Modus operandi of empirical (data) work:

- We observe bids $b_1, \ldots, b_n$, and we want to recover valuations $v_1, \ldots, v_n$.
- Why? Analogously to demand estimation, we can evaluate the "market power" of bidders, as measured by the margin $v - p$.

  Could be interesting to examine: how fast does margin decrease as n (number of bidders) increases?
- Useful for the optimal design of auctions:
  1. What is auction format which would maximize seller revenue?
  2. What value for reserve price would maximize seller revenue?
- Testing between CV and PV models

  Very different behavioral implications

1 Nonparametric Identification and Estimation in IPV First-price Auction Model

- Main reference: Guerre, Perrigne, and Vuong (2000)
- Recall first-order condition for equilibrium bid (general affiliated values case):

$$b'(x) = (v(x,x) - b(x)) \cdot \frac{f_{y_i|x_i}(x|x)}{F_{y_i|x_i}(x|x)}; \tag{1}$$

  where $y_i \equiv \max_{j \neq i} x_i$ (highest among rivals' signals) and $b(\cdot)$ denotes the equilibrium bidding strategy.
- In IPV case: $V_i = X_i$, so that

$$\begin{aligned} v(x,x) &= x \\ F_{y_i|x_i}(x|x) &= F(x)^{n-1} \\ f_{y_i|x_i}(x|x) &= \frac{\partial}{\partial x} F(x)^{n-1} = (n-1)F(x)^{n-2} f(x). \end{aligned}$$

Hence, first-order condition becomes

$$\begin{aligned} b'(x) &= (x - b(x)) \cdot (n-1)\frac{F(x)^{n-2}f(x)}{F(x)^{n-1}} \\ &= (x - b(x)) \cdot (n-1)\frac{f(x)}{F(x)}. \end{aligned} \tag{2}$$

- Now, note that because equilibrium bidding function $b(x)$ is just a monotone increasing function of the valuation x. Hence, for $b_i \equiv b(x_i)$:
  - The cumulative distribution function of the bids is:

$$G(b_i) = P(b \le b_i) = P(x \le x_i) = F(x_i) \tag{3}$$

  - Correspondingly, the bid density function can be obtained by differentiation:

$$g(b_i) = \frac{\partial G(b_i)}{\partial b_i} = \frac{\partial F(x_i)}{\partial x_i} \cdot \frac{\partial x_i}{\partial b_i} = f(x_i) \cdot \frac{\partial b^{-1}(b_i)}{\partial b_i} = f(x_i) \cdot 1/b'(x_i). \tag{4}$$

Hence, substituting the above into Eq. (2):

$$\begin{aligned} \frac{1}{g(b_i)} &= (n-1)\frac{x_i - b_i}{G(b_i)} \\ \Leftrightarrow x_i &= b_i + \frac{G(b_i)}{(n-1)g(b_i)}. \end{aligned} \tag{5}$$

Everything on the RHS of the preceding equation is observed: the equilibrium bid CDF G and density g can be estimated directly from the data *nonparametrically.* Assuming a dataset consisting of T n-bidder auctions:

$$\begin{aligned} \hat{g}(b) &\approx \frac{1}{T \cdot n}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{1}{h}\mathcal{K}\left(\frac{b - b_{it}}{h}\right) \\ \hat{G}(b) &\approx \frac{1}{T \cdot n}\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbf{1}(b_{it} \le b). \end{aligned} \tag{6}$$

The first is a *kernel density estimate* of bid density. The second is the *empirical distribution function (EDF).*

- In the above, $\mathcal{K}$ is a "kernel function". A kernel function is a function satisfying the following conditions:

1. It is a probability density function, ie: $\int_{-\infty}^{+\infty} \mathcal{K}(d)du = 1$, and $\mathcal{K}(u) \geq 0$ for all u.
2. It is symmetric around zero: $\mathcal{K}(u) = \mathcal{K}(-u)$.
3. h is bandwidth: describe below
4. Examples:
   (a) $\mathcal{K}(u) = \phi(u)$ (standard normal density function);
   (b) $\mathcal{K}(u) = \frac{1}{2}\mathbf{1}(|u| \leq 1)$ (uniform kernel);
   (c) $\mathcal{K}(u) = \frac{3}{4}(1-u^2)\mathbf{1}(|u| \leq 1)$ (Epanechnikov kernel)

- To get some intuition for the kernel estimate of $\hat{g}(b)$, consider the histogram

$$h(b) = \frac{1}{Tn}\sum_t \sum_i \mathbf{1}(b_{it} \in [b-\epsilon, b+\epsilon])$$

  for some small $\epsilon > 0$. The histogram at b, $h(b)$ is the frequency with which the observed bids land within an ϵ-neighborhood of b.

- In comparison, the kernel estimate of $\hat{g}(b)$ replaces $\mathbf{1}(b_{it} \in [b-\epsilon, b+\epsilon])$ with $\frac{1}{h}\mathcal{K}\left(\frac{b-b_{it}}{h}\right)$. This is:

  - always ≥ 0
  - takes large values for b_{it} close to b; small values (or zero) for b_{it} far from b
  - takes values in $\mathbb{R}+$ (can be much larger than 1)
  - h is bandwidth, which blows up $\frac{1}{h}\mathcal{K}\left(\frac{b-b_{it}}{h}\right)$: when it is smaller, then this quantity becomes larger.

    Think of h as measuring the "neighborhood size" (like ϵ in the histogram). When $T \to \infty$, then we can make h smaller and smaller.

    Bias/variance tradeoff.
  - Roughly speaking, then, $\hat{g}(b)$ is a "smoothed" histogram,

- For $\hat{G}(b)$, recall definition of the CDF:

$$G(\tilde{b}) = Pr(b \leq \tilde{b}).$$

  The EDF measures these probabilities by the (within-sample) frequency of the events.

- Hence, the IPV first-price auction model is *nonparametrically identified*. For each observed bid b_{it}, the corresponding valuation $x_{it} = b^{-1}(b_{it})$ can be recovered as:

$$\hat{x}_{it} = b_{it} + \frac{\hat{G}(b_{it})}{(n-1)\hat{g}(b_{it})}. \tag{7}$$

Hence, GPV recommend a two-step approach to estimating the valuation distribution $f(x)$:

1. In first step, estimate $G(b)$ and $g(b)$ nonparametrically, using Eqs. (6).

2. In second step, estimate the density $f(x)$ and CDF $F(x)$ of valuations by using kernel density estimator of recovered valuations:

$$\begin{aligned} \hat{f}(x) &\approx \frac{1}{T \cdot n} \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{1}{h} \mathcal{K}\left(\frac{x - \hat{x}_{it}}{h}\right). \\ \hat{F}(x) &\approx \frac{1}{T \cdot n} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathbf{1}\left(\hat{x}_{it} \leq x\right). \end{aligned} \tag{8}$$

1.1 Optimal reserve price

With knowledge of $f(x)$ and $F(x)$, you can compute the optimal reserve price:

$$\hat{r}:\ \hat{r} - \frac{1 - \hat{F}(\hat{r})}{f(\hat{r})} = 0.$$

1.2 Only winning bids observed

As a simple extension, we see that identification continues to hold, even when only the highest-bid in each auction is observed. Specifically, if only $b_{n:n} \equiv \max(b_1, \ldots, b_n)$ is observed, we can estimate $G_{n:n}$, the CDF of the maximum bid, from the data. Note that the relationship between the CDF of the maximum bid and the marginal CDF of an equilibrium bid is

$$G_{n:n}(b) = G(b)^n$$

implying that $G(b)$ can be recovered from knowledge of $G_{n:n}(b)$. Once $G(b)$ is recovered, the corresponding density $g(b)$ can also be recovered, and we could solve Eq. (7) for every b to obtain the inverse bid function.

■■■

2 Affiliated values models

Can this methodology be extended to affiliated values models (including common value models)?

To prepare what follows, we introduce n subscript (so we index distributions according to the number of bidders in the auction).

Go back to first order condition for this model is: for bidder i

$$b'(x,n) = (v(x,x,n) - b(x,n)) \cdot \frac{f_{y_i|x_i,n}(x|x)}{F_{y_i|x_i,n}(x|x)};$$

where $y_i \equiv \max_{j \neq i} \{x_1, \ldots, x_n\}$, and $v(x,x,n) = E[V_i|X_i = x, Y_i = x]$.

As before, because of the monotonicity of the bidding strategy $b(x,n)$ in x, we can exploit the following change of variable formulas:

- $$G_{b^*|b,n}(b|b) = F_{y|x,n}(x|x)$$
- $$g_{b^*|b,n}(b|b) = f_{y|x,n}(x|x) \cdot 1/b'(x)$$

where b^* denotes (for a given bidder), the highest bid submitted by this bidder's rivals: for a given bidder i, $b_i^* = \max_{j \neq i} b_j$.

Hence, by considering some bid $b = b(x,n)$, and substituting the above into the first-order condition, we obtain:

$$v(x,x,n) = b + \frac{G_{b^*|b}(b|b,n)}{g_{b^*|b}(b|b,n)}. \tag{9}$$

Procedure similar to GPV can be used here to recover, for each bid b_i, the corresponding quantity $\frac{G_{b^*|b}(b|b,n)}{g_{b^*|b}(b|b,n)}$ (see below).

That is, for a given bid b, we can recover the corresponding $v(x,x,n)$. We cannot recover the signal x which caused this bidder to submit a bid equal to $b = b(x,n)$, but we can recover the "expected valuation conditional on winning".

But it turns out this is enough for determining whether the bids came from a common value or private value environment.

2.1 Testing between CV and PV

Recall the winner's curse: it implies that $v(x,x,n)$ is invariant to n for all x in a PV model but strictly decreasing in n for all x in a CV model.

In Haile, Hong, and Shum (2003), we use this intuition to develop a test for CV:

$$H_0 \text{ (PV)}: \; E\left[v(X,X;\underline{n})\right] = E\left[v(X,X;\underline{n}+1)\right] = \cdots = E\left[v(X,X;\bar{n})\right]$$
$$H_1 \text{ (CV)}: \; E\left[v(X,X;\underline{n})\right] > E\left[v(X,X;\underline{n}+1)\right] > \cdots > E\left[v(X,X;\bar{n})\right]$$

2.2 Technical details

Recall the fundamental probability laws,

$$g_{b^*,b,n}(b,b) = g_{b^*|b,n}(b|b) \cdot g_n(b)$$

where $g_n(b)$ denotes the marginal density of bids. Then the fraction in the key equation (9) is equivalent to

$$\frac{G_n(b;b)}{g_n(b;b)} = \frac{G_n(b|b)}{g_n(b|b)}. \tag{10}$$

Li, Perrigne, and Vuong (2000) suggest kernel-based nonparametric estimates for $g_n(b;b)$ and $G_n(b;b)$ where

$$G_n(b;b) \equiv G_n(b|b)g_n(b) = \frac{\partial}{\partial b}\Pr(B^*_{it} \leq m, B_{it} \leq b)|_{m=b}$$

and

$$g_n(b;b) \equiv g_n(b|b)g_n(b) = \frac{\partial^2}{\partial m \partial b}\Pr(B^*_{it} \leq m, B_{it} \leq b)|_{m=b}$$

and $g_n(\cdot)$ is the marginal density of bids in equilibrium. The kernel-based estimators are:

$$\begin{aligned} \hat{G}_n(b;b) &= \frac{1}{T_n \times h \times n}\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{b-b_{it}}{h}\right)\mathbf{1}\left(b^*_{it} < b, n_t = n\right) \\ \hat{g}_n(b;b) &= \frac{1}{T_n \times h^2 \times n}\sum_{t=1}^{T}\sum_{i=1}^{n} \mathbf{1}(n_t = n) K\left(\frac{b-b_{it}}{h}\right) K\left(\frac{b-b^*_{it}}{h}\right). \end{aligned} \tag{11}$$

Here, as above, h is a bandwidth and $K(\cdot)$ is a kernel function.

Hence, by evaluating $\hat{G}_n(\cdot,\cdot)$ and $\hat{g}_n(\cdot,\cdot)$ at each observed bid b_{it}, we can construct a pseudo-sample of estimates of

$$\hat{v}_{it} = b_{it} + \frac{\hat{G}_n(b_{it};b_{it})}{\hat{g}_n(b_{it};b_{it})}. \tag{12}$$

where each $v_{it} = E[V_i|X_i = x_{it}, Y_i = x_{it}]$, the expected value of winning for a bidder who submitted the bid b_{it}.

References

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–74.

HAILE, P., H. HONG, AND M. SHUM (2003): "Nonparametric Tests for Common Values in First-Price Auctions," NBER working paper #10105.

LI, T., I. PERRIGNE, AND Q. VUONG (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129–161.