Online Appendix to "Identification of First-Price Auctions with Non-Equilibrium Beliefs: A Measurement Error Approach"

Yonghong An[*]
Texas A&M University

July 11, 2017

Abstract

This online Appendix contains a comparison between our main paper and Gillen (2009) and asymptotic properties of estimators omitted from the main paper.

1 A comparison with Gillen (2009)

We provide some details on the identification results in Gillen (2009), and show that Gillen's methodology is not applicable to the level-k auction models we considered where the number of types, the specification of all possible types and type-$L0$ are unknown to econometricians.

We first summarize Gillen's main results on identification using the notation in our main paper. The objectives of identification in Gillen (2009) are value distribution $F(\cdot)$ and type distribution $p(\cdot)$ under the assumption that the econometrician observes the specification of type-$L0$ and all the other possible types, and the number of types m. The author employs a two-step procedure to achieve identification. In the first step, it is shown that if all the bidders are homogenous (of the same type) and the type is known, then value distribution of bidders can be identified from the observed bid distribution (Section 4). The identification of this step depends crucially on the assumption that bidding strategy $s_k(\cdot)$ is known to the econometrician. Consequently $F(s_k^{-1}(b)) = G(b|\tau = Lk)$ is used to identify $F(\cdot)$ from $G(\cdot|\tau = Lk)$. As the author

[*]Department of Economics, Texas A&M University, 3051 Allen Building, College Station , TX 77843. Email: y.an@tamu.edu.

admitted, the setting of homogenous bidders is hardly true even in laboratories (p.14), but the identification results in the first step can be employed in the second-step identification where bidders are assumed to be heterogenous (of different types).

In the second-step of identification, the author considers a population of heterogenous bidders, and each bidder is of type Lk with probability $p(\tau = Lk)$, where Lk is known and $p(\tau = Lk)$ is the component of the identification objective $p(\cdot) \equiv (p(\tau = L1), \cdots, p(\tau = Lk))$. For the purpose of identification, the author considers three scenarios with different observed information by the econometrician. First, when bid distribution for each individual in the heterogenous population is observed then the value distribution is identified using the relationship $F(s_k^{-1}(b)) = G(b|\tau = Lk)$. This identification result requires a large number of observations for each individual, which is rarely observed. Second, if the econometrician observes the bid distribution for a fixed population of bidders $G(\cdot)$, then only *partial* identification can be achieved, i.e., for any given distribution over types $p(\cdot)$, there exists a value distribution $F(\cdot)$ such that $G(b) = \sum_\tau p(\tau = Lk)G(b|\tau = Lk)$ holds. This partial identification result is due to (6) in the present paper $G(b) = \sum_{\tau \in \mathcal{K}} p(\tau)G(b|\tau) = \sum_{\tau \in \mathcal{K}} p(\tau)F(s_\tau^{-1}(b))$: if $p(\cdot)$ and $s_\tau^{-1}(\cdot)$ are known, then $F(\cdot)$ is identified based on the results of the first-step in Gillen (2009). Third, when the econometrician observes variation of number of bidders and the variation does not change the distribution of types and values, the distribution of values $F(\cdot)$ and types $p(\cdot)$ are identified. This result holds because an identification set is obtained for N heterogenous bidders, and when N varies, the intersection of identification sets for different N may lead to point identification.

In summary, the identification results in Gillen (2009) rely on the assumption that the econometrician observes the number of types m, the set of all the types $\mathcal{K}$, the specification of type-$L0$, and the variation of number of bidders across auctions. Hence Gillen's methodology cannot be used in the setting of the present paper where the components listed above are unknown and need to be identified from the data. According to the comparison above, this present paper identifies more components under less restrictive assumptions on the model. For example, the identification of type-$L0$ achieved in this paper but not in Gillen (2009) is crucial in the literature for us to understand bidders' strategic behavior. Of course, the advantages of the present paper on identification comes at the price of requiring a richer data structure, i.e., known identity of bidders and three bids for each bidder. Nevertheless, the essential difference between the two papers is not the data structure but the econometric methodology. In empirical application, Gillen (2009) also employs USFS timber data. However, the focus is to study optimal reserve price under the assumption that bidders are all of the same type (homogenous). The present paper analyzes the bidders' behavior and tests whether they behave according to level-k auction models.

All in all, Gillen (2009) and the present paper have main differences in both methodology and results even though both papers focus on level-k auction models.

2 Asymptotic Properties of Estimators

This section summarizes properties of the proposed estimators in the main paper, and the details of the proofs. We show the uniform consistency of estimators of bid distribution $\widehat{G}(b|\tau)$, bid density $\widehat{g}(b|\tau)$, the pseduovalue estimator $\widehat{\xi}_{Lk}(\cdot)$, and the estimator for value distribution $\widehat{F}(\cdot)$, and as well as the consistency of the boundary estimators $\widehat{\underline{b}}$ and $\widehat{\overline{b}}_{Lk}$. We also prove that the estimators of bid distribution $\widehat{G}(b|\tau)$, and bid density $\widehat{g}(b|\tau)$ are asymptotically normal.

The uniform consistency of estimators $\widehat{G}(b|\tau)$ holds without additional assumptions, and the proof relies on the result that the estimator of eigenvector matrix $\widehat{B}_{d_1|\tau}$ is uniformly consistent. We summarize the uniform consistency of $\widehat{G}(b|\tau)$ as follows.

Proposition 1. *For any realization of type $\tau \in \mathcal{K}$, $\widehat{G}(b|\tau)$ is uniformly consistent, i.e., $\sup_b \big|\widehat{G}(b|\tau) - G(b|\tau)\big| = O_p(N^{-1/2})$.*

Proof To prove the uniform consistency of estimators $\widehat{G}(b|\tau)$ and $\widehat{g}(b|\tau)$, we need following preliminary results.

Lemma 1. *For any realization of type $\tau \in \mathcal{K}$, the estimates $\widehat{B}_{d_1|\tau}$ and $\widehat{p}(\tau)$ are both uniformly consistent. Specifically $\widehat{B}_{d_1|\tau} - B_{d_1|\tau} = O_p(N^{-1/2})$, and $\widehat{p}(\tau) - p(\tau) = O_p(N^{-1/2})$.*

Proof The matrix $B_{d_1|\tau}$ is estimated as

$$\widehat{B}_{d_1|\tau} := \phi\left(\widehat{B}_{Eb_2,d_1,d_3}\widehat{B}^{-1}_{d_1,d_3}\right), \tag{1}$$

where $\phi(\cdot)$ is an analytical function as described in Andrew et al. (1993). The continuity of $\phi(\cdot)$ implies that it is sufficient to prove the uniform consistency of $\widehat{B}_{Eb_2,d_1,d_3}$ and $\widehat{B}^{-1}_{d_1,d_3}$. It is straightforward to show $\widehat{B}_{d_1,d_3} - B_{d_1,d_3} = O_p(N^{-1/2})$. Similarly, we can also obtain that $\widehat{B}_{Eb_2,d_1,d_3} - B_{Eb_2,d_1,d_3} = O_p(N^{-1/2})$. Consequently, we conclude that $\widehat{B}_{d_1|\tau} - B_{d_1|\tau} = O_p(N^{-1/2})$ holds.

The type distribution $\widehat{p}(\tau)$ is estimated by $\widehat{p}(\tau) = \widehat{B}^{-1}_{d_1|\tau}\widehat{p}(d_1)$. Then the uniform consistency of $\widehat{p}(\tau)$ immediately follows the consistency of $\widehat{B}^{-1}_{d_1|\tau}$. ■

To prove the uniform consistency of $\widehat{G}_{b|\tau}$, first consider that $\widehat{p}(\tau) = O_p(1)$ and $\sup_{b_2} |\widehat{G}(b_2, d_1)| < \infty$, then the estimate of $\widehat{G}_{b|\tau}$ can be rewritten as

$$\widehat{G}(b_2|Lk) = \frac{e_k^T B^{-1}_{d_1|\tau}\widehat{G}(b_2, d_1)}{e_k^T B^{-1}_{d_1|\tau} p(d_1)} + O_p(N^{-1/2}). \tag{2}$$

According to the Dvoretzky-Kiefer-Wolfowitz inequality, we have

$$\Pr\left(\sup_{b_2}\left|\widehat{G}(b_2,d_1)-G(b_2,d_1)\right|>\epsilon\right)\le 2e^{-2N\epsilon^2},\ \text{ for every } \epsilon>0.$$

If one chooses $\epsilon=\frac{C}{\sqrt{N}}, C>0$, an immediate result is

$$\sup_{b_2}\left|\widehat{G}(b_2,d_1)-G(b_2,d_1)\right|=O_p(N^{-1/2}).$$

Consequently, it is straightforward to show that $\sup_{b_2}\left|\widehat{G}(b_2|\tau)-G(b_2|\tau)\right|=O_p(N^{-1/2})$. ■

To prove the uniform consistency of density estimator $g(b|\tau)$, we define the norm $\|\cdot\|_\infty$ as $\|\widehat{g}-g\|_\infty=\sup_b|\widehat{g}-g|$, and impose the following assumptions on the bid density $g(b|\tau)$ and the kernel function $K(\cdot)$:

Assumption 1. *For any realization of type $\tau\in\mathcal{K}$, $g(\cdot,\tau)$ has a second derivative which is bounded and continuous at an interior point of the support of $g(\cdot,\tau)$ and $\int|g(b,\tau)|db<+\infty$.*

Assumption 2. *The kernel function $K(\cdot)$ is a symmetric function around zero satisfying:*

(i) $\int_{-\infty}^{+\infty}K(u)du=1$. *(ii)* $|uK(u)|\to 0$ *as* $|u|\to+\infty$, *and* $\sup|K(u)|<+\infty$. *(iii)* $\int_{-\infty}^{+\infty}|K(u)|du<+\infty$, *and* $\int_{-\infty}^{+\infty}K^2(u)du<+\infty$. *(iv)* $\int_{-\infty}^{+\infty}|u^2K(u)|du<+\infty$.

Assumption 3. *There exists some $\delta>0$ such that $\int K^{2+\delta}(u)du<+\infty$.*

Given the assumptions above, we have the following result regarding $\widehat{g}(b|\tau)$.

Proposition 2. *Suppose Assumptions $1-3$ hold, then for all realizations of $\tau\in\mathcal{K}$, $\|\widehat{g}(b|\tau)-g(b|\tau)\|_\infty=O_p\left(\frac{1}{\sqrt{Nh}}+h^2\right)$, where h the bandwidth in the kernel density estimator $\widehat{g}(b|\tau)$.*

Proof The proof of uniform consistency of kernel density $\widehat{g}(b|\tau)$ is similar to that of the distribution function $\widehat{G}(b|\tau)$ except that the kernel function $K(\cdot)$ and bandwidth h have to be taken into account. It is readily to derive the following relationship analogous with (2),

$$\widehat{g}(b_2|Lk)=\frac{e_k^TB_{d_1|\tau}^{-1}\widehat{g}(b_2,d_1)}{e_k^TB_{d_1|\tau}^{-1}p(d_1)}+O_p(N^{-1/2}). \tag{3}$$

According to the arguments in the proof of uniform consistency of $\widehat{G}(b|\tau)$, we only need to focus on the properties of the joint kernel density $\widehat{g}(b_2,d_1)$ in order to prove the uniform consistency of $\widehat{g}(b_2|\tau)$. For ease of notation, we suppress the arguments in $g(\cdot|\cdot)$ and $\widehat{g}(\cdot|\cdot)$, and all the asymptotic properties of $\hat{g}$ is conditional on τ. According to the triangular inequality

$$\|\widehat{g} - g\|_\infty \leq \|E\widehat{g} - g\|_\infty + \|\widehat{g} - E\widehat{g}\|_\infty \,, \tag{4}$$

it is sufficient to show the uniform consistency of the two terms on the right-hand side.

According to Lemma 5.1 in Fan and Yao (2005), under Assumption 1 and (i), (iv) of Assumption 2, the bias of the estimator $\widehat{g}$ satisfies $\|E\widehat{g} - g\|_\infty = O(h^2)$. If Assumptions 1, 3, and (ii), (iii) of Assumption 2 hold, then Theorem 2.8 and 2.9 in Pagan and Ullah (1999) shows that $\|\widehat{g} - E\widehat{g}\|_\infty = O_p\left(\frac{1}{\sqrt{Nh}}\right)$. Consequently, (4) implies that $\|\widehat{g} - g\|_\infty = O_p\left(\frac{1}{\sqrt{Nh}} + h^2\right)$ hold. This completes the proof of uniform consistency. If the density $g(\cdot, \tau)$ has a higher-order derivative (say qth, $q \geq 3$), then a bias $\|E\widehat{g} - g\|_\infty = O(h^q)$ can be obtained by further requiring $\int u^j K(u)du = 0$, and $\int |u^q K(u)|du < +\infty$, where $j = 1, ..., q-1$. The details can be found on page 205 in Fan and Yao (2005). ■

Next, we proceed to analyze the consistency of the boundary estimators. Among all the boundary estimators $\widehat{\underline{b}}$ and $\widehat{\overline{b}}_{Lk} (k = 1, 2, ..., m)$, $\widehat{\underline{b}}$ and $\widehat{\overline{b}}_{L1}$ are estimated by sample minimum and maximum, respectively, while $\widehat{\overline{b}}_{Lk} (k = 2, ..., m)$ is estimated from the bid distribution for type-Lk. Thus the first two estimators have similar properties, which are different from the remaining $m-1$ ones, $\widehat{\overline{b}}_{Lk}, k = 2, ..., m$. The asymptotic properties of $\widehat{\underline{b}}$ is summarized in the following proposition.

Proposition 3. *Suppose the value distribution $F(\cdot)$ has a finite support, and the bidding function $s_{Lk}(\cdot)$ has a bounded derivative, then the estimate $\widehat{\underline{b}}$ satisfies $(\widehat{\underline{b}} - \underline{b})/w(n) \xrightarrow{d} W$, where $w(n)$ is a sequence dependent on both $F(\cdot)$ and $s_{Lk}(\cdot)$; W is a nondegenerate random variable with Weibull distribution function $H(\cdot; 1)$.*

Proof To prove the proposition, we first consider the case where there is only one type of bidders, say type-Lk. Denote this type's bid distribution $G(b|Lk)$, and its support $[\underline{b}, \overline{b}_{Lk}] = [\underline{v}, s_{Lk}(\overline{v})]$. This support is finite because $s_{Lk}(\overline{v}) \leq \overline{v} < \infty$. Moreover, we have shown in the paper that the density of bid $g(b|Lk) > 0$ on $[\underline{b}, \overline{b}_{Lk}]$. Under these conditions we consider the limit

$$\lim_{\epsilon \to 0+} \frac{G\big(G^{-1}(0|Lk) + \epsilon b|Lk\big)}{G\big(G^{-1}(0|Lk) + \epsilon|Lk\big)} = \lim_{\epsilon \to 0+} \frac{G\big(\underline{b} + \epsilon b|Lk\big)}{G\big(\underline{b} + \epsilon|Lk\big)} = \lim_{\epsilon \to 0+} \frac{g\big(\underline{b} + \epsilon b|Lk\big)}{g\big(\underline{b} + \epsilon|Lk\big)} b = b, \tag{5}$$

where the last equality is due to the fact that $g(b|Lk)$ is bounded away from zero. According to Theorem 8.3.6 in Arnold et al. (2008), the finite support of $G(b|Lk)$ and the limit condition suffice the asymptotic distribution of $(\widehat{\underline{b}} - \underline{b})/w(n)$ being Weibull distribution function with shape parameter 1. In particular, $w(n)$ can be chosen as $G^{-1}(1/n|Lk) - \underline{b}$. Apparently, without further specification of $F(\cdot)$, it is not possible to get the rate of convergence of the boundary estimator $\widehat{\underline{b}}$.

Now we turn to the case of a mixture model $G(b) = \sum_{\tau \in \mathcal{K}} G(b|\tau)p(\tau)$. AL-Hussaini and El-Adll (2004) and Sreehari and Ravi (2010) show[1] that if $(\widehat{\underline{b}} - \underline{b})/w_\tau(n) \xrightarrow{d} W$ holds for each component $G(b|\tau)$ of the mixture model, then it also holds for $G(b)$. Of course, the sequence $w_\tau(n)$ may differ for different realizations of $\tau \in \mathcal{K}$. The argument above implies that for the mixture model, $(\widehat{\underline{b}} - \underline{b})/w(n) \xrightarrow{d} W$ holds, where $w(n)$ is a common positive sequence for all the possible types. The sequence $w(n)$ depends on both the components $G(b|\tau)$ and the distribution of types, $p(\tau)$, while $G(b|\tau)$ relying on value distribution $F(\cdot)$ and bidding strategy of bidders $s_\tau(\cdot)$. Without imposing further restrictions on the model, the rate of convergence $w(n)$ is unattainable. ■

The upper bound of bid distribution for type-1 bidders, $\bar{b}_{L1}$, is estimated as sample maximum $\widehat{\bar{b}}$. The properties of sample maximum is similar to that of sample minimum. Actually, a sample minimum from a cdf $Y(x)$ has the same distribution as the negative of the sample maximum from the cdf $\widetilde{Y}(x)$, where $\widetilde{Y}(x) = 1 - Y(-x)$. Therefore, the asymptotic properties of $\widehat{\bar{b}}$ and the proof is similar to that of $\widehat{\underline{b}}$ and they are not reproduced here in its entirety. We show in the following that all of the other $m-1$ estimates $\widehat{\bar{b}}_{Lk}$ converge in probability to the upper bound $\bar{b}_{Lk}$.

Proposition 4. *$\widehat{\bar{b}}_{Lk} \xrightarrow{p} \bar{b}_{Lk}$, $Lk \in \mathcal{K}$, $k > 1$.*

Proof For ease of notation, the subscript Lk is dropped and the estimator is denoted ,$\widehat{\bar{b}}_n$ where n is the sample size. We prove $p\lim_{n\to\infty} \widehat{\bar{b}}_n = \bar{b}$ by seeking a contradiction. Suppose $p\lim_{n\to\infty} \widehat{\bar{b}}_n = \bar{b}$ does not hold, then there exists a subsequence of estimators $\widehat{\bar{b}}_{n_k}$ such that for any $\epsilon > 0$, $\lim_{n_k\to\infty} P\big(|\widehat{\bar{b}}_{n_k} - \bar{b}| > \epsilon\big) > 0$, which implies that either one of the following cases or both of them are true:

(i) $\lim_{n_k\to\infty} P\big(\widehat{\bar{b}}_{n_k} < \bar{b} - \epsilon\big) > 0$; (ii) $\lim_{n_k\to\infty} P\big(\widehat{\bar{b}}_{n_k} > \bar{b} + \epsilon\big) > 0$.

Suppose that case (i) is true, then because $G(\cdot)$ is strictly increasing on $[\underline{b}, \bar{b}]$, I have

$$\begin{aligned} 0 < \lim_{n_k\to\infty} P\big(\widehat{\bar{b}}_{n_k} < \bar{b} - \epsilon\big) &= \lim_{n_k\to\infty} P\big(G(\widehat{\bar{b}}_{n_k}) < G(\bar{b} - \epsilon)\big) \\ &\le \lim_{n_k\to\infty} P\big(G(\widehat{\bar{b}}_{n_k}) < 1\big). \end{aligned} \tag{6}$$

The last inequality holds because $G(\bar{b} - \epsilon) < G(\bar{b}) = 1$. On the other hand, since $\widehat{G}(\cdot)$ converges to $G(\cdot)$ on $[\underline{b}, \bar{b}]$, then $\widehat{G}(\widehat{\bar{b}}_{n_k})$ converges to $G(\widehat{\bar{b}}_{n_k})$ since $\lim_{n_k\to\infty} P\big(\widehat{\bar{b}}_{n_k} < \bar{b} - \epsilon\big) > 0$. The convergence constructs a contradiction because $\widehat{G}(\widehat{\bar{b}}_{n_k}) = 1$, while (6) implies that $G(\widehat{\bar{b}}_{n_k}) < 1$ with positive

[1] Both references deal with the sample maximum but the argument can be readily extended to the sample minimum. The reason is that a sample minimum from a cdf Y has the same distribution as the negative of the sample maximum from the cdf Y^*, where $Y^* = 1 - Y(-x)$.

probability. Therefore, the first case is not true, i.e., $\lim_{n_k\to\infty} P\big(\widehat{\bar{b}}_{n_k} < \bar{b} - \epsilon\big) = 0$,

Suppose the case (ii) is true, then $\widehat{\bar{b}}_{n_k} \equiv \inf\{b : \widehat{G}(b) = 1\}$ implies that $\lim_{n_k\to\infty} P\big(\widehat{G}(\bar{b}) < \widehat{G}(\widehat{\bar{b}}_{n_k})\big) > 0$. Employing the fact that $\widehat{G}(\cdot)$ converges to $G(\cdot)$ on $[\underline{b}, \bar{b}]$, we conclude that $\widehat{G}(\bar{b})$ converges to $G(\bar{b}) = 1$. Apparently, the result contradicts to $\lim_{n_k\to\infty} P\big(\widehat{G}(\bar{b}) < \widehat{G}(\widehat{\bar{b}}_{n_k})\big) > 0$. Therefore, we showed that $\lim_{n_k\to\infty} P\big(|\widehat{\bar{b}}_{n_k} - \bar{b}| > \epsilon\big) = 0$ for any $\epsilon > 0$, i.e., $\widehat{\bar{b}}_{Lk} \xrightarrow{p} \bar{b}_{Lk}$. ■

Next we establish the uniform consistency of the estimator $\widehat{\xi}_{Lk}(\cdot)$ with its rate of convergence. In establishing and proving the results, we follow Proposition 3 in GPV, and the proof there. Define $\mathscr{C}(v)$ as any closed inner subset of $[\underline{v}, \bar{v}]$, $\mathscr{C}_{Lk}(b) = \big\{b \in [\underline{v}, \bar{b}_{Lk}] : \xi_{Lk} \in \mathscr{C}(v)\big\}$, which also depends on v. The strict monotonicity and continuity of ξ_{Lk}, together with the definition that $\mathscr{C}(v)$ is a closed inner subset of $[\underline{v}, \bar{v}]$, implies that $\mathscr{C}_{Lk}$ is also a closed inner subset of $[\underline{v}, \bar{b}_{Lk}]$. According to (B.4) in the Appendix of the main paper, and the convergence of lower and upper bounds of bids, it is readily shown that when the sample size is large enough, then $\widehat{\xi}_{Lk} \neq +\infty$ if $b \in \mathscr{C}_{Lk}(b)$.

Proposition 5. *Suppose* (i) *Conditions of Proposition 2 hold;* (ii) $g(b|Lk) > c_g > 0$ *for all bids* b *and types* $Lk \in \mathcal{K}, k > 1$. *Then* $\sup_i \mathbf{1}_{\mathscr{C}(v)}(v_i)|\widehat{\xi}_{Lk}(b_i, \cdot, \cdot) - \xi_{Lk}(b_i, \cdot, \cdot)| = O(h^2 + \frac{1}{\sqrt{Nh}})$ *a.s.*

Proof For simplicity of the proof, we drop the Lk in all the related variables and let $\widehat{\xi}(b_i, \cdot, \cdot) = \widehat{\xi}_i$, $\xi(b_i, \cdot, \cdot) = \xi_i$, $G_i = G(b_{i2}|Lj)$, $g_i = g(b_{i2}|Lj)$, $\widehat{G}_i = \widehat{G}(b_{i2}|Lj)$, and $\widehat{g}_i = \widehat{g}(b_{i2}|Lj)$. Define $\widehat{c}_g = \min\{|\widehat{g}_i|, i = 1, 2, ..., N\}$, where $j = k - 1$.

To show the convergence of $\widehat{\xi}_i$, I consider the difference $|\widehat{\xi}_i - \xi_i|$ on $\mathscr{C}(v)$, i.e., $\mathbf{1}_{\mathscr{C}(v)}(v_i)|\widehat{\xi}_i - \xi_i|$. Employing (B.4) and the fact that $I \geq 2$, I have

$$\begin{aligned}\mathbf{1}_{\mathscr{C}(v)}(v_i)|\widehat{\xi}_i - \xi_i| &\leq \mathbf{1}_{\mathscr{C}(v)}(v_i)\left|\widehat{G}_i/\widehat{g}_i - G_i/g_i\right| \\ &= \mathbf{1}_{\mathscr{C}_{Lk}(b)}(b_i)\left|\widehat{G}_i/\widehat{g}_i - G_i/g_i\right|,\end{aligned}$$

where the equality is due to the the relationship between $\mathscr{C}_{Lk}(b)$ and $\mathscr{C}(v)$. We further work on the last term above. Under the assumptions imposed in the proposition, we get

$$\begin{aligned}\mathbf{1}_{\mathscr{C}_{Lk}(b)}(b_i)\left|\widehat{G}_i/\widehat{g}_i - G_i/g_i\right| &= \frac{\mathbf{1}_{\mathscr{C}_{Lk}(b)}(b_i)}{g_i|\widehat{g}_i|}\left|(\widehat{G}_i - G_i)g_i + (g_i - \widehat{g}_i)G_i\right| \\ &\leq \frac{\mathbf{1}_{\mathscr{C}_{Lk}(b)}(b_i)}{c_g\widehat{c}_g}\left||\widehat{G}_i - G_i||\bar{g}| + |g_i - \widehat{g}_i|\right|,\end{aligned}$$

where the inequality holds because of the assumption (ii), and the fact that $G_i \leq 1$. It is proved previously that the rate of convergence of $\widehat{G}_i$ is faster than that of $\widehat{g}_i$. Thus the last term implies that

$$\sup_i \mathbf{1}_{\mathscr{C}(v)}(v_i)|\widehat{\xi}_i - \xi_i| = O\left(h^2 + \frac{1}{\sqrt{Nh}}\right).$$

■

Using the above asymptotic properties of pseudovalues, we show that the estimator for value distribution, $\widehat{F}(\cdot)$, which is estimated by $\widehat{F}(\cdot) = \widehat{G}(\widehat{\xi}_{Lk}^{-1}(\cdot)|Lk)$ is uniformly consistent.

Proposition 6. *Suppose the conditions in Proposition 5 hold, then*

$$\sup_{v_i \in \mathscr{C}(v)} |\widehat{F}(v_i) - F(v_i)| = O(h^2 + \frac{1}{\sqrt{Nh}}) \ a.s.$$

Proof First consider that for any type Lk and every chosen design point $v_i \in \mathscr{C}(v)$, the value of $F(\cdot)$ is equivalently estimated as $\widehat{F}(v_i) = \widehat{G}(\widehat{\xi}^{-1}(v_i)|Lk)$, while the theoretical relationship between $F(\cdot)$ and $G(\cdot)$ is $F(v) = G(\xi^{-1}(v)|Lk)$. For simplicity, we do not specify the type Lk and denote $G(\cdot|Lk)$ and $\widehat{G}(\cdot|Lk)$ by $G(\cdot)$ and $\widehat{G}(\cdot)$, respectively. Using the relationship between $F(\cdot)$ and $G(\cdot)$, the pointwise difference between $F(\cdot)$ and $\widehat{F}(\cdot)$ can be computed as

$$\begin{aligned} & \left|\widehat{F}(v_i) - F(v_i)\right| \\ = & \left|\widehat{G}\big(\widehat{\xi}^{-1}(v_i)\big) - G\big(\widehat{\xi}^{-1}(v_i)\big) + G\big(\xi^{-1}(v_i)\big) - G\big(\widehat{\xi}^{-1}(v_i)\big)\right| \\ \leq & \left|\widehat{G}\big(\widehat{\xi}^{-1}(v_i)\big) - G\big(\widehat{\xi}^{-1}(v_i)\big)\right| + \left|G\big(\xi^{-1}(v_i)\big) - G\big(\widehat{\xi}^{-1}(v_i)\big)\right|. \end{aligned}$$

Given $\xi^{-1}(v_i) \in \mathscr{C}_{Lk}(b)$ or equivalently $v_i \in \mathscr{C}(v)$, the uniform consistency of the first term is implied by Proposition 1 for all $v_i \in \mathscr{C}(v)$. That is

$$\sup_{v_i \in \mathscr{C}(v)} \left|\widehat{G}\big(\widehat{\xi}^{-1}(v_i)\big) - G\big(\widehat{\xi}^{-1}(v_i)\big)\right| = O(\frac{1}{\sqrt{N}}).$$

To deal with the second term, we use the first order approximation of $G(\widehat{\xi}^{-1}(v_i))$ at $\xi^{-1}(v_i)$:

$$G\big(\widehat{\xi}^{-1}(v_i)\big) = G\big(\xi^{-1}(v_i)\big) + g(\eta)\big(\widehat{\xi}^{-1}(v_i) - \xi^{-1}(v_i)\big),$$

where η is between $\xi^{-1}(v_i)$ and $\widehat{\xi}^{-1}(v_i)$. Thus

$$\sup_{v_i \in \mathscr{C}(v)} \left|G\big(\widehat{\xi}^{-1}(v_i)\big) - G\big(\xi^{-1}(v_i)\big)\right| = \sup_{v_i \in \mathscr{C}(v)} \left|g(\eta)\big(\widehat{\xi}^{-1}(v_i) - \xi^{-1}(v_i)\big)\right|$$

For all $v_i \in \mathscr{C}(v)$, $g(\eta)$ is bounded by assumption, hence

$$\begin{aligned} \sup_{v_i \in \mathscr{C}(v)} \left| G\big(\widehat{\xi}^{-1}(v_i)\big) - G\big(\xi^{-1}(v_i)\big) \right| &= \sup_{v_i \in \mathscr{C}(v)} \left| \widehat{\xi}^{-1}(v_i) - \xi^{-1}(v_i) \right| \\ &= O\left(h^2 + \frac{1}{\sqrt{Nh}} \right). \end{aligned}$$

■

Asymptotic Normality of $\widehat{G}(b_2|\tau)$ and $\widehat{g}(b_2|\tau)$ To prove the asymptotic normality of the estimator $\widehat{G}_{b_2|\tau}$, we first define $\rho_0(b_2)$ and $\widehat{\rho}(b_2)$ as $m \times 1$ column vectors contain all the elements of $G(b_2, d_1)$ and $\widehat{G}(b_2, d_1)$, respectively. Since $\widehat{p}(\tau) = O_p(1)$ and $\sup_{b_2} |\widehat{G}(b_2, d_1)| < \infty$, then the estimate of $\widehat{G}_{b|\tau}$ can be rewritten as

$$\widehat{G}(b_2|Lk) = \frac{e_k^T B_{d_1|\tau}^{-1} \widehat{G}(b_2, d_1)}{e_k^T B_{d_1|\tau}^{-1} p(d_1)} + O_p(N^{-1/2}), \tag{7}$$

where e_k is defined as a unit column vector with k-th component is one and all other components are zeros. Furthermore, the equation above can be rewritten as follows.

$$\widehat{G}(b_2|Lk) = \psi\big(\widehat{\rho}(b_2)\big) + O_p(N^{-1/2}),$$

where

$$\psi\big(\widehat{\rho}(b_2)\big) \equiv \frac{e_k^T B_{d_1|\tau}^{-1} \widehat{G}(b_2, d_1)}{e_k^T B_{d_1|\tau}^{-1} p(d_1)}.$$

The function $\psi(\cdot)$ defined above is non-stochastic since it only depends on $B_{d_1|\tau}^{-1}$ and $p(d_1)$. Also $\psi(\cdot)$ is linear in each entry of the argument $\widehat{\rho}(b_2)$. Next we employ the delta method to show the asymptotic normality of $\widehat{G}_{b_2|\tau}$. The asymptotic normality of $\widehat{G}_{b_2|\tau}$ is summarized as follows.

Proposition 7. *The estimator of conditional bid distribution $\widehat{G}(b|\tau)$ as a vector is asymptotically normally distributed. More specifically, $\sqrt{N}\big[\widehat{G}(b_2|\tau) - G(b_2|\tau)\big] \xrightarrow{d} \mathcal{N}\big(0, \Omega\big)$, where $\Omega = \left(\frac{d\psi}{d\rho}\right)^T \big|_{\rho=\rho_0} V \left(\frac{d\psi}{d\rho}\right) \big|_{\rho=\rho_0}$, and V is the asymptotic variance of $\widehat{\rho}(b_2)$.*

Proof Taylor expansion of $\psi\big(\widehat{\rho}(b_2)\big)$ gives

$$\begin{aligned} \widehat{G}(b_2|\tau) - G(b_2|\tau) &= \psi\big(\widehat{\rho}(b_2)\big) - \psi\big(\rho_0(b_2)\big) \\ &= \left(\frac{d\psi}{d\rho}\right)^T \Bigg|_{\rho=\rho_0} \big(\widehat{\rho}(b_2) - \rho_0(b_2)\big) + o_p(N^{-1/2}). \end{aligned} \tag{8}$$

The asymptotic normality of $\widehat{\rho}(b_2)$ is a well-known result, i.e.,

$$\sqrt{N}\Big[\widehat{\rho}(b_2) - \rho_0(b_2)\Big] \xrightarrow{d} \mathcal{N}\big(0, V\big), \tag{9}$$

where the variance matrix $V \equiv \mathbb{E}\Big[G(b_2, d_1)\big(1 - G(b_2, d_1)\big)^T\Big]$.

(8) and (9) together imply that

$$\sqrt{N}\big[\widehat{G}(b_2|\tau) - G(b_2|\tau)\big] \xrightarrow{d} \mathcal{N}\big(0, v\big),$$

where the variance matrix Ω is

$$\Omega = \left(\frac{d\psi}{d\rho}\right)^T\Big|_{\rho=\rho_0} V \left(\frac{d\psi}{d\rho}\right)\Big|_{\rho=\rho_0}.$$

■

Let $\gamma(\cdot)$ and $\phi(\cdot)$ be the counterparts of $\rho(\cdot)$ and $\psi(\cdot)$, respectively, and $\phi(\gamma(b_2))$ is also defined analogously to $\psi(\rho(b_2))$. Then the asymptotic normality of $\widehat{g}(b_2|\tau)$ is similar to that of $\widehat{G}(b_2|\tau)$, and the proof is similar, too. The result of asymptotic normality is presented in the following proposition.

Proposition 8. *Under Assumptions* $1-3$*, the estimator for conditional bid density* $\widehat{g}(b|\tau)$ *as a vector is asymptotically normally distributed. More specifically,* $\sqrt{N}\big[\widehat{g}(b_2|\tau) - g(b_2|\tau)\big] \xrightarrow{d} \mathcal{N}\big(0, \Sigma'\big)$*, where* $\Sigma' = \left(\frac{d\phi}{d\gamma}\right)^T\big|_{\gamma=\gamma_0}\Sigma\left(\frac{d\phi}{d\gamma}\right)\big|_{\gamma=\gamma_0}$*, and* Σ *is the asymptotic variance of* $\widehat{\gamma}(b_2)$*.*

Proof The strategy of proving the asymptotic normality of $\widehat{g}(b|\tau)$ is similar to that of $\widehat{G}(b|\tau)$. Hence we will only sketch the proof. First let γ and ϕ be the counterparts of ρ and ψ, respectively, and $\phi(\gamma(b_2))$ is also defined analogously. Then we have

$$\begin{aligned} \widehat{g}(b_2|\tau) - g(b_2|\tau) &= \phi\big(\widehat{\gamma}(b_2)\big) - \phi\big(\gamma_0(b_2)\big) \\ &= \left(\frac{d\phi}{d\gamma}\right)^T\Big|_{\gamma=\gamma_0}\big(\widehat{\gamma}(b_2) - \gamma_0(b_2)\big) + o_p(N^{-1/2}), \end{aligned} \tag{10}$$

The asymptotic normality of $\widehat{\gamma}(b_2) - \gamma_0(b_2)$ can be readily derived based on Theorem 2.10 in Pagan and Ullah (1999) which shows that each component of $\sqrt{Nh}\big(\widehat{\gamma}(b_2) - \gamma_0(b_2)\big)$ is asymptotically normal distributed. Therefore, $\sqrt{Nh}\big(\widehat{\gamma}(b_2) - \gamma_0(b_2)\big)$ is distributed according to a multidimensional normal distribution with variance matrix Σ. Consequently, we can obtain the asymptotic normality of $\widehat{g}(b_2|\tau) - g(b_2|\tau)$ as

$$\sqrt{Nh}\big[\widehat{g}(b_2|\tau) - g(b_2|\tau)\big] \xrightarrow{d} \mathcal{N}\big(0, \Sigma'\big), \tag{11}$$

where the variance matrix Σ' is $\Sigma' = \left(\frac{d\phi}{d\gamma}\right)^T\Big|_{\gamma=\gamma_0} \Sigma \left(\frac{d\phi}{d\gamma}\right)\Big|_{\gamma=\gamma_0}$. ■

References

AL-Hussaini, E., El-Adll, M., 2004. Asymptotic Distribution of Normalized Maximum under Finite Mixture Models. Statistics & Probability Letters 70 (1), 109–117.

Andrew, A., Chu, K., Lancaster, P., 1993. Derivatives of Eigenvalues and Eigenvectors of Matrix Functions. SIAM Journal on Matrix Analysis and Applications 14 (4), 903–926.

Arnold, B., Balakrishnan, N., Nagaraja, H., 2008. A First Course in Order Statistics. Society for Industrial Mathematics.

Fan, J., Yao, Q., 2005. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

Gillen, B., 2009. Identification and Estimation of Level-k Auctions, manuscript, Department of Economics, UCSD.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press.

Sreehari, M., Ravi, S., 2010. On Extremes of Mixtures of Distributions. Metrika 71 (1), 117–123.