}$$

where $b(\cdot|I_k)$ is the quantile function of bids generated from auctions with I_k bidders and valuation distribution $F(\cdot|I_k)$.

Under the null hypothesis, $V(\cdot|I_1)$ and $V(\cdot|I_2)$ are identical on the whole domain $[0,1]$. In other words, the distance between $V(\cdot|I_1)$ and $V(\cdot|I_2)$ is zero under the null and positive under the alternative. Naturally, we propose a test statistic which measures the difference between their sample analogue[4]

$$t \equiv \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \int_0^1 \left(\widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2)\right)^2 d\beta \tag{5}$$

where $\widehat{V}(\cdot|I_1)$ and $\widehat{V}(\cdot|I_2)$ are the sample analogues of $V(\cdot|I_1)$ and $V(\cdot|I_2)$, respectively. When sample size is large, $\widehat{V}(\cdot|I_1)$ and $\widehat{V}(\cdot|I_2)$ will be close to their true functions. Thus, the test statistic has a small value under the null hypothesis, and diverges to infinity under the alternative. Consequently, our test rejects the null hypothesis when the test statistic is large enough.

$\widehat{V}(\cdot|I_k)$ is piecewise linear with possible jumps at $\{1/N_k, 2/N_k, \dots, N_k/N_k\}$. To see this, we order the bids in each sample. Denote $B_{1,(1)} \leq \dots \leq B_{1,(N_1)}$ and $B_{2,(1)} \leq \dots \leq B_{2,(N_2)}$ as the order statistics of sample 1 and 2, respectively. Simple algebra yields

$$\widehat{V}(\beta|I_k) = B_{k,(i_k(\beta))} \times \beta + \frac{I_k - 2}{N_k(I_k - 1)} \Big[\sum_{j=1}^{i_k(\beta)} B_{k,(j)} - i_k(\beta) \times B_{k,(i_k(\beta))}\Big],$$

where $i_k(\beta)$ is chosen such that $\frac{i_k(\beta)-1}{N_k} < \beta \leq \frac{i_k(\beta)}{N_k}$.[5]

Thanks to this property, our test statistic is easy to implement in practice. The knots $\{1/N_1, 2/N_1, \dots, N_1/N_1\}$ and $\{1/N_2, 2/N_2, \dots, N_2/N_2\}$ divide the interval $(0, 1]$ into subintervals. Denote these subintervals as $\{(\tau_\ell, \tau_{\ell+1}]\}_{\ell=0}^{N_0-1}$, where N_0 is the number of subintervals, $\tau_\ell < \tau_{\ell+1}$, $\tau_0 = 0$ and $\tau_{N_0} = 1$. Since $\widehat{V}(\cdot|I_1)$ and $\widehat{V}(\cdot|I_2)$ are both linear in $(\tau_\ell, \tau_{\ell+1}]$, the integral $\int_{\tau_\ell}^{\tau_{\ell+1}} \left(\widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2)\right)^2 d\beta$ is essentially the area below a quadratic function on $(\tau_\ell, \tau_{\ell+1}]$. Denote $d_\ell^- \equiv \widehat{V}(\tau_{\ell-}|I_1) - \widehat{V}(\tau_{\ell-}|I_2)$ and $d_\ell^+ \equiv \widehat{V}(\tau_{\ell+}|I_1) - \widehat{V}(\tau_{\ell+}|I_2)$, where $\widehat{V}(\beta_-|I_k) \equiv \lim_{u\uparrow\beta} \widehat{V}(u|I_k)$ and $\widehat{V}(\beta_+|I_k) \equiv \lim_{u\downarrow\beta} \widehat{V}(u|I_k)$. Our test statistic has an explicit formula:

$$t = \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \sum_{\ell=0}^{N_0-1} \left(\frac{\tau_{\ell+1} - \tau_\ell}{3}\right) \cdot \left(d_\ell^{+2} + {d_{\ell+1}^-}^2 + d_\ell^+ \cdot d_{\ell+1}^-\right). \tag{6}$$

[4]Notice that the square root of test statistic t becomes an L^2-metric based statistic $t' = \sqrt{\frac{N_1 \cdot N_2}{N_1+N_2}} \cdot \left[\int_0^1 \left(\widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2)\right)^2 d\beta\right]^{1/2}$.

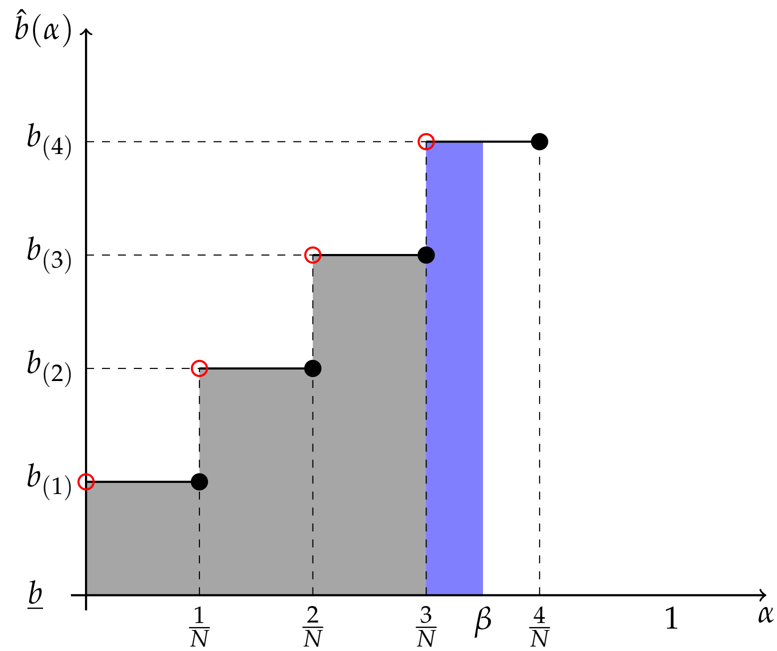
[5]For a given sample $X_1, \dots, X_n$, we define the empirical quantile function F_n^{-1} as the inverse mapping of the empirical distribution function F_n, i.e.

$$F_n^{-1}(\alpha) = inf\{x : F_n(x) \geq \alpha\} = X_{(i)}$$

where i is chosen such that $\frac{i-1}{n} < \alpha \leq \frac{i}{n}$, and $X_{(1)}, \dots, X_{(n)}$ are the order statistics of the sample; that is, $X_{(1)} \leq \dots \leq X_{(n)}$ and $(X_{(1)}, \dots, X_{(n)})$ is a permutation of the sample $X_1, \dots, X_n$. As Figure 1 shows,

$$\int_0^\beta F_n^{-1}(\alpha) d\alpha = X_{(i)}(\beta - \frac{i-1}{n}) + \frac{1}{n}\sum_{k=1}^{i-1} X_{(k)}.$$

Figure 1: Integration of the Empirical Bid Quantile Function



Alternatively, we can use a test statistic based on the L^1 metric as follows

$$t'' = \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \cdot \int_0^1 \left| \widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2) \right| d\beta.$$

We can establish its asymptotic null distribution in a similar way to the case of the statistic t. To implement in practice, this test statistic also has an explicit formula similar to Equation (6) as follows:[6]

$$t'' = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \cdot \sum_{\ell=0}^{N_0 - 1} \left(\frac{\tau_{\ell+1} - \tau_\ell}{2} \right) \cdot \left(|d_\ell^+| + |d_{\ell+1}^-| - \mathbf{1}\left(d_\ell^+ \cdot d_{\ell+1}^- < 0 \right) \cdot \frac{2|d_\ell^+| \cdot |d_{\ell+1}^-|}{|d_\ell^+| + |d_{\ell+1}^-|} \right).$$

3 Asymptotic Properties

Following the literature, we make the following regularity assumptions on how the bids are generated:

Assumption 1. *For any $I \in \{I_1, I_2\}$, the random variables $B_1, \ldots, B_I$ are independent and identically distributed (i.i.d.) with common true distribution $G(\cdot|I)$.*

Assumption 1 is a regularity assumption concerning the bid data-generating process. In addition, we impose some specific properties on the bid distribution in Assumption 2.

[6] Notice that, in this case, the integral $\int_{\tau_\ell}^{\tau_{\ell+1}} \left| \widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2) \right| d\beta$ is essentially the area of a trapezoid if the curves of $\widehat{V}(\cdot|I_1)$ and $\widehat{V}(\cdot|I_2)$ do not cross in $(\tau_\ell, \tau_{\ell+1}]$, or two similar triangles otherwise.

Assumption 2. *For any $I \in \{I_1, I_2\}$, the bid distribution $G(\cdot \mid I)$ is continuously differentiable with a density $g(\cdot \mid I)$ on a support of $\left[\underline{b}, \overline{b}\right]$. In addition, the bid density $g(b|I) \geq c_g > 0$ for any $b \in [\underline{b}, \overline{b}]$.*

Assumption 2 requires that the bid distribution has a continuous density function bounded away from zero on a compact support. It guarantees that the inverse bid function $\xi(b) = b + \frac{1}{I-1} \cdot \frac{G(b|I)}{g(b|I)}$ is well defined on the whole support of the bid distribution.

Assumption 3. $\frac{N_1}{N_1+N_2} \to \lambda \in [0,1]$ *as* $\min\{N_1, N_2\} \to \infty$.

Assumption 3 means that the size of the first bids sample is proportional to the second one in the limit when the parameter λ is in the interior of $[0,1]$. The second sample tends to have many more (fewer) observations than the first one when the parameter λ is 0 (is 1).

3.1 Asymptotic Distribution under the Null Hypothesis

In this subsection, we provide the asymptotic null distribution of our test statistic. First, we define two mappings T_k, $k = 1, 2$, as

$$T_k(f)(\beta) = \frac{I_k - 2}{I_k - 1} \int_0^\beta f(\alpha) d\alpha + \frac{1}{I_k - 1} \cdot f(\beta) \cdot \beta, \qquad \beta \in [0,1], \tag{7}$$

where $f(\cdot)$ is any integrable function defined on $[0,1]$. The following lemma shows that the mappings T_k, $k = 1, 2$, are linear.

Lemma 1. *The mappings T_k, $k = 1, 2$, defined in Equation (7) are linear, i.e., for any $c \in \mathbb{R}$ and any integrable functions $f(\cdot)$ and $h(\cdot)$ defined on $[0,1]$,*

$$T_k\left(c \cdot f + h\right)(\beta) = c \cdot T_k(f)(\beta) + T_k(h)(\beta), \quad \forall \beta \in [0,1].$$

Proof. See Appendix B.1. □

Since a linear mapping of a Gaussian process is still a Gaussian process,[7] we now show that the asymptotic null distribution of our test statistic is the square of the L^2-norm of a Gaussian process. This Gaussian process is the sum of two independent Gaussian processes generated by applying T_1 and T_2 on two quantile processes. Let $\mathbb{B}(\cdot)$ denote the standard Brownian bridge, $\mathbb{G}_k(\cdot)$ denote $\mathbb{B}(\cdot)/g(b(\cdot \mid I_k) \mid I_k)$ for $k = 1, 2$, $\hat{b}(\cdot|I_k)$ be the empirical bid quantile function of sample k for $k = 1, 2$, $\xrightarrow{d}$ denote convergence in distribution, and $\rightsquigarrow$ be weak convergence.

Theorem 1. *Suppose Assumptions 1, 2 and 3 hold. Then, under H_0, the test statistic $t \xrightarrow{d} \int_0^1 \mathbb{G}(\beta)^2 d\beta$ as $\min\{N_1, N_2\} \to \infty$, where $\mathbb{G}$ is a Gaussian process with a mean of zero and a covariance function of*

$$cov\left(\mathbb{G}(t), \mathbb{G}(s)\right) = (1-\lambda) \cdot cov\left(T_1\left(\mathbb{G}_1\right)(t), T_1\left(\mathbb{G}_1\right)(s)\right) + \lambda \cdot cov\left(T_2\left(\mathbb{G}_2\right)(t), T_2\left(\mathbb{G}_2\right)(s)\right).$$

[7] See Proposition 7.5 of Kosorok (2008).

Proof. See Appendix B.3. □

Theorem 1 provides the asymptotic distribution of our test statistic under the null hypothesis. It states that our test statistic converges in distribution to the square of the L^2-norm of a Gaussian process with mean zero when sample sizes are large enough. Theorem 1 is important for two reasons. First, the test statistic converges to the asymptotic distribution at a parametric rate, although our testing procedure is fully nonparametric. This feature makes our test applicable in data sets with moderate sample sizes. Second, the asymptotic distribution of the test statistic under the null hypothesis can be well characterized since it is essentially the integral of a squared Gaussian process.

Based on the asymptotic null distribution in Theorem 1, one candidate of critical values can be given as follows

$$\{c_{1-\alpha} : c_{1-\alpha} \text{ is the } (1-\alpha)\text{-th quantile of the distribution of } \int_0^1 \widehat{\mathbb{G}}(\beta)^2 d\beta\},$$

where $\widehat{\mathbb{G}}(\cdot)$ is obtained by replacing $g(b(\cdot|I_k)|I_k)$ with its nonparametric estimate $\widehat{g}(\widehat{b}(\cdot|I_k)|I_k)$ in $\mathbb{G}(\cdot)$. However, this plug-in type of critical values is undesirable in practice, since it involves the estimation of bid density. Instead, we propose data-driven critical values by a bootstrap approach in next subsection.

3.2 Bootstrap Critical Value

In this subsection, we investigate the properties of critical value for the test statistic from bootstrapping. The bootstrap treats the given data as if they were the population. It develops the bootstrap empirical distribution of the test statistic by repeatedly sampling the given data and computing the test statistic from the resulting bootstrap samples.

Specifically, we can bootstrap by independently drawing samples of size N_k with replacement from each of the two original samples for $k = 1, 2$. Let $\mathcal{D}_k$, $k = 1, 2$, denote the original sample $\{B_{k,1}, \ldots, B_{k,N_k}\}$. Let $\{B^m_{1,1}, \ldots, B^m_{1,N_1}\}$ and $\{B^m_{2,1}, \ldots, B^m_{2,N_2}\}$ be the pair of bootstrap samples, where $m = 1, \ldots, M$ and M is the number of bootstrap sample pairs. For each bootstrap sample pair m, we can compute a value of the bootstrap test statistic as

$$t^m = \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \int_0^1 \left(\left(\widehat{V}^m(\beta|I_1) - \widehat{V}^m(\beta|I_2) \right) - \left(\widehat{V}(\beta|I_1) - \widehat{V}(\beta|I_2) \right) \right)^2 d\beta,$$

where $\widehat{V}^m(\cdot|I_k)$ is computed by the mth bootstrap sample drawn from original sample k, and $\widehat{V}(\cdot|I_k)$ is computed by the original sample k. We then define the bootstrap critical value as

$$\{c^M_{1-\alpha} : (1-\alpha)\text{-th quantile of bootstrap statistics} \{t^1, \ldots, t^M\}\},$$

whose asymptotic properties are given in the following theorem:

Theorem 2. *Suppose Assumptions 1, 2 and 3 hold. Then the bootstrapping critical value $c_{1-\alpha}^M$ satisfies:*

1. *For any $(b(\cdot|I_1), b(\cdot|I_2)) \in H_0$, $\lim_{N_1,N_2,M\to\infty} \Pr(t > c_{1-\alpha}^M) = \alpha$ for any $\alpha \in (0,1)$;*
2. *For any $(b(\cdot|I_1), b(\cdot|I_2)) \in H_1$, $\lim_{N_1,N_2,M\to\infty} \Pr(t > c_{1-\alpha}^M) = 1$ for any $\alpha \in (0,1)$.*

Proof. See Appendix B.4. □

Theorem 2 shows that the bootstrap critical value has two properties. Property 1 says that the critical value has the correct size asymptotically, and Property 2 shows that, as sample sizes are large enough, the test based on the bootstrap critical value will almost surely reject the null hypothesis for any size value in $(0,1)$ and any Data Generating Process (DGP) in the alternative. Consequently, our test based on the bootstrap critical value has correct size asymptotically and is also consistent; i.e. our bootstrap critical value is asymptotically valid.

3.3 Asymptotic Local Power

We now study the asymptotic local power properties of our test. We consider the following class of local alternatives,

$$H_{1n}: \quad V(\beta|I_2) = V(\beta|I_1) + n^{-\gamma} \cdot h(\beta), \quad \forall \beta \in [0,1],$$

where $n = \frac{N_1 \cdot N_2}{N_1+N_2}$ and $h(\cdot)$ is nonzero for some β and is differentiable on $[0,1]$.[8] These local alternatives are equivalent to $v(\cdot|I_2) = v(\cdot|I_1) + n^{-\gamma} h'(\cdot)$. If $h(\cdot) = 0$, they degenerate to the null hypothesis H_0. The following theorem describes the local alternatives our test can detect.

Theorem 3. *Suppose that the DGPs satisfy the local alternative hypothesis H_{1n}. Then the following statements hold: Let $n \to \infty$,*

1. *If $\gamma < \frac{1}{2}$, then the test statistic $t \xrightarrow{p} +\infty$;*
2. *If $\gamma = \frac{1}{2}$, then $t \xrightarrow{d} \int_0^1 (\mathbb{G}(\beta) - h(\beta))^2 d\beta$, where $\mathbb{G}(\cdot)$ is the process defined by Theorem 1;*
3. *If $\gamma > \frac{1}{2}$, then the test statistic t converges in distribution to $\int_0^1 \mathbb{G}(\beta)^2 d\beta$, which is the asymptotic distribution of t under H_0.*

Proof. See Appendix B.5. □

Theorem 3 shows that, despite our test being fully nonparametric, it has non-trivial power against local alternatives approaching the null hypothesis at a rate of root-N. When the local alternatives approach the null hypothesis at a rate slower than root-N, our test will almost surely reject them under large samples. However, the power of our test becomes the nominal size as the local alternatives go to the null hypothesis at a rate faster than root-N.

[8] Notice that $h(0) = 0$ since $V(0|I_1) = V(0|I_2) = 0$. Moreover, the differentiability of $h(\cdot)$ is due to the differentiability of $V(\cdot|I_1)$ and $V(\cdot|I_2)$ under Assumption 2.

4 Finite Sample Performance

To study the finite sample performance of our testing procedure, we conduct Monte Carlo experiments. Unless stated otherwise, we consider two groups of auctions: one group has $I_1 = 3$ bidders, and the other group has $I_2 = 7$ bidders. The true valuation distribution of group k is

$$F(v|I_k) = \begin{cases} 0 & \text{if } v < 0, \\ v^{\gamma_k} & \text{if } 0 \leq v \leq 1, \\ 1 & \text{if } v > 1, \end{cases} \tag{8}$$

where $\gamma_k > 0$ and $k = 1, 2$.[9] Such a choice of private value distributions is convenient since they correspond to linear bidding strategies as follows:

$$s(v|I_k) = \big(1 - \frac{1}{\gamma_k(I_k - 1) + 1}\big) \cdot v, \tag{9}$$

which is derived in Appendix A.2.

The number of Monte Carlo replications is 1000. For each replication, we first generate randomly $N_1 = I_1 \cdot L_1$ and $N_2 = I_2 \cdot L_2$ private values from $F(\cdot|I_1)$ and $F(\cdot|I_2)$, respectively. Second, we calculate the corresponding bids $B_{1,i}$ and $B_{2,i}$ using the linear bidding strategies in Equation (9). Third, we compute the one-step test statistic t using Equation (5). Fourth, we obtain the bootstrap critical value by applying the bootstrapping procedure described in Section 3.2 with 1000 pairs of bootstrap samples. Comparing the test statistic and the bootstrap critical value, we conclude whether the null hypothesis H_0 can be rejected for this Monte Carlo replication. We can then obtain the simulated rejection rate based on the rejection rate of these 1000 Monte Carlo replications.

We now conduct several experiments to study the size and local power of our test in finite samples.

4.1 Size

We first study the size of our test, that is, the probability that the test will reject the null hypothesis when it is true. In this experiment, we consider $\gamma_1 = \gamma_2 \in \{0.25, 0.50\}$, $N_1 = N_2 \in \{105, 525, 735\}$, and the size $\alpha \in \{0.10, 0.05, 0.01\}$. The results are summarized in Table 1.

Table 1 shows that our test has good size properties with moderate sample sizes. Consider $\gamma_1 = \gamma_2 = 0.5$. When $N_1 = N_2 = 105$ (i.e., the number of auctions $L_1 = 35$ and $L_2 = 15$), the simulated rejection rates are 0.0920, 0.0480 and 0.0080, respectively. They are close to their nominal values.

[9]We adopt the setup of the Monte Carlo simulations from Marmer and Shneyerov (2012).

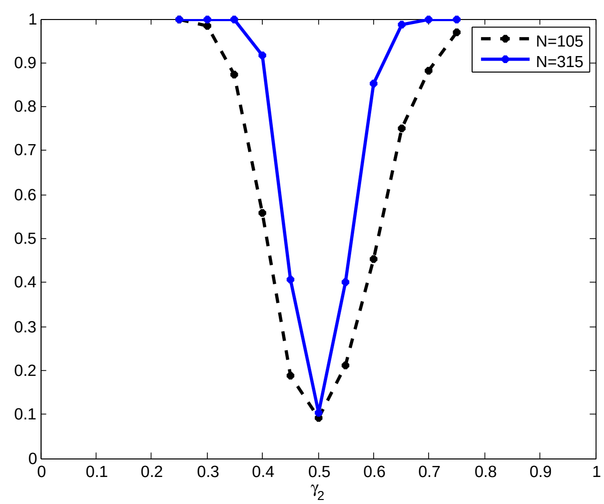
Table 1: Simulated Size for $I_1 = 3, I_2 = 7$ and $\gamma_1 = \gamma_2 = \gamma$

	$N = 105$			$N = 525$			$N = 735$		
Nominal size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\gamma = 0.25$	0.1060	0.0550	0.0090	0.0880	0.0470	0.0040	0.1000	0.0550	0.0120
$\gamma = 0.5$	0.0920	0.0480	0.0080	0.1020	0.0490	0.0130	0.1030	0.0610	0.0150

4.2 Power

We now study the power of our test, namely, the probability that the test will reject the null hypothesis when it is false. First, we examine the power of our test against fixed alternatives. Figure 2 displays the simulated rejection rate for a nominal size of $\alpha = 0.10$. We fix $\gamma_1 = 0.5$, and let γ_2 vary between 0.25 and 0.75 with a step size of 0.05 and sample size $N_1 = N_2 \in \{105, 315\}$.

Figure 2: Simulated Rejection Rate ($\gamma_1 = 0.5$, $\gamma_2 \in [0.25, 0.75]$, $\alpha = 0.10$)



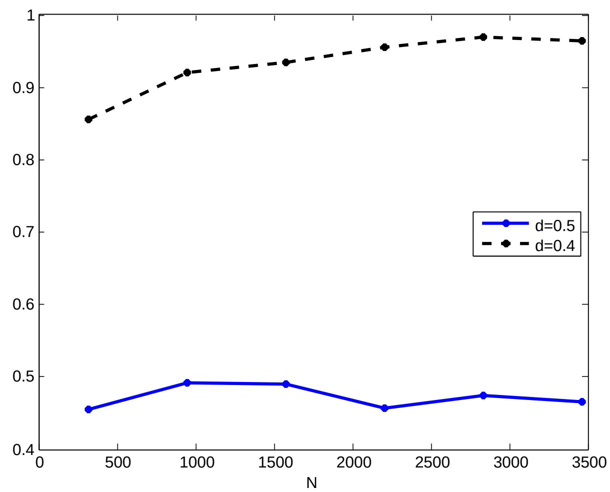
As shown in Figure 2, for a given sample size, the rejection rate converges to 100% when γ_2 moves further away from $\gamma_1 = 0.5$; and for a given value of γ_2, the rejection rate is higher when the sample size is larger.

Now we examine the local power of our test, that is, the power of the test when the alternatives approach the null hypothesis at some given rate when the sample size increases. In particular, we fix $\gamma_1 = 0.5$, and let γ_2 approach γ_1 as the sample size increases

$$\gamma_2 = 0.5 + \frac{1}{N^d},$$

where $N_1 = N_2 = N$. Note that $v(\beta|I_2) = \beta^{1/\gamma_2}$. Consider a large N. A Taylor expansion gives $v(\beta|I_2) \approx v(\beta|I_1) - \frac{\beta^{1/\gamma_1}\log\beta}{\gamma_1^2}\frac{1}{N^d}$. Thus, Theorem 3 applies. We let the significance level $\alpha = 0.10$, and the sample size N increase from 315 to 3465 in increments of 630. Figure 3

Figure 3: Simulated Local Power ($\gamma_1 = 0.5$, $\gamma_2 = 0.5 + 1/N^d$, $\alpha = 0.10$)



displays the simulated local power of our test for $d = 0.5$ (solid line) and $d = 0.4$ (dashed line). The former shows the probabilities of detecting the root-N local alternatives. Its simulated rejection rate is greater than 40%, which is significantly higher than the nominal size $\alpha = 0.10$. When $d = 0.4$, the rejection rate increases towards 100% as the sample size N increases. In sum, Figure 3 offers some evidence that our test can detect the local alternatives converging to the null hypothesis at a rate of root-N.

5 Extensions

Up to now, we have assumed that the auction objects are identical, the reserve price is nonbinding, and the bidders are symmetric and risk neutral. It is trivial to generalize our testing procedure to allow for auction-specific heterogeneity. Let $X \in \mathbb{R}^d$ be a random vector that describes the heterogeneity of auctions. Notice that, with auction-specific heterogeneity, our testing procedure involves the curse of dimensionality due to the estimation of the conditional quantile function $b(\cdot|I_k, x)$. However, its convergence rate is still faster than estimating $g(\cdot|I_k, x)$ if one were to compare the valuation distribution (quantile) functions based on the sample of pseudo values.

In this section, we discuss how to adapt our test to a variety of settings.

5.1 General Sample Case

There can be more than two samples in many bid data sets. We now generalize our testing problem to this case. Suppose that there are $K \geq 2$ independent i.i.d. samples $\{B_{k,1}, ..., B_{k,N_k}\}$,

$k = 1, \ldots, K$, where $N_k = I_k \times L_k$ is the number of bids generated from L_k first-price auctions with I_k bidders and valuation distribution $F(\cdot|I_k)$ for $k = 1, \ldots, K$. The hypothesis of interest becomes:

$$H_0^K : F(v|I_1) = \cdots = F(v|I_K), \forall v \in [\underline{v}, \overline{v}] \quad \text{v.s.} \quad H_1^K : \text{not } H_0^K,$$

which is equivalent to

$$H_0^K : V(\beta|I_1) = \cdots = V(\beta|I_K), \forall \beta \in [0,1] \quad \text{v.s.} \quad H_1^K : \text{not } H_0^K,$$

where $V(\cdot|I_k)$ is defined by Equation (4). We propose the following test statistic for this general problem:

$$t_K = \sum_{i=1}^{K} \sum_{j=i+1}^{K} w_{ij} \cdot \frac{N_i \cdot N_j}{N_i + N_j} \int_0^1 \left(\hat{V}(\beta|I_i) - \hat{V}(\beta|I_j) \right)^2 d\beta,$$

where w_{ij}s are some deterministic weights by choice (i.e., $w_{ij} \geq 0$ and $\sum_{i=1}^{K} \sum_{j=i+1}^{K} w_{ij} = 1$). The test statistic t_K can be viewed as a weighted average of the two-sample statistic t defined by Equation (5) across different pairs of bidder numbers. In Appendix A.3, we provide a characterization of the asymptotic null distribution of the test statistic t_K. In Appendix A.4, we compare three alternative proposals of weights in Monte Carlo simulations, namely: (i) sample size weights $w_{ij}^{\text{s.z.}} = (N_i + N_j) / \sum_{i=1}^{K} \sum_{j=i+1}^{K} (N_i + N_j)$ which give more weight to a pair with more observations; (ii) uniform weights $w_{ij}^{\text{U}} = 2/[K \cdot (K-1)]$; and (iii) the inverse asymptotic standard error weights $w_{ij}^{\text{i.s.e.}} = (1/\hat{\sigma}_{ij}) \Big/ \sum_{i=1}^{K} \sum_{j=i+1}^{K} 1/\hat{\sigma}_{ij}$ where $\hat{\sigma}_{ij}{}^2$ is the estimator of asymptotic variance of $N_i \cdot N_j / (N_i + N_j) \cdot \int_0^1 (\hat{V}(\beta|I_i) - \hat{V}(\beta|I_j))^2 d\beta$. Notice that both the sample size and inverse standard error weights are data-driven.

5.2 Endogenous Entry

Another possible extension of our approach is to models of endogenous entry discussed in Marmer, Shneyerov, and Xu (2013). They considered the selective entry model (SEM), which nests the Levin and Smith (1994) model of entry (LS) and the Samuelson (1985) model (S). In a selective entry model, a potential bidder observes a private signal correlated with his valuation of the good at the entry stage, which can be learned upon incurring entry cost. Bidders only know the number of potential bidders rather than the active ones.

Assume that the researcher knows the number of potential bidders I as well as the active bidders' bids.[10] Therefore, the participation probability $p(I)$ is identified. Marmer, Shneyerov,

[10]When the number of potential bidders I is not observed, An, Hu, and Shum (2010) showed how to apply the results of Hu (2008) to identify and estimate the model. See also Section 5.6 on the application of Hu (2008) to deal with unobserved heterogeneity. Under the exogenous participation assumption, Shneyerov and Wong (2011) adopted a different identification strategy which allows asymmetric bidders. Moreover, they generalized their results to the case where only the winning bid is observed.

and Xu (2013) showed that the inverse bidding strategy is identifiable as

$$\xi(b|I) = b + \frac{1}{I-1}\left(\frac{G^*(b|I)}{g^*(b|I)} + \frac{1-p(I)}{p(I)}\frac{1}{g^*(b|I)}\right),$$

where $p(I)$ is the equilibrium probability of bidding and $G^*(\cdot|I)$ and $g^*(\cdot|I)$ are the conditional distribution and density of active bidders' bids, respectively. They also show that, as the number of potential bidders increases, those who enter tend to have larger valuations. Denote $v(\alpha|I) = \xi(b(\alpha|I)|I)$, where $b(\alpha|I) \equiv G^{*-1}(\cdot|I)$. By the selection effect, the restriction of the SEM, LS, and S model can be stated as follows: if $I_1 < I_2$,

$$\begin{aligned} H_{SEM}: &\quad v(\alpha|I_1) \le v(\alpha|I_2), \\ H_{LS}: &\quad v(\alpha|I_1) = v(\alpha|I_2), \\ H_S: &\quad v\left(1 - \frac{p(I_2)}{p(I_1)}(1-\alpha)|I_1\right) = v(\alpha|I_2) \end{aligned}$$

They consider testing H_{SEM}, H_{LS} and H_S against their corresponding unrestricted alternatives. Test statistics are based on pairwise differences between the sample quantiles corresponding to different numbers of potential bidders. Note that H_{LS} and H_S are equivalent to

$$\begin{aligned} \int_0^\beta v(\alpha|I_1)d\alpha &= \int_0^\beta v(\alpha|I_2)d\alpha, \\ \int_0^\beta v\left(1 - \frac{p(I_2)}{p(I_1)}(1-\alpha)|I_1\right)d\alpha &= \int_0^\beta v(\alpha|I_2)d\alpha, \end{aligned}$$

respectively. On the other hand, H_{SEM} implies $\int_0^\beta v(\alpha|I_1)d\alpha \le \int_0^\beta v(\alpha|I_2)d\alpha$, but the reverse is not true.

We now consider the integrated quantile functions. After some algebra, we obtain:

$$\begin{aligned} \int_0^\beta v(\alpha|I_k)d\alpha &= \frac{I_k-2}{I_k-1}\int_0^\beta b(\alpha|I_k)d\alpha + \frac{1}{I_k-1}b(\beta|I_k)\left[\beta + \frac{1-p(I_k)}{p(I_k)}\right], \\ \int_0^\beta v\left(1 - \frac{p(I_2)}{p(I_1)}(1-\alpha)|I_1\right)d\alpha &= \frac{p(I_1)}{p(I_2)}\int_{1-\frac{p(I_2)}{p(I_1)}}^{1-\frac{p(I_2)}{p(I_1)}(1-\beta)} v(\alpha|I_1)d\alpha, \end{aligned}$$

where $\beta \in [0,1]$ and $k = 1, 2$. Both functions can be estimated at a root-N rate, which involves no density estimation.

In sum, we can construct tests which are equivalent to testing H_{LS} and H_S. Test statistics can be defined similarly as in Equation (5). However, our method can only test an implication of H_{SEM}. A test statistic can be defined similarly as in Equation (11).

5.3 Only Winning Bids are Recorded

In practice, the researcher might only observe the winning bids. Guerre, Perrigne, and Vuong (2000) showed that winning bids are sufficient to identify the model. Our method can be adapted to this problem. By definition, the distribution of a winning bid is $G^*(b) = G(b)^I$. This implies that the bid quantile function can be defined as follows

$$b(\alpha) \equiv G^{-1}(\alpha) = G^{*-1}(\alpha^I) \equiv b^*(\alpha^I).$$

The integrated-quantile function then becomes

$$V(\beta|I) = \frac{I-2}{I-1}\int_0^\beta b^*(\alpha^I|I)d\alpha + \frac{1}{I-1}b^*(\beta^I|I)\beta = \frac{I-2}{I(I-1)}\int_0^{\beta^I} \frac{b^*(x|I)}{x^{(I-1)/I}}dx + \frac{1}{I-1}b^*(\beta^I|I)\beta,$$

where the last equation follows from the change-of-variable. $V(\beta|I)$ can be estimated by replacing $b^*(\cdot|I)$ by its empirical counterpart $\widehat{b}^*(\cdot|I)$. Note that $\widehat{b}^*(\cdot|I)$ is piecewise constant. The integration on the right-hand side is easy to obtain. Finally, a test statistic can be defined similarly as in Equation (5).

5.4 Reserve Price

We now consider a binding reserve price b_0, which introduces a truncation because a potential bidder with a valuation lower than b_0 does not bid. Guerre, Perrigne, and Vuong (2000) showed that the inverse bidding strategy is identifiable as

$$\xi(b|I) = b + \frac{1}{I-1}\left(\frac{G^*(b|I)}{g^*(b|I)} + \frac{1-F(b_0|I)}{F(b_0|I)}\frac{1}{g^*(b|I)}\right),$$

where I is the number of potential bidders, $F(b_0)$ is the equilibrium probability of bidding and $G^*(\cdot|I)$ and $g^*(\cdot|I)$ are the conditional distribution and density of active bidders' bids, respectively. Note that I and $F(b_0)$ are unknown. Since the distribution of the number of actual bidders I^*_ℓ is binomial with parameters $(I, 1-p)$, Guerre, Perrigne, and Vuong (2000) showed that I is identified and $F(\cdot)$ is identified on $[b_0, \overline{v}]$.

After some algebra, the quantile representation of the identification equation implies that

$$V^*(\beta|I) \equiv \int_0^\beta v^*(\alpha|I)d\alpha = \frac{I-2}{I-1}\int_0^\beta b^*(\alpha|I)d\alpha + \frac{1}{I-1}b^*(\beta|I)\left[\beta + \frac{1-F(b_0|I)}{F(b_0|I)}\right],$$

which can be estimated by replacing I, $b^*(\alpha|I)$, and $F(b_0|I)$ by their empirical counterparts $\widehat{I}$, $\widehat{b}^*(\alpha|I)$, and $\widehat{F}(b_0|I)$. In particular, we follow Guerre, Perrigne, and Vuong (2000) to: first estimate I by $\widehat{I} = \max_{\ell=1,\ldots,L} I^*_\ell$, where I^*_ℓ is the number of active bids in auction ℓ; then estimate $F(b_0|I)$ by $\widehat{F}(b_0|I) = 1 - \frac{\sum_{\ell=1}^L I^*_\ell}{\widehat{I}L}$. The later follows from the fact that $E[I^*_\ell] = I(1-F(b_0))$ and $E[I^*_\ell]$ can be estimated by $\sum_{\ell=1}^L I^*_\ell / L$. Let the estimate of $V^*(\beta|I)$ be $\widehat{V}^*(\beta|I)$. Finally, a test statistic can be defined similarly as in Equation (5).

5.5 Risk Aversion

Risk aversion has been shown to be an important component of bidders' behavior in auctions. See, e.g., Athey and Levin (2001), Lu and Perrigne (2008), and Campo, Guerre, Perrigne, and Vuong (2011) for empirical literature; and Cox, Smith, and Walker (1988) and Bajari and Hortaçsu (2005) for experimental literature.

We now generalize our test to allow for risk averse bidders. Let $U(\cdot) : R_+ \to R$ be the bidders' utility function with $U(0) = 0$, $U'(\cdot) > 0$ and $U''(\cdot) < 0$. Let $\lambda(\cdot) = U(\cdot)/U'(\cdot)$ and $\lambda^{-1}(\cdot)$ be its inverse function. Note that $\lambda^{-1\prime}(\cdot) \in (0,1)$ for risk averse bidders.[11] Denote $R(\alpha|I_k) \equiv \frac{1}{I_k-1}\frac{\alpha}{g(b(\alpha|I_k)|I_k)}$, where $k = 1,2$ and $I_1 < I_2$.

Under the null hypothesis that $F(\cdot|I_1) = F(\cdot|I_2)$, Guerre, Perrigne, and Vuong (2009) gave the compatibility conditions:

$$b(\alpha|I_1) + \lambda^{-1}(R(\alpha|I_1)) = b(\alpha|I_2) + \lambda^{-1}(R(\alpha|I_2)),$$

where $\alpha \in [0,1]$. They also showed that $R(\alpha|I_1) > R(\alpha|I_2) > 0$. Therefore,

$$\begin{aligned} b(\alpha|I_2) - b(\alpha|I_1) = \lambda^{-1}(R(\alpha|I_1)) - \lambda^{-1}(R(\alpha|I_2)) =& \lambda^{-1\prime}(\widetilde{R(\alpha)}) \times (R(\alpha|I_1) - R(\alpha|I_2)) \\ <& R(\alpha|I_1) - R(\alpha|I_2), \end{aligned}$$

where $\widetilde{R(\alpha)} \in (R(\alpha|I_1), R(\alpha|I_2))$. The second equality follows from the Mean Value Theorem. The inequality follows from $\lambda^{-1\prime}(\cdot) \in (0,1)$. Thus,

$$b(\alpha|I_1) + R(\alpha|I_1) > b(\alpha|I_2) + R(\alpha|I_2),$$

which implies that, for risk averse bidders,

$$\tilde{V}(\beta|I_1) > \tilde{V}(\beta|I_2), \quad \textit{for any } \beta \in (0,1], \tag{10}$$

where $\tilde{V}(\beta|I_k) \equiv \int_0^{\beta}[b(\alpha|I_k) + R(\alpha|I_k)]d\alpha$.[12] In addition, our previous discussion shows that, for risk neutral bidders, the null hypothesis of $F(\cdot|I_1) = F(\cdot|I_2)$ is equivalent to $\tilde{V}(\cdot|I_1) = \tilde{V}(\cdot|I_2)$. Consequently, for both the risk averse and risk neutral bidders, the null hypothesis of $F(\cdot|I_1) = F(\cdot|I_2)$ implies that

$$\tilde{V}(\beta|I_1) \geq \tilde{V}(\beta|I_2), \quad \textit{for any } \beta \in (0,1],$$

which can be tested by a procedure adapted from ours. In particular, we consider a test

[11] Without the normalization of $U(0) = 0$, it is easy to show that $\lambda(\cdot) = [U(\cdot) - U(0)]/U'(\cdot)$ and $\lambda'(\cdot) = 1 - [U(\cdot) - U(0)] \cdot U''(\cdot)/U'(\cdot)^2$. In this case, we still have $\lambda'(x) > 1$ for all $x > 0$ (and hence $\lambda^{-1\prime}(\cdot) \in (0,1)$) for risk averse bidders. Moreover, $\lambda(\cdot)$ and $\lambda'(\cdot)$ are unaffected by a strictly increasing affine transformation of $U(\cdot)$.

[12] When bidders are risk neutral, it is easy to see that $\tilde{V}(\cdot|I_k) = V(\cdot|I_k)$, where $V(\cdot|I_k)$ is the integrated quantile function of bidders' valuation distribution.

statistic that only penalizes one-sided deviations:[13]

$$t_+ = \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \int_0^1 \left| \widehat{\tilde{V}}(\beta|I_2) - \widehat{\tilde{V}}(\beta|I_1) \right|_+ d\beta. \tag{11}$$

If the bidders have a Constant Relative Risk Aversion (CRRA) utility function $U(x) = x^\theta$ where $\theta \in (0, 1]$, we have $\lambda^{-1}(x) = \theta x$ and

$$V(\beta; \theta|I) = \frac{I - 1 - \theta}{I - 1} \cdot \int_0^\beta b(\alpha|I) d\alpha + \frac{\theta}{I - 1} \cdot b(\beta|I) \cdot \beta.$$

The hypothesis of $F(\cdot|I_1) = F(\cdot|I_2)$ is then equivalent to $V(\cdot; \theta_1|I_1) = V(\cdot; \theta_2|I_2)$, where θ_1 and θ_2 are the CRRA parameters in the two samples. Therefore, a statistic for testing equality of valuation distributions can be defined similarly as in Equation (5).

On the other hand, if one maintains the assumption that $F(\cdot|I_1) = F(\cdot|I_2)$, our benchmark test can be used to detect risk aversion. Specifically, if $F(\cdot|I_1) = F(\cdot|I_2)$, risk neutrality can then be discriminated from risk aversion by testing $\tilde{V}(\cdot|I_1) = \tilde{V}(\cdot|I_2)$ against (10).

5.6 Unobserved Heterogeneity

There are two main approaches to dealing with unobserved heterogeneity in auctions. The first is the deconvolution method, in which the unobserved heterogeneity has either an additive or multiplicative effect on bidder values (See, Li and Vuong (1998) and Krasnokutskaya (2011)). This method requires at least two bids per auction. The second method, proposed by Hu, McAdams, and Shum (2013) (who built upon earlier work by Hu (2008)), allows a finite number of non-separable states (i.e. unobserved heterogeneity), and requires at least three bids per auction. In this latter method, the distribution of bidder values is monotonic in the state variable. While the deconvolution method requires fewer bidders than the second method, it is known to suffer from a slow convergence rate. The Hu, McAdams, and Shum (2013) method, on the other hand, leads to a root-N convergence rate for the empirical conditional CDF of the bids. See Hu (2008) and An, Hu, and Shum (2010) for details.

In light of these results, we adapt the Hu, McAdams, and Shum (2013) method for incorporating unobserved heterogeneity. Bidders' private values are i.i.d. conditional on an auction-specific state $\omega \in \{1, \ldots, \Omega\}$, where ω is common information among bidders but unknown to the analyst. Moreover, Ω is known to the analyst. Denote the conditional distribution as $F_\omega(\cdot)$. We assume that $F_\omega(v) \leq F_{\omega'}(v)$ for all v and $\omega < \omega'$.

We now discuss how our test can be adapted. In the first step, we invoke the estimation method in Hu, McAdams, and Shum (2013) to obtain the empirical conditional CDF for the bids $\widehat{G}_\omega(\cdot|I)$, based on which a sample of bids are constructed $\widehat{b}_{\omega j} \equiv \widehat{G}_\omega^{-1}(j/N|I)$, where $j = 1, \ldots, N$. In the second step, we apply our method based on the samples of bids.

[13] We thank referees for this point.

5.7 Asymmetric Bidders

Asymmetry among bidders can arise from: (i) different valuation distributions and/or (ii) different utility functions. We focus on case (i) here and will discuss the case where bidders have a Constant Relative Risk Aversion (CRRA) utility function in the empirical application.

A single and indivisible object is auctioned to I risk neutral bidders. For notation convenience, we consider two types of bidders. Types 1 and 0 consist of I_1 and I_0 bidders, respectively, with $I_1 + I_0 = I \geq 2$. For notational convenience, let $\mathbb{I}$ represent the vector (I_1, I_0). Denote the bidder's valuation in group 1 and 0 by v_{1i} and v_{0i}, which are i.i.d. draws from the distributions $F_1(\cdot)$ and $F_0(\cdot)$, respectively. Both distributions are absolutely continuous with support $[\underline{v}, \overline{v}]$.

Without loss of generality, we focus on bidder i of type 1 hereafter. He/She chooses a bid b_{1i} to maximize his/her expected profit

$$(v_{1i} - b_{1i}) \Pr(b_{1i} \geq B_{1i} \textit{ and } b_{1i} \geq B_0),$$

where $B_{1i} \equiv \max_{j \neq i} b_{1j}$ and $B_0 \equiv \max_i b_{0i}$. Following Guerre, Perrigne, and Vuong (2000) and Flambard and Perrigne (2006), the FOC can be rewritten as

$$v_{1i} = b_{1i} + \frac{1}{(I_1 - 1)\frac{g_1(b_{1i})}{G_1(b_{1i})} + I_0 \frac{g_0(b_{1i})}{G_0(b_{1i})}}, \tag{12}$$

where $g_\tau(\cdot)$ and $G_\tau(\cdot)$ are the bid density and distribution functions of type-τ bidders, and $\tau = 1, 0$.

Consider two independent i.i.d. samples who differ in competition $\mathbb{I}$. Without loss of generality, we consider testing whether the type-1 bidders in the two samples have the same value distribution, i.e. $H_0 : F_1(\cdot|\mathbb{I}_1) = F_1(\cdot|\mathbb{I}_2)$, where $\mathbb{I}_1 \neq \mathbb{I}_2$.

We now consider the integration of the quantile representation of Equation (12). In contrast to symmetric auctions, note that the denominator of the second term on the right-hand side is a mixture of the two ratios from the two types of bidders. Therefore, $g_k(\cdot)$ remains in the integrated-quantile function. Thus, our integrated-quantile-based test may be extended to asymmetric auctions; but this is done at the cost of including density estimation.

5.8 Parametric Specification in Value Distribution

Many auction datasets are relatively small, and parametric estimation is often used to analyze these datasets. In this case, our integrated value quantile approach can be used to test a parametric specification on the private value distribution.[14] Suppose that the distribution of bidders' valuations is specified up to a finite dimensional vector of parameters $\theta \in \Theta$ where Θ is a compact subset of $\mathbb{R}^n$. Let $V(\cdot|\theta)$ denote the integrated quantile function of a parametric value distribution $F(\cdot|\theta)$ for $\theta \in \Theta$. The null and alternative hypotheses of this

[14]We thank one referee for pointing this out.

specification test are respectively given as follows:

$$H_0^{\Theta} : V(\cdot) \in \{V(\cdot|\theta) : \theta \in \Theta\} \quad \text{v.s.} \quad H_1^{\Theta} : \text{not } H_0^{\Theta},$$

where $V(\cdot)$ is the integrated quantile function of the true value distribution.

For this specification test, we can propose a quadratic type of statistic as $t^{\Theta} = N \cdot \int (\hat{V}(\beta) - V(\beta|\hat{\theta}))^2 d\beta$, where $\hat{V}(\cdot)$ is the unrestricted estimator of the integrated value quantile function obtained by replacing the bid quantile with its sample analogue in Equation (4), $V(\cdot|\hat{\theta})$ is the restricted estimator of the integrated quantile function of value distribution in $\{V(\cdot|\theta) : \theta \in \Theta\}$, and N is the sample size. It is easy to see that the null hypothesis H_0^{Θ} will be rejected when the test statistic t^{Θ} is large enough. Establishing the asymptotic properties of such a testing procedure is left for future research.

6 Application

In this section, we apply our test to some real-life data. We study the United States Forest Service (USFS) timber auctions, whereby timber harvesting rights on publicly owned forests are sold to private sector entities. Detailed descriptions of such timber auctions can be found in, e.g., Baldwin, Marshall, and Richard (1997), Haile (2001) and Haile and Tamer (2003). In particular, we analyze the sealed-bid auction data from 1982 to 1990.

In a typical scenario, the USFS conducts a cruise of the tract to be auctioned prior to announcement of the auction. On the basis of the cruise report, the USFS mechanically determines a reserve price. At least 30 days prior to the auction, USFS advertises the sale, announces the reserve price for the tract, and makes the fully disaggregated information in the cruise report publicly available.

6.1 Data

We now describe how we construct our dataset from the raw data available on Philip A. Haile's website. Following the literature, we focus on a subset of the timber auctions that are most likely to satisfy the independent private value assumption. See, e.g., Baldwin, Marshall, and Richard (1997), Haile (2001) and Haile and Tamer (2003). Following Haile, Hong, and Shum (2003), we consider only scaled sales from 1982 to 1990 in Forest Service regions 1 and 5. In view of policy effected changes in 1981 (see, e.g., Haile (2001)), this restriction minimizes the significance of subcontracting/resale, and thus common value. Moreover, we exclude salvage sales and sales set aside for small businesses. We also drop auctions with one bidder.

Table 2 describes the resulting sample sizes for auctions at each competition level (I_1, I_0), where I_1 and I_0 denote the number of small and big bidders, respectively. There are fairly few auctions with more than 5 bidders, which are neglected from our analysis.

The timber tracts are highly heterogeneous. Failing to control for auction-specific observ-

Table 2: Number of Auctions

$I_1 \backslash I_0$	0	1	2	3	4	5	Total
0	0	0	88	49	48	18	203
1	0	50	39	25	18	9	141
2	59	29	28	12	10	9	147
3	42	24	13	10	4	1	94
4	24	9	4	4	4	1	46
5	28	5	3	1	0	0	37
Total	153	117	175	101	84	38	668

ables would limit the relevance of our empirical results. Table 3 provides summary statistics on the winning bids and the auction-specific covariates. All dollar values are nominal and all volume values are in thousand board-feet (MBF) of timber. [15] We control for a list of auction-specific observables: the size of the tract (in acres), the estimated volume of timber (in MBF), the appraisal value (per MBF), the estimated selling value (per MBF), the estimated harvesting cost (per MBF), the estimated manufacturing cost (per MBF), the species concentration index (HHI) and the Forest Service region.

Table 3: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
winning bid	125773.1	238563.7	2193.75	1968985
acres	511.38	897.80	4	7000
volume	1641.37	2332.19	43	13181.86
AppValue_avg	34.00	31.29	0.5	189.76
SellValue_avg	351.06	64.16	153.05	539.83
LogCost_avg	132.67	29.98	53.69	290.72
MfgCost_avg	170.08	32.55	3.71	271.67
HHI	0.5656	0.2501	0.1457	1
D5	0.5210	0.4999	0	1

Note: D5=1 if the Forest Service region is 5, =0 otherwise.

Several relevant points are important to note. First, we consider a non-binding reserve price. As described in Baldwin, Marshall, and Richard (1997), the reserve price is calculated according to the “residual method” and known to be low. Second, bidders are asymmetric and potentially risk averse. Campo (2012) argued that bidders’ distributions of private information may depend on the firm’s size, distance from the lot, and physical and financial constraints. Moreover, attitudes toward risk may vary with their portfolio of assets or

[15]The volumes reported in the raw data are expressed in varying units of measurement. In most cases, non-MBF units can be readily converted to MBF units in a straightforward manner. There are, however, a few cases where the units cannot be converted (for e.g., Christmas trees, which are measured in tons, or "for each"). In these cases, we have used our best judgment in the conversion to MBF. Such cases are few and far between, so we do not expect the measurement error arising from them to be significant.

experience. Thus, she considered a general first-price auction model where bidders differ not only in their utility functions but also in their value distribution. As discussed in Athey, Levin, and Seira (2011), the bidders in timber auctions range from large mills to individually owned logging companies. We classify them into two groups by small business status, which requires no more than 500 employees. We allow the two types of bidders to differ in their valuation distributions and utility functions. For reasons that we will explain later, we allow risk aversion by assuming a CRRA utility function for the bidders.

Third, we abstract from unobserved heterogeneity in auctions. Although Hu, McAdams, and Shum (2013)'s approach can be generalized to asymmetric auctions, incorporating both unobserved heterogeneity and asymmetric bidders significantly complicates our analysis. Moreover, the cruise report contains detailed disaggregated information about the auction. This enables us to condition our analysis on a large number of auction covariates. We expect that there is little space for unobserved heterogeneity, and thus valuations are independent conditional on observables.

6.2 Testing Procedure

In the application, we consider auctions with two types of risk averse bidders: small and big. We classify bidders based on their small-business status, which requires no more than 500 employees. Bidders of the same type are symmetric. Moreover, we assume that bidders have a CRRA utility function: $U(x) = x^{\theta}$ where $\theta \in (0, 1]$. We allow small and big bidders to have different CRRA parameters θ_1 and θ_0 as well as different value distributions $F_1(\cdot)$ and $F_0(\cdot)$ with support $[\underline{v}, \overline{v}]$. The CRRA assumption comes with the benefit that the Haile, Hong, and Shum (2003) method of "homogenizing" the bids still applies.

Homogenizing the bids

Following Haile, Hong, and Shum (2003), we homogenize the bids before implementing our method to control for observable heterogeneity. If bidder i of type 1 has a CRRA utility function with parameter θ_1, Equation (12) can be written as

$$\frac{\theta_1}{v_{1i} - b_{1i}} = (I_1 - 1)\frac{g_1(b_{1i})}{G_1(b_{1i})} + I_0\frac{g_0(b_{1i})}{G_0(b_{1i})}, \tag{13}$$

which defines the equilibrium bidding strategy $s_1(\cdot)$ along with the boundary condition $\underline{b} = \underline{v}$. Following Haile, Hong, and Shum (2003), one can show that in another auction where value is a random variable $\widetilde{v}_i = \delta v_i$, the equilibrium bidding strategy is $\widetilde{b}_i = \delta s_1(v_i)$. To see this, note that $\widetilde{G}_j(\widetilde{b}_i) = Pr(\delta b_i \leq \widetilde{b}_i) = G(\widetilde{b}_i/\delta) = G(b_i)$ and $\widetilde{g}_j(\widetilde{b}_i) = g(\widetilde{b}_i/\delta)/\delta = g(b_i)/\delta$. It is obvious that the boundary condition $\underline{\widetilde{b}} = \underline{\widetilde{v}}$ and the following equation are both true:

$$\frac{\theta_1}{\widetilde{v}_{1i} - \widetilde{b}_{1i}} = (I_1 - 1)\frac{\widetilde{g}_1(\widetilde{b}_{1i})}{\widetilde{G}_1(\widetilde{b}_{1i})} + I_0\frac{\widetilde{g}_0(\widetilde{b}_{1i})}{\widetilde{G}_0(\widetilde{b}_{1i})}.$$

In particular, assume that

$$v(\epsilon|x,\mathbb{I}) = \delta(x) \times v(\epsilon|\mathbb{I}),$$

where ϵ is bidder's private value, which is independent of auction heterogeneity x. Our previous discussion implies that bidders' bidding strategy satisfies

$$b(\epsilon|x,\mathbb{I}) = \delta(x) \times b(\epsilon|\mathbb{I}),$$

which we rewrite as

$$\log b(\epsilon|x,\mathbb{I}) = \delta_0 + b_0(\mathbb{I}) + \widetilde{\delta}(x) + \widetilde{b}(\epsilon|\mathbb{I}),$$

where $\delta_0 = E[\log \delta(X)]$, $b_0(\mathbb{I}) = E[\log b(\epsilon|\mathbb{I})|\mathbb{I}]$, $\widetilde{\delta}(x) = \log \delta(x) - \delta_0$ and $\widetilde{b}(\epsilon|\mathbb{I}) = \log b(\epsilon|\mathbb{I}) - b_0(\mathbb{I})$. Moreover, the term $\widetilde{b}(\epsilon|\mathbb{I})$ has mean zero conditional on $(\mathbb{I}, X)$ by independence of X and ϵ.

We define the homogenized bid as the bid that a bidder would have submitted in equilibrium if she were in an "average" auction (i.e., $\log \delta(x) = E[\log \delta(X)]$) with $\mathbb{I}$ bidders:

$$b^*(\epsilon|\mathbb{I}) \equiv \exp(\delta_0 + b_0(\mathbb{I}) + \widetilde{b}(\epsilon|\mathbb{I})).$$

Implementation

Consider K samples with different competition levels: $\mathbb{I}_k$, where $k \in \{1,\ldots,K\}$. Our objective is to test whether the two types of bidders have the same value distributions in these samples. That is, $H_0 : F_\tau(\cdot|\mathbb{I}_k) = F_{\tau'}(\cdot|\mathbb{I}_{k'})$, where $k \neq k'$ and $\tau, \tau' \in \{1, 0\}$.

Our analysis involves the following steps:

Step 1: We apply the Haile, Hong, and Shum (2003) method to control for observed auction characteristics.

In particular, we regress the logarithm of the total bids ($logb$) on a host of control variables. Table 4 displays the regression results from the original sample. Regression (1) includes the following regressors: log(volume) and log(appraisal value per MBF). An R^2 of 86% indicates that the auction-specific heterogeneity is captured quite well by the appraisal value, in agreement with the literature. Regression (2) includes all the control variables as well as competition fixed effects (i.e., $b_0(\mathbb{I})$). All of the fitted coefficients have the expected signs. Bidders bid more when the selling value or the appraisal value is higher and bid less when the harvesting cost or the manufacturing cost is higher. Regression (3) adds year dummies which control for dynamic effects common to timber auctions, such as inflation. We use regression (3) as the baseline regression. The homogenized bids are calculated as the exponential of the differences between the logarithm of the original total bids and the demeaned fitted values of regression (3). Table 5 tabulates the means and standard deviations of the homogenized bids obtained from the original sample.

Step 2: We estimate the CRRA parameter θ_τ by comparing two samples with different

Table 4: Regression Results

VARIABLES	(1) OLS	(2) Fixed Effects	(3) Fixed Effects
log_acres		0.0190**	0.00724
		(0.00908)	(0.00935)
log_volume	1.103***	1.037***	1.054***
	(0.00883)	(0.0132)	(0.0135)
log_AppValue_avg	0.649***	0.484***	0.478***
	(0.0151)	(0.0242)	(0.0240)
log_SellValue_avg		1.097***	1.178***
		(0.113)	(0.114)
log_LogCost_avg		-0.366***	-0.399***
		(0.0584)	(0.0585)
log_MfgCost_avg		-0.262***	-0.322***
		(0.0377)	(0.0375)
HHI		-0.174***	-0.127***
		(0.0445)	(0.0440)
Observations	2,382	2,382	2,382
R-squared	0.859	0.882	0.885
Competition FE		YES	YES
Year FE			YES

Note: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

competition levels $\mathbb{I}_k$ and $\mathbb{I}_{k'}$ under the null hypothesis that bidders of the same type are symmetric. Such a strategy was proposed in Guerre, Perrigne, and Vuong (2009) and Campo, Guerre, Perrigne, and Vuong (2011). The idea is to exploit the property that the bid distribution varies with $\mathbb{I}$ while the valuation distribution does not. Applications include Bajari and Hortaçsu (2005) and Campo (2012).

Here, we utilize the property that the integrated quantile functions are the same for the same type of bidders, i.e. $V(\cdot;\theta_\tau|\mathbb{I}_k) = V(\cdot;\theta_\tau|\mathbb{I}_{k'})$. We use the symmetric auctions to obtain our estimates as they constitute the majority of our sample. Moreover, we do not need to worry about choosing a bandwidth. In particular, our estimator of (θ_1, θ_0) is

$$(\widehat{\theta}_1, \widehat{\theta}_0) = \arg\min_{\theta_1,\theta_0\in[0,1]} \widehat{Q}(\theta_1, \theta_0),$$

where $\widehat{Q}(\theta_1, \theta_0) = \sum_{\tau=0,1}\sum_{I=2}^{4}\sum_{I'=I+1}^{5} w_{I,I',\tau} \cdot \frac{N_{I\tau}\cdot N_{I'\tau}}{N_{I\tau}+N_{I'\tau}} \int_0^1 \left(\widehat{V}(\beta;\theta_\tau|I) - \widehat{V}(\beta;\theta_\tau|I')\right)^2 d\beta$. We use the sample size weights defined in Section 5.1. That is, $w_{I,I',\tau} = \frac{N_{I\tau}+N_{I'\tau}}{\sum_{I=2}^{4}\sum_{I'=I+1}^{5}(N_{I\tau}+N_{I'\tau})}$, where $N_{I\tau}$ is the number of bids in symmetric auctions with I type-τ bidders. In symmetric auctions, simple algebra yields the empirical counterpart of the integrated valuation quantile

Table 5: Means and Standard Deviations of the Homogenized Bids

$I_1 \backslash I_0$	0	1	2	3	4	5	Total
0	–	–	43696	44965	44307	44379	44300
	–	–	(24169)	(26070)	(25529)	(23724)	(24956)
1	–	43797	43884	44475	43952	45236	44165
	–	(22788)	(23084)	(25093)	(21662)	(26046)	(23470)
2	41745	44763	43875	42779	44623	45079	43637
	(18493)	(24828)	(23970)	(19684)	(24925)	(26918)	(22914)
3	41550	44397	45998	44223	42560	42899	43539
	(17239)	(21367)	(29717)	(23553)	(18150)	(20290)	(21861)
4	42529	41653	40831	41032	39983	39271	41534
	(19594)	(16742)	(12546)	(15489)	(10315)	(6685)	(16417)
5	44024	43276	40951	39692	–	–	43413
	(25220)	(20224)	(14160)	(9226)	–	–	(23047)
Total	42515	43879	43823	44029	43809	44524	43666
	(20563)	(21970)	(23862)	(23651)	(23193)	(24562)	(22853)

Note: Standard deviations in parentheses.

function

$$\widehat{V}(\beta;\theta|I_k) = B_{k,(i_k(\beta))} \times \beta + \frac{I_k - (1+\theta)}{N_k(I_k-1)}\Big[\sum_{j=1}^{i_k(\beta)} B_{k,(j)} - i_k(\beta) \times B_{k,(i_k(\beta))}\Big],$$

where $i_k(\beta)$ is chosen such that $\frac{i_k(\beta)-1}{N_k} < \beta \leq \frac{i_k(\beta)}{N_k}$.

Step 3: We calculate the test statistics for every pair of samples. In symmetric auctions, we follow Step 2 to calculate the test statistics. In asymmetric auctions, we first construct the pseudo values following Flambard and Perrigne (2006), which leads to a piecewise constant quantile function. We use the triweight kernel function and the rule-of-thumb bandwidth to estimate the bid density functions. The empirical counterpart of the integrated valuation quantile function can be calculated from these pseudo values. Finally, the test statistics are calculated similarly as in Equation (6).

We use a bootstrap procedure to obtain p-values. We repeat the above three steps 1000 times. For each replication, we independently draw samples of size $L_{\mathbb{I}}$, with replacement, from the auctions under competition level $\mathbb{I}$, where $L_{\mathbb{I}}$ is the number of auctions under competition level $\mathbb{I}$.[16] For each bootstrap sample pair, we can compute a value of the bootstrap test statistic $\widehat{t}^{m}_{\mathbb{I},\mathbb{I}'}$. Finally, we define the bootstrapped p-value as $\sum_{m=1}^{1000} 1(\widehat{t}^{m}_{\mathbb{I},\mathbb{I}'} > \widehat{t}_{\mathbb{I},\mathbb{I}'})/1000$, where $1(\cdot)$ is an indicator function.

[16]Our results are robust to alternative bootstrap resampling schemes: (1) resampling on auctions conditioning on competition level and bidder's type; (2) resampling on bids conditioning on competition level and bidder's type.

6.3 Testing Results

Before we present our empirical results, we add a few remarks on our method. Although our test can detect root-N local alternatives, the small sample size associated with some competition levels $\mathbb{I}$ may make it difficult to detect small differences between timber auction valuation distributions. In such cases, failure to reject the null hypothesis should be viewed as evidence that the difference in the valuation distributions is too small relative to other sources of variation in these auctions. In order to draw conclusions from our results with more confidence, we focus on the sample pairs where there are more than 50 bids of each type in the sample of the same $\mathbb{I}$. Moreover, we report the results by aggregating the test statistics across all sample pairs. Even though our method allows for a rich class of models, when one assumption is tested, others must be held fixed. A rejection of the null hypothesis may thus indicate violation of some maintained assumptions, such as private value. As described in Subsection 6.1, we focus on a subset of the timber auctions to minimize the significance of alternative explanations.

We now report the results for symmetric auctions. Before applying the formal test, we construct the pseudo values using the homogenized bids obtained from the original sample. In particular, we rely on Equation (13) to construct those pseudo values, and use the triweight kernel function and the rule-of-thumb bandwidth to estimate the bid density functions. Figure 4 shows the estimated valuation distributions based on those psudo values conditional on different competition levels. In general, they look similar.

Figure 4: Estimated Valuation Distributions

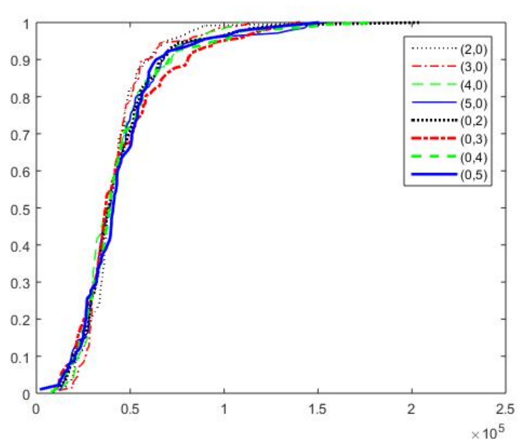


Table 6 reports the formal testing results. The top-left box compares auctions with only small bidders, while the bottom-right box compares auctions with only large bidders. All p-values are larger than 10%. Therefore, we do not reject the null hypothesis at conventional significance levels. Moreover, we use sample size weights to aggregate the test statistics for all samples pairs. The aggregated test statistic is $4.82e + 08$, which gives a p-value of 0.422. We obtain similar results when we compare small bidders and big bidders. On this basis, we conclude that the samples have similar valuation distributions after controlling for auction-specific covariates.

Table 6: Testing Results: p-value

(I_1, I_0)	(2,0)	(3,0)	(4,0)	(5,0)	(0,2)	(0,3)	(0,4)	(0,5)
(2,0)								
(3,0)	0.408							
(4,0)	0.382	0.334						
(5,0)	0.474	0.167	0.799					
(0,2)	0.519	0.338	0.761	0.941				
(0,3)	0.396	0.012	0.245	0.678	0.693			
(0,4)	0.515	0.132	0.720	0.991	0.415	0.582		
(0,5)	0.395	0.142	0.330	0.537	0.848	0.473	0.567	

We apply the same procedure to test whether the valuation distributions are the same for the bidders of the same type in the two sets of asymmetric auctions where $\mathbb{I} = (1, 1)$ and $(2, 2)$, respectively. The p-value is 0.346 for small bidders and 0.693 for big bidders. Again, we fail to reject the null hypothesis at conventional significance levels.

In sum, after controlling for the observed auction-specific heterogeneity, the difference between valuation distributions across competition levels and bidder types in our timber auction dataset appears to be small.

7 Conclusion

In this paper, we develop an integrated-quantile-based nonparametric test for comparing valuation distributions in first-price auctions. One important application of such a test is to justify by data the exogenous participation which assumes that the bidders' private value distributions are the same across auctions with different numbers of bidders. Our test is convenient in practice since it only involves one step estimation of quantile functions of bids. We show that, at a parametric rate, the test statistic converges to the square of the L^2-norm of a Gaussian process with mean zero under the null hypothesis. We propose bootstrap critical values, and we show that our test has the correct size and is consistent against all fixed alternatives. Moreover, our test detects local alternatives approaching the null at a parametric rate despite its nonparametric nature. Our simulation results show that our test behaves well in finite samples. Additionally, we demonstrate how our test can be extended to auctions with an arbitrary number of samples, auctions with endogenous entry, auctions where only winning bids are recorded; and we demonstrate how the test can be accommodated to cases where assumptions regarding reserve price, risk aversion, observed and unobserved heterogeneity, and asymmetric bidders are relaxed. Finally, we apply our method to data from U.S. Forest Service timber auctions.

A Supplementary Results

A.1 Weak Convergence of the Empirical Bid Quantile Process

Let $D[\underline{b}, \overline{b}]$ denote the collection of cadlag functions on $[\underline{b}, \overline{b}]$, $\mathbb{D}$ be the set of all distribution functions of measures that concentrate on $(\underline{b}, \overline{b}]$, $\ell^\infty(0,1)$ denote the set of all uniformly bounded real functions on $(0,1)$, and $C[\underline{b}, \overline{b}]$ be the collection of continuous functions from $[\underline{b}, \overline{b}]$ to $\mathbb{R}$. Our next lemma concerns the differentiability of the mapping from the bid distribution function to its quantile function. It is a direct application of part (ii) of Lemma 3.9.23 in van der Vaart and Wellner (1996) (or part (ii) of Lemma 12.8 in Kosorok (2008)).

Lemma 2. *Suppose a bid distribution $G(\cdot|I)$ satisfies Assumption 2. Then the inverse map $F \to F^{-1}$ as a map $\mathbb{D} \subset D[\underline{b}, \overline{b}] \to \ell^\infty(0,1)$ is Hadamard-differentiable at $G(\cdot|I)$ tangentially to $C[\underline{b}, \overline{b}]$. In addition, the derivative is the map $\alpha \to -(\alpha/g(\cdot|I)) \circ G^{-1}(\cdot|I)$.*

Proof. See Appendix B.2. □

Lemma 2 shows that the mapping from a cumulative distribution function (CDF) with compact support $[\underline{b}, \overline{b}]$ to its quantile function is Hadamard-differentiable at the bid CDF $G(\cdot|I)$ over the index set $(0,1)$.[17] Consequently, it may be used to obtain uniform convergence of the whole quantile process for bid distribution $G(\cdot|I)$. See, e.g., the (last four lines of) remark in Example 3.9.24 of van der Vaart and Wellner (1996). This lemma extends the well known result on quantile process which provides the Hadamard differentiability of the mapping between the CDF and the quantile function when the index set $[p,q]$ is in the interior of $[0,1]$. Nevertheless, it needs additional conditions for this extension: (i) the bid CDF $G(\cdot|I)$ has a compact support $[\underline{b}, \overline{b}]$; (ii) $G(\cdot|I)$ is continuously differentiable on its support with strictly positive derivative $g(\cdot|I)$. These two conditions are imposed by many papers in empirical auctions, see, e.g., Guerre, Perrigne, and Vuong (2000, 2009); Marmer and Shneyerov (2012); Marmer, Shneyerov, and Xu (2013). As an implication of Lemma 2, the following corollary characterizes the limiting behavior of the empirical bid quantile process.

Corollary 1. *Suppose Assumptions 1 and 2 hold. For $k = 1, 2$, as $N_k \to \infty$, the empirical quantile process $\widehat{\mathcal{G}}_k(\cdot) \equiv \sqrt{N_k}\left(\widehat{b}(\cdot \mid I_k) - b(\cdot \mid I_k)\right)$, weakly converges on $(0,1)$ to the ratio of $\mathbb{B}(\cdot)$ and $g(b(\cdot|I_k)|I_k)$, namely $\widehat{\mathcal{G}}_k(\cdot) \rightsquigarrow \mathbb{B}(\cdot)/g(b(\cdot|I_k)|I_k)$ on $(0,1)$.*

Corollary 1 obtains uniform convergence of the whole empirical bid quantile process. It is an immediate result from Lemma 2. Its proof is hence omitted.

A.2 Derivation of Linear Bidding Strategy in Simulation

We derive the linear bidding strategy Equation (9) in this subsection. Suppose that the reserve price is not binding, and the private value $v \in (0,1)$. By Riley and Samuelson (1981), in an auction with I_k bidders, the bidding strategy of a bidder with value of v is then given by

$$s(v|I_k) = v - \frac{1}{F(v)^{I_k-1}} \int_0^v F(u)^{I_k-1} du$$

[17] As a matter of fact, such a Hadamard-differentiability can be extended to the whole index set $[0,1]$ if we modify the definition of $G^{-1}(0|I) = \underline{b}$.

Consequently, under value distribution specified by Equation (8), the bidding strategy evaluated at v can be simplified as

$$\begin{aligned} s(v|I_k) &= v - \frac{1}{v^{\gamma_k(I_k-1)}} \int_0^v u^{\gamma_k(I_k-1)} du \\ &= \left(1 - \frac{1}{\gamma_k(I_k-1)+1}\right) \cdot v \end{aligned}$$

where the first equality is obtained by replacing the value CDF by Equation (8), and the last equality is got by simplifying the integration in the first one. This therefore completes the derivation of the linear bidding strategy (9) under value distribution (8).

A.3 Asymptotic Null Distribution of K-Sample Statistic

This subsection aims to provide the asymptotic null distribution of the weighted test statistic t_K defined in Section 5.1. We first give some regularity assumptions on the bid data-generating process.

Assumption 4. *(i) For any $I \in \{I_1, \ldots, I_K\}$, the random variables $B_1, \ldots, B_I$ are independent and identically distributed (i.i.d.) with common true distribution $G(\cdot|I)$.*

(ii) For any $I \in \{I_1, \ldots, I_K\}$, the bid distribution $G(\cdot \mid I)$ is continuously differentiable with a density $g(\cdot \mid I)$ on a support of $\left[\underline{b}, \overline{b}\right]$. In addition, the bid density $g(b|I) \geq c_g > 0$ for any $b \in [\underline{b}, \overline{b}]$.

(iii) For any $i < j \in \{1, \ldots, K\}$, $\frac{N_i}{N_i+N_j} \to \lambda_{ij} \in [0,1]$ as $\min\{N_1, \ldots, N_K\} \to \infty$.

Parts (i)-(iii) of Assumption 4 extend Assumptions 1 to 3 to the case of K samples, respectively. They require that each bids sample is an i.i.d. sample with a probability density function bounded away from zero on a compact support and the size ratio of any sample pair has a limit in large sample.

The next lemma characterizes the asymptotic null distribution of test statistic t_K under Assumption 4.

Lemma 3. *Suppose that Assumption 4 holds. Then, under $H_0^K : V(\cdot|I_1) = \cdots = V(\cdot|I_K)$, the K-sample test statistic t_K converges in distribution to*

$$\sum_{i=1}^{K} \sum_{j=i+1}^{K} w_{ij} \cdot \int_0^1 \left[\sqrt{1-\lambda_{ij}} \cdot T_i(\mathbb{G}_i)(\beta) - \sqrt{\lambda_{ij}} \cdot T_j(\mathbb{G}_j)(\beta)\right]^2 d\beta,$$

where mapping $T_k, k = 1, \ldots, K$, are defined by Equation (7), and $(\mathbb{G}_1(\cdot), \ldots, \mathbb{G}_K(\cdot))$ is a K-variate Gaussian process with independent components $\mathbb{G}_k(\cdot) = \mathbb{B}(\cdot)/g(b(\cdot|I_k)|I_k)$, $k = 1, \ldots K$.

Proof. See Appendix B.6. □

Lemma 3 shows that the asymptotic null distribution of test statistic t_K is a weighted average of the squared L^2-norm of a Gaussian process which is a difference of two independent Gaussian processes. Such an asymptotic null distribution can also be viewed as a continuous mapping of a K-variate Gaussian process whose components are independent (limiting) quantile processes.[18]

[18]See Barmi and Mukerjee (2005) for another example of K-variate Gaussian process with independent components.

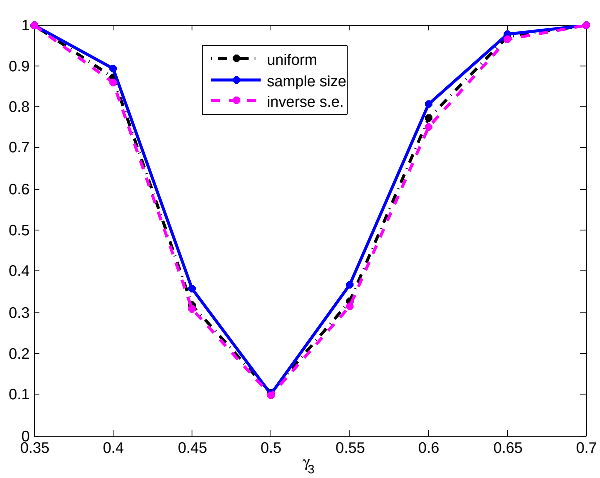
Based on Lemma 3, we can establish the validity of bootstrap in K-sample case. It can be accomplished by following similar proof to Theorem 2 which is in two sample case, and hence is omitted here.

A.4 Weights Comparison in Three Samples

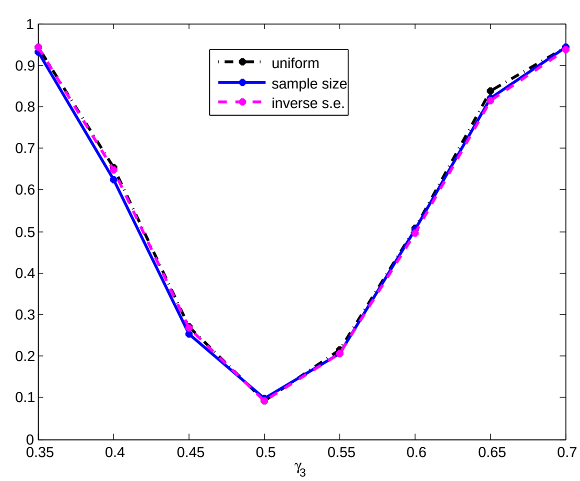
Section 5.1 discusses how to generalize our testing procedure to compare valuation distributions in more than two bid samples. This generalized testing procedure involves in a free choice of weights which can affect the power property. In the case of three samples, we use simulation to compare the following three proposals of weights which are defined in Section 5.1: (i) sample size weights, (ii) uniform weights, and (iii) inverse (asymptotic) standard error weights. In particular, the sample size weights give more weight to sample pair with larger size, and the inverse (asymptotic) standard error weights take the estimation error into account and put more weight on the pair which has smaller variance.[19]

In this experiment, we specify $\gamma_1 = \gamma_2 = 0.5$, and let γ_3 vary between 0.35 and 0.7 with a step size of 0.05 and number of bidders $I_1 = 3, I_2 = 5, I_3 = 7$. Furthermore, we consider two cases of sample sizes: (i) $N_1 = 105, N_2 = 210, N_3 = 420$, i.e. sample size is bigger when number of bidders is bigger, (ii) $N_1 = 420, N_2 = 210, N_3 = 105$, i.e. sample size is smaller when number of bidders is bigger. Their results are given in Figures 5a and 5b, respectively. They display the power curve of the sample size weights in solid line, the uniform weights in dash-dot line, and the inverse standard error weights in dashed line. Figures 5a and 5b show that the best choice of weights is case by case. Specifically, in case (i), Figure 5a shows that the sample size weights have the highest power among all of the three, and the uniform weights perform slightly better than the inverse standard error weights. In case (ii), as shown by Figure 5b, the uniform weights have the best performance in power, and there is no clear winner between the sample size and inverse standard error weights. Nevertheless, the power difference among these three proposals of weights is small in case (ii).

Figure 5: Simulated Rejection Rates under Different Weights ($\gamma_1 = \gamma_2 = 0.5, \gamma_3 \in [0.35, 0.7], I_1 = 3, I_2 = 5, I_3 = 7, \alpha = 0.10$)



(a) $N_1 = 105, N_2 = 210, N_3 = 420$



(b) $N_1 = 420, N_2 = 210, N_3 = 105$

[19]The inverse (asymptotic) standard error weights borrow the idea of studentization of a test statistic whose asymptotic null distribution is normal.

B Proofs

B.1 Proof of Lemma 1

Proof. For any $\beta \in [0,1]$, any $c \in \mathbb{R}$, and any integrable functions $f(\cdot)$ and $h(\cdot)$ defined on $[0,1]$, by definition, we have

$$
\begin{aligned}
T_k\left(c \cdot f + h\right)(\beta) &= \frac{I_k - 2}{I_k - 1} \int_0^{\beta} \left[c \cdot f(\alpha) + h(\alpha)\right] d\alpha + \frac{1}{I_k - 1} \cdot \left[c \cdot f(\beta) + h(\beta)\right] \beta \\
&= c \cdot \left[\frac{I_k - 2}{I_k - 1} \int_0^{\beta} f(\alpha) d\alpha + \frac{1}{I_k - 1} \cdot f(\beta)\beta\right] + \frac{I_k - 2}{I_k - 1} \int_0^{\beta} h(\alpha) d\alpha + \frac{1}{I_k - 1} \cdot h(\beta)\beta \\
&= c \cdot T_k(f)(\beta) + T_k(h)(\beta),
\end{aligned}
$$

which says that T_k is linear. □

B.2 Proof of Lemma 2

Proof. We first verify that the conditions in part (ii) of Lemma 3.9.23 in van der Vaart and Wellner (1996) (or part (ii) of Lemma 12.8 in Kosorok (2008)) hold in auction setup under our assumptions. In auction setup, part (ii) of Lemma 3.9.23 in van der Vaart and Wellner (1996) requires that the bid distribution $G(\cdot|I)$ must have compact support and be continuously differentiable on its support with strictly positive derivative. Under our Assumption 2, the bid distribution function $G(\cdot|I)$ has a compact support $[\underline{b}, \overline{b}]$ and is continuously differentiable with a density $g(\cdot|I)$ on its support. Furthermore, the bid density (i.e., the derivative) $g(b|I) \geq c_g > 0$ for any $b \in [\underline{b}, \overline{b}]$. Consequently, the conditions in part (ii) of Lemma 3.9.23 of van der Vaart and Wellner (1996) hold in auctions under our Assumption 2.

We then apply part (ii) of Lemma 3.9.23 of van der Vaart and Wellner (1996) to our mapping from the bid distribution function to its quantile function, and can obtain the following results: (i) the inverse map $F \to F^{-1}$ as a map $\mathbb{D} \subset D[\underline{b}, \overline{b}] \to \ell^{\infty}(0,1)$ is Hadamard-differentiable at $G(\cdot|I)$ tangentially to $C[\underline{b}, \overline{b}]$; (ii) the derivative is the map $\alpha \to -(\alpha/g(\cdot|I)) \circ G^{-1}(\cdot|I)$. The desired conclusion therefore follows. □

B.3 Proof of Theorem 1

Proof. Under the null hypothesis H_0, we have $V_1(\beta) = V_2(\beta)$ for any $\beta \in [0,1]$. Consequently, we can rewrite the test statistic as

$$
\begin{aligned}
t &= \int_0^1 \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \left(\widehat{V_1}(\beta) - \widehat{V_2}(\beta)\right)^2 d\beta \\
&= \int_0^1 \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \left(\widehat{V_1}(\beta) - V_1(\beta) - \left(\widehat{V_2}(\beta) - V_2(\beta)\right)\right)^2 d\beta \\
&= \int_0^1 \left(\frac{I_1 - 2}{I_1 - 1} \int_0^{\beta} \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \left(\widehat{b}(\alpha|I_1) - b(\alpha|I_1)\right) d\alpha + \frac{1}{I_1 - 1} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \left(\widehat{b}(\beta|I_1) - b(\beta|I_1)\right) \beta\right. \\
&\qquad \left. - \frac{I_2 - 2}{I_2 - 1} \int_0^{\beta} \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \left(\widehat{b}(\alpha|I_2) - b(\alpha|I_2)\right) d\alpha - \frac{1}{I_2 - 1} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \left(\widehat{b}(\beta|I_2) - b(\beta|I_2)\right) \beta\right)^2 d\beta,
\end{aligned}
$$

where the second equality holds due to $V_1(\beta) = V_2(\beta)$ for any $\beta \in [0,1]$ under the null H_0, and the last equality comes from the definitions of $\widehat{V}_k(\cdot)$ and $V_k(\cdot)$ for $k = 1, 2$.

The test statistic can then be rewritten as

$$t = \int_0^1 \left(\sqrt{\frac{N_2}{N_1 + N_2}} \cdot T_1 \left(\widehat{\mathcal{G}_1} \right) (\beta) - \sqrt{\frac{N_1}{N_1 + N_2}} \cdot T_2 \left(\widehat{\mathcal{G}_2} \right) (\beta) \right)^2 d\beta,$$

where $\widehat{\mathcal{G}_k}(\cdot) \equiv \sqrt{N_k} \left(\widehat{b}(\cdot \mid I_k) - b(\cdot \mid I_k) \right)$, $k = 1, 2$, are the empirical quantile processes. Notice that, by Corollary 1, the empirical quantile processes $\widehat{\mathcal{G}_k}(\cdot) \rightsquigarrow \mathbb{B}(\cdot)/g(b(\cdot \mid I_k) \mid I_k)$ on $(0, 1)$ as $N_k \to \infty$. In addition, both mappings T_1 and T_2 are linear by Lemma 1, and the empirical quantile processes $\widehat{\mathcal{G}_1}(\cdot)$ and $\widehat{\mathcal{G}_2}(\cdot)$ are independent of each other because the bids under $I = I_1$ are independent of bids under $I = I_2$. Consequently,

$$\sqrt{\frac{N_2}{N_1 + N_2}} \cdot T_1 \left(\widehat{\mathcal{G}_1} \right) (\cdot) - \sqrt{\frac{N_1}{N_1 + N_2}} \cdot T_2 \left(\widehat{\mathcal{G}_2} \right) (\cdot) \rightsquigarrow \mathbb{G}(\cdot) \qquad \text{on } (0, 1),$$

where $\mathbb{G}(\cdot)$ is a Gaussian process with mean zero and covariance function of

$$\text{cov} \left(\mathbb{G}(t), \mathbb{G}(s) \right) = (1 - \lambda) \cdot \text{cov} \left(T_1 \left(\mathbb{G}_1 \right) (t), T_1 \left(\mathbb{G}_1 \right) (s) \right) + \lambda \cdot \text{cov} \left(T_2 \left(\mathbb{G}_2 \right) (t), T_2 \left(\mathbb{G}_2 \right) (s) \right);$$

and $\mathbb{G}_k(\cdot) \equiv \frac{\mathbb{B}(\cdot)}{g(b(\cdot|I_k)|I_k)}$ for $k = 1, 2$. Notice that the covariance function of $T_k(\mathbb{G}_k)(\cdot)$ for $k = 1, 2$, is given by

$$\begin{aligned} \text{cov}_{T_k(\mathbb{G}_k)}(t, s) \quad = \quad & \left(\frac{I-2}{I-1} \right)^2 \int_0^s \int_0^t \text{cov}_{\mathbb{G}_k}(\alpha, \beta) d\alpha d\beta + \frac{I-2}{(I-1)^2} \cdot t \int_0^s \text{cov}_{\mathbb{G}_k}(\alpha, t) d\alpha \\ & + \frac{I-2}{(I-1)^2} \cdot s \int_0^t \text{cov}_{\mathbb{G}_k}(\alpha, s) d\alpha + \frac{1}{(I-1)^2} \cdot st \cdot \text{cov}_{\mathbb{G}_k}(s, t), \end{aligned}$$

where $\text{cov}_{T_k(\mathbb{G}_k)}$ and $\text{cov}_{\mathbb{G}_k}$ denote the covariance functions of $T_k(\mathbb{G}_k)$ and $\mathbb{G}_k$, respectively. As a matter of fact, the covariance function of $T_k(\mathbb{G}_k)$ and hence $\mathbb{G}$ is well defined even when s, t approach 0 or 1, since the covariance function of $\mathbb{G}_k$ is well defined as s, t approach 0 or 1 under Assumption 2.

By continuous mapping theorem, our test statistic $t \xrightarrow{d} \int_0^1 \mathbb{G}(\beta)^2 d\beta$. □

B.4 Proof of Theorem 2

Proof. Lemma 1 shows that the mapping T_k is linear for $k = 1, 2$. In addition, by Lemma 2, the bid quantile function is a Hadamard differentiable functional of the bid CDF over the index set $(0, 1)$. Consequently, $V_k(\cdot) = T_k(b(\cdot|I_k))(\cdot)$ is a Hadamard differentiable functional of the bid CDF over the index set $(0, 1)$. By functional delta method, it suffices to show that the bootstrap applied to the empirical distributions yields processes with the same asymptotic covariance properties as those for the empirical distributions of the original sample.

Notice that the bootstrap empirical processes are (for details, see Chapter 3.6 of van der Vaart and Wellner (1996)),

$$\mathbf{G}_1^*(b_1) = \frac{1}{\sqrt{N_1}} \sum_{j=1}^{N_1} \mathbf{1}\{B_{1,j}^m \le b_1\} - \widehat{G_1}(b_1) = \frac{1}{\sqrt{N_1}} \sum_{j=1}^{N_1} \left(M_{1j} - 1 \right) \cdot \mathbf{1}\{B_{1,j} \le b_1\},$$

$$\mathbf{G}_2^*(b_2) = \frac{1}{\sqrt{N_2}} \sum_{j=1}^{N_2} \mathbf{1}\{B_{2,j}^m \le b_2\} - \widehat{G_2}(b_2) = \frac{1}{\sqrt{N_2}} \sum_{j=1}^{N_2} \left(M_{2j} - 1 \right) \cdot \mathbf{1}\{B_{2,j} \le b_2\},$$

where M_{1j} and M_{2j} are independent multinomial random variables with N_1 and N_2 cells and success probabilities of $1/N_1$ and $1/N_2$, respectively. Notice that M_{1j} and M_{2j} are also independent of the original sample. It is easy to show that, given the original samples $\mathcal{D}_1$ and $\mathcal{D}_2$, $\mathbf{G}_1^*(\cdot)$ and $\mathbf{G}_2^*(\cdot)$ are independent mean zero processes with covariance kernels of:

$$E\left(\mathbf{G}_k^*(b_k)\mathbf{G}_k^*(b_k')|B_{k,1},\ldots,B_{k,N_k}\right) = \widehat{G_k}(b_k) - \widehat{G_k}(b_k)\widehat{G_k}(b_k'),$$

for $b_k \leq b_k'$ and $k = 1,2$. Such covariance kernels converge to the ones of the limiting processes of the corresponding empirical processes. Consequently, the bootstrap empirical processes $\mathbf{G}_1^*(\cdot)$ and $\mathbf{G}_2^*(\cdot)$ have the same limiting processes as the empirical processes based on the corresponding empirical distributions of the original samples $\mathcal{D}_1$ and $\mathcal{D}_2$.

Following the delta method for bootstrap as shown by Theorem 3.9.11 of van der Vaart and Wellner (1996) and the continuous mapping theorem, we can show that, given the original samples $\mathcal{D}_1$ and $\mathcal{D}_2$, the bootstrap test statistic

$$t^m \xrightarrow{d} \int_0^1 \mathbb{G}(\beta)^2 d\beta$$

as $N_1, N_2 \to \infty$. Consequently, given $\mathcal{D}_1$ and $\mathcal{D}_2$, the asymptotic distribution of bootstrap test statistic is the same as the asymptotic null distribution of the original test statistic.

We now show part 1. Denote n' as the minimum of N_1 and N_2. Under the null hypothesis H_0,

$$\begin{aligned}
& \lim_{n',M\to\infty} \Pr\left(t > c_{1-\alpha}^M\right) \\
= \; & 1 - \lim_{n',M\to\infty}\left[\Pr(t \leq c_{1-\alpha}^M) - \Pr(t_\infty \leq c_{1-\alpha}^M)\right] - \lim_{n',M\to\infty}\Pr(t_\infty \leq c_{1-\alpha}^M) \\
= \; & 1 - \lim_{n',M\to\infty}\Pr(t_\infty \leq c_{1-\alpha}^M) \\
= \; & \alpha,
\end{aligned}$$

where t_∞ is distributed as $\int_0^1 \mathbb{G}(\beta)^2 d\beta$, namely the asymptotic null distribution of t; the second-to-last equality holds by Polya's theorem since the asymptotic null distribution of t (i.e. $\int_0^1 \mathbb{G}(\beta)^2 d\beta$) is continuous; and the last equality holds due to the fact that, given the original samples $\mathcal{D}_1$ and $\mathcal{D}_2$, $\lim_{n',M\to\infty} c_{1-\alpha}^M = c_{1-\alpha}^\infty$ where $c_{1-\alpha}^\infty$ is the $(1-\alpha)$-quantile of distribution $\int_0^1 \mathbb{G}(\beta)^2 d\beta$. [20] The conclusion of part 1 therefore follows.

We then show part 2. Under the alternative hypothesis H_1, there exists a β in $[0,1]$, denoted as β^*, such that $V_1(\beta^*) \neq V_2(\beta^*)$. Consequently, there must exist an interval $\mathcal{C}^*$ with positive measure around β^* such that $V_1(\beta) \neq V_2(\beta)$ for $\beta \in \mathcal{C}^*$, since V_1 and V_2 are

[20] Notice that $c_{1-\alpha}^M$ is essentially the $(1-\alpha)$-th quantile of the i.i.d. bootstrap test statistic sample $\{t^1,\ldots,t^M\}$ where $t^m \xrightarrow{d} t_\infty$ as $n' \to \infty$, $m = 1,\ldots,M$. Consequently, as $n', M \to \infty$, the empirical quantile value $c_{1-\alpha}^M$ converges to the $(1-\alpha)$-quantile of t_∞.

continuous by Assumption 2. We therefore have

$$
\begin{aligned}
t &= \int_0^1 \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \left(\widehat{V_1}(\beta) - \widehat{V_2}(\beta)\right)^2 d\beta \\
&\geq \int_{\mathcal{C}^*} \frac{N_1 \cdot N_2}{N_1 + N_2} \cdot \left(\widehat{V_1}(\beta) - \widehat{V_2}(\beta)\right)^2 d\beta \\
&= \int_{\mathcal{C}^*} \left(\sqrt{\frac{N_1 N_2}{N_1 + N_2}}[\widehat{V_1}(\beta) - V_1(\beta)] - \sqrt{\frac{N_1 N_2}{N_1 + N_2}}[\widehat{V_2}(\beta) - V_2(\beta)] + \sqrt{\frac{N_1 N_2}{N_1 + N_2}}[V_1(\beta) - V_2(\beta)]\right)^2 d\beta
\end{aligned} \tag{14}
$$

where the first two terms on the right-hand side of last equality in Equation (14) are stochastically bounded, but its third term diverges to infinity as $n' \to \infty$ since $V_1(\beta) \neq V_2(\beta)$ for any β in the positively measured interval $\mathcal{C}^*$. Thus, the right-hand side of last equality in Equation (14) diverges to $+\infty$ as n' goes to infinity, which implies that the statistic t goes to $+\infty$ as $n' \to \infty$. On the other hand, as $n', M \to \infty$, $c_{1-\alpha}^M$ converges to the $(1-\alpha)$-quantile of $\int_0^1 \mathbb{G}(\beta)^2 d\beta$ which is finite. We therefore have that, for any $\alpha \in (0, 1)$,

$$
\lim_{n', M \to \infty} \Pr(t > c_{1-\alpha}^M) = 1
$$

under the alternative hypothesis H_1. Part 2 follows immediately. □

B.5 Proof of Theorem 3

Proof. Notice that there must exist an interval $\mathcal{C}^*$ with positive measure such that $h(\beta) \neq 0$ for any $\beta \in \mathcal{C}^*$, because that $h(\cdot)$ is nonzero and is differentiable on $[0, 1]$. Under the local alternative hypothesis H_{1n}, we can rewrite the test statistic t as follows:

$$
\begin{aligned}
t &= \int_0^1 n \cdot \left(\widehat{V_1}(\beta) - \widehat{V_2}(\beta)\right)^2 d\beta \\
&= \int_0^1 \left(\sqrt{n}\left(\widehat{V_1}(\beta) - V_1(\beta)\right) - \sqrt{n}\left(\widehat{V_2}(\beta) - V_2(\beta)\right) - n^{\frac{1}{2}-\gamma} \cdot h(\beta)\right)^2 d\beta
\end{aligned}
$$

As shown by the proof of Theorem 1 in Appendix B.3,

$$
\sqrt{n}\left(\widehat{V_1}(\cdot) - V_1(\cdot)\right) - \sqrt{n}\left(\widehat{V_2}(\cdot) - V_2(\cdot)\right) \rightsquigarrow \mathbb{G}(\cdot) \qquad \text{on } (0, 1).
$$

Consequently, if $\gamma < \frac{1}{2}$, then the test statistic $t \xrightarrow{p} +\infty$, since $n^{\frac{1}{2}-\gamma} \cdot h(\beta) \longrightarrow \infty$ for all β in the positively measured interval $\mathcal{C}^*$; and by the continuous mapping theorem, if $\gamma = \frac{1}{2}$, then $t \xrightarrow{d} \int_0^1 (\mathbb{G}(\beta) - h(\beta))^2 d\beta$, due to the fact that $n^{\frac{1}{2}-\gamma} \cdot h(\beta) = h(\beta)$ for all $\beta \in [0, 1]$; if $\gamma > \frac{1}{2}$, then $t \xrightarrow{d} \int_0^1 \mathbb{G}(\beta)^2 d\beta$ for that $n^{\frac{1}{2}-\gamma} \cdot h(\cdot) \to 0$ uniformly on $[0, 1]$. The desired result therefore follows. □

B.6 Proof of Lemma 3

Proof. For $k = 1, \ldots, K$, let $\widehat{\mathcal{G}_k}(\cdot) \equiv \sqrt{N_k} \cdot (\hat{b}(\cdot|I_k) - b(\cdot|I_k))$ be the empirical quantile process of sample k. Under H_0^K, we can rewrite the test statistic t_K as follows:

$$
\begin{aligned}
t_K &= \sum_{i=1}^{K}\sum_{j=i+1}^{K} w_{ij} \cdot \frac{N_i \cdot N_j}{N_i + N_j} \int_0^1 \left(\hat{V}(\beta|I_i) - \hat{V}(\beta|I_j)\right)^2 d\beta \\
&= \sum_{i=1}^{K}\sum_{j=i+1}^{K} w_{ij} \int_0^1 \left[\sqrt{\frac{N_j}{N_i+N_j}} \cdot \sqrt{N_i}\left(\hat{V}(\beta|I_i) - V(\beta|I_i)\right) - \sqrt{\frac{N_i}{N_i+N_j}} \cdot \sqrt{N_j}\left(\hat{V}(\beta|I_j) - V(\beta|I_j)\right)\right]^2 d\beta \\
&= \sum_{i=1}^{K}\sum_{j=i+1}^{K} w_{ij} \int_0^1 \left[\sqrt{\frac{N_j}{N_i+N_j}} \cdot T_i\left(\widehat{\mathcal{G}}_i\right)(\beta) - \sqrt{\frac{N_i}{N_i+N_j}} \cdot T_j\left(\widehat{\mathcal{G}}_j\right)(\beta)\right]^2 d\beta,
\end{aligned}
$$

where the second equality holds because $V(\cdot|I_i) = V(\cdot|I_j)$ for any $i, j = 1, \ldots, K$ under H_0^K, and the last equality obtains by the definitions of mapping T_k and empirical quantile process $\widehat{\mathcal{G}_k}$ for $k = 1, \ldots, K$.

By Corollary 1, it is easy to show that $\widehat{\mathcal{G}_k}(\cdot) \rightsquigarrow \mathbb{G}_k(\cdot) \equiv \mathbb{B}(\cdot)/g(b(\cdot|I_k)|I_k)$ on $(0,1)$ as $N_k \to \infty$ for $k = 1, \ldots, K$. Moreover, $\mathbb{G}_i(\cdot)$ is independent of $\mathbb{G}_j(\cdot)$ for any $i \neq j \in \{1, \ldots, K\}$, since $\widehat{\mathcal{G}}_i(\cdot)$ is independent of $\widehat{\mathcal{G}}_j(\cdot)$ by the independence between samples i and j. Therefore, $(T_1(\widehat{\mathcal{G}_1})(\cdot), \ldots, T_K(\widehat{\mathcal{G}_K})(\cdot)) \rightsquigarrow (T_1(\mathbb{G}_1)(\cdot), \ldots, T_K(\mathbb{G}_K)(\cdot))$ on $(0,1)$ as $\min\{N_1, \ldots, N_K\} \to \infty$ by linearity of T_k, $k = 1, \ldots, K$. By continuous mapping theorem, we conclude that

$$
\begin{aligned}
t_K &= \sum_{i=1}^{K}\sum_{j=i+1}^{K} w_{ij} \int_0^1 \left[\sqrt{\frac{N_j}{N_i+N_j}} \cdot T_i\left(\widehat{\mathcal{G}}_i\right)(\beta) - \sqrt{\frac{N_i}{N_i+N_j}} \cdot T_j\left(\widehat{\mathcal{G}}_j\right)(\beta)\right]^2 d\beta \\
&\xrightarrow{d} \sum_{i=1}^{K}\sum_{j=i+1}^{K} w_{ij} \int_0^1 \left[\sqrt{1-\lambda_{ij}} \cdot T_i\left(\mathbb{G}_i\right)(\beta) - \sqrt{\lambda_{ij}} \cdot T_j\left(\mathbb{G}_j\right)(\beta)\right]^2 d\beta,
\end{aligned}
$$

where $(\mathbb{G}_1(\cdot), \ldots, \mathbb{G}_K(\cdot))$ is a K-variate Gaussian process with independent components $\mathbb{G}_k(\cdot) = \mathbb{B}(\cdot)/g(b(\cdot|I_k)|I_k)$, $k = 1, \ldots K$. This completes the proof. □

References

AN, Y., Y. HU, AND M. SHUM (2010): "Estimating first-price auctions with an unknown number of bidders: A misclassification approach," *Journal of Econometrics*, 157(2), 328–341.

ARADILLAS-LÓPEZ, A., A. GANDHI, AND D. QUINT (2013): "Identification and inference in ascending auctions with correlated private values," *Econometrica*, 81(2), 489–534.

ARYAL, G., AND M. F. GABRIELLI (2012): "Testing for collusion in asymmetric first-price auctions," *International Journal of Industrial Organization*.

ARYAL, G., S. GRUNDL, D.-H. KIM, AND Y. ZHU (2015): "Empirical Relevance of Ambiguity in First Price Auctions," *Working Paper*.

ATHEY, S., AND P. A. HAILE (2002): "Identification of standard auction models," *Econometrica*, 70(6), 2107–2140.

ATHEY, S., AND J. LEVIN (2001): "Information and Competition in US Forest Service Timber Auctions," *Journal of Political Economy*, 109(2), 375–417.

ATHEY, S., J. LEVIN, AND E. SEIRA (2011): "Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions," *The Quarterly Journal of Economics*, 126(1), 207–257.

BAJARI, P., AND A. HORTAÇSU (2005): "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data," *Journal of Political Economy*, 113(4), 703–741.

BALDWIN, L. H., R. C. MARSHALL, AND J.-F. RICHARD (1997): "Bidder collusion at forest service timber sales," *Journal of Political Economy*, 105(4), 657–699.

BARMI, H. E., AND H. MUKERJEE (2005): "Inferences under a Stochastic Ordering Constraint: The k-Sample Case," *Journal of the American Statistical Association*, pp. 252–261.

BARRETT, G. F., S. G. DONALD, AND D. BHATTACHARYA (2014): "Consistent nonparametric tests for Lorenz dominance," *Journal of Business & Economic Statistics*, 32(1), 1–13.

CAMPO, S. (2012): "Risk aversion and asymmetry in procurement auctions: Identification, estimation and application to construction procurements," *Journal of Econometrics*, 168(1), 96–107.

CAMPO, S., E. GUERRE, I. PERRIGNE, AND Q. VUONG (2011): "Semiparametric estimation of first-price auctions with risk-averse bidders," *The Review of Economic Studies*, 78(1), 112–147.

COX, J. C., V. L. SMITH, AND J. M. WALKER (1988): "Theory and individual behavior of first-price auctions," *Journal of Risk and Uncertainty*, 1(1), 61–99.

FANG, H., AND X. TANG (2014): "Inference of bidders' risk attitudes in ascending auctions with endogenous entry," *Journal of Econometrics*, 180(2), 198–216.

FLAMBARD, V., AND I. PERRIGNE (2006): "Asymmetry in procurement auctions: evidence from snow removal contracts," *The Economic Journal*, 116(514), 1014–1036.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-price Auctions," *Econometrica*, 68(3), 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77(4), 1193–1227.

HAILE, P. A. (2001): "Auctions with resale markets: an application to US forest service timber sales," *American Economic Review*, pp. 399–427.

HAILE, P. A., H. HONG, AND M. SHUM (2003): "Nonparametric tests for common values at first-price sealed-bid auctions," Discussion paper, National Bureau of Economic Research.

HAILE, P. A., AND E. TAMER (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111(1), 1–51.

HICKMAN, B. R., AND T. P. HUBBARD (2014): "Replacing Sample Trimming with Boundary Correction in Nonparametric Estimation of First-Price Auctions," *Journal of Applied Econometrics*.

HU, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144(1), 27–61.

HU, Y., D. MCADAMS, AND M. SHUM (2013): "Identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, 174(2), 186–193.

JUN, S. J., J. PINKSE, AND Y. WAN (2010): "A consistent nonparametric test of affiliation in auction models," *Journal of Econometrics*, 159(1), 46 – 54.

KOSOROK, M. R. (2008): *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media.

KRASNOKUTSKAYA, E. (2011): "Identification and estimation of auction models with unobserved heterogeneity," *The Review of Economic Studies*, 78(1), 293–327.

LEVIN, D., AND J. L. SMITH (1994): "Equilibrium in auctions with entry," *The American Economic Review*, pp. 585–599.

LI, H., AND N. LIU (2015): "Nonparametric Identification and Estimation of Double Auctions with Bargaining," Working Paper.

LI, T., AND Q. VUONG (1998): "Nonparametric estimation of the measurement error model using multiple indicators," *Journal of Multivariate Analysis*, 65(2), 139–165.

LI, T., AND B. ZHANG (2010): "Testing for affiliation in first-price auctions using entry behavior," *International Economic Review*, 51(3), 837–850.

LIU, N., AND Q. VUONG (2013): "Nonparametric Test of Monotonicity of Bidding Strategy in First-price Auctions," Working Paper, New York University.

LU, J., AND I. PERRIGNE (2008): "Estimating risk aversion from ascending and sealed-bid auctions: the case of timber auction data," *Journal of Applied Econometrics*, 23(7), 871–896.

LUO, Y., I. PERRIGNE, AND Q. VUONG (2014): "Structural Analysis of Nonlinear Pricing," Working Paper, New York University.

MARMER, V., AND A. SHNEYEROV (2012): "Quantile-based nonparametric inference for first-price auctions," *Journal of Econometrics*, 167(2), 345–357.

MARMER, V., A. SHNEYEROV, AND P. XU (2013): "What model for entry in first-price auctions? A nonparametric approach," *Journal of Econometrics*, 176(1), 46–58.

RILEY, J. G., AND W. F. SAMUELSON (1981): "Optimal auctions," *The American Economic Review*, pp. 381–392.

SAMUELSON, W. F. (1985): "Competitive bidding with entry costs," *Economics Letters*, 17(1), 53–57.

SHNEYEROV, A., AND A. C. L. WONG (2011): "Identification in first-price and dutch auctions when the number of potential bidders is unobservable," *Games and Economic Behavior*, 72(2), 574–582.

VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.