Semi-Nonparametric Modeling and Estimation of First-Price Auction Models with Auction-Specific Heterogeneity

Herman J. Bierens and Hosin Song

Introduction

- In many repeated first-price auctions the objects to be auctioned off are different across auctions.
- The question then arises how to incorporate the observable auction-specific characteristics into the auction model.
- Laffont et al. (1995) incorporate covariates in the value distribution by specifying a linear regression model for the log of values with zero-mean normal errors.
- Donald and Paarsch (1996) parameterize the upper bound of the values as a function of covariates.
- Li (2005) specifies the value distribution as the exponential distribution with mean a linear function of covariates.

- Guerre et al. (2000) propose a two-stage nonparametric kernel density estimation approach, where in the first stage the bid distribution and density conditional on the covariates are estimated nonparametrically, which then is used in inverse form to generate values given the actual bids and the covariates.
  The generated values are then used to estimate the conditional value distribution nonparametrically.
- The problem with the latter approach is that the results can only be displayed in the form of graphs.
- But how do you display functions of more than two variables?
- Therefore, in this paper we propose a semi-nonparametric (SNP) approach, where similar to Laffont et al. (1995) the log-values are specified by a linear model in the auctions-specific covariates, but with unknown error distribution.

- In this paper we will set forth mild conditions such that the parameter vector in the log-linear value model can be estimated consistently and asymptotic normally without knowing the exact form of the error distribution.
- Given this consistent estimator, the error distribution will be modeled semi-nonparametrically, via an infinite series expansion.
- This error distribution can now be consistently estimated via a simulated method of moments sieve estimation approach, (somewhat) similar to:

  Bierens, H. J. and H. Song (2012), ”Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method”, *Journal of Econometrics* 168, 108-119.

- The semi-nonparametric (SNP) analysis involved is an adaptation to SNP auction models of the approach in

  Bierens, H. J. (2014), ”Consistency and Asymptotic Normality of Sieve ML Estimators Under Low-Level Conditions”, *Econometric Theory* 30, 1021-1076.

  Bierens, H. J. (2014), ”The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling”, Chapter 1 in Racine, J., L. Su & A. Ullah (eds.), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press.

First price auctions with observed heterogeneity

In this paper we consider first price sealed bid auctions where values are independent and private, the bidders are symmetric and risk-neutral, and the objects to be auctioned off are heterogenous across auctions.

As is well-known, the equilibrium bid function takes the form

$$b_0\,(v|X) = v - \Gamma_0(v|X)^{1-I_X}\int_{p_X}^{v}\Gamma_0(y|X)^{I_X-1}\mathrm{d}y$$

for $v > p_X$, where

- p_X is the seller's reservation price,
- $I_X \geq 2$ is the number of potential bidder, and
- $\Gamma_0(v|X)$ is the conditional distribution of the private value V that each potential bidder has for the object to be auctioned off.

However, in the paper it will be assumed that there is no reservation price, so that the equilibrium bid function becomes

$$b_0\left(v|X, I_X\right) = v - \Gamma_0(v|X)^{1-I_X} \int_0^v \Gamma_0(y|X)^{I_X-1} \mathrm{d}y,$$

where now I_X is the actual number of bidders in this auction. Thus, for each auction $k = 1, ..., L$ with auction-specific covariates X_k we observe $I_k = I_{X_k}$ bids

$$\begin{aligned} B_{k,j} &= b_0(V_{k,j}|X_k, I_k) \\ &= V_{k,j} - \Gamma_0(V_{k,j}|X)^{1-I_k} \int_0^{V_{k,j}} \Gamma_0(y|X_k)^{I_k-1} dy, \\ j &= 1, ..., I_k, \end{aligned}$$

where given X_k the values $V_{k,j}$, $j = 1, ..., I_k$ are independent random drawings from the true conditional value distribution $\Gamma_0(v|X_k)$, and $b_0(v|X_k, I_k)$ is the corresponding true bid function.

More precisely, it will be assumed that

Assumption 1. *We observe* L *auctions, labeled* $k = 1, 2, ..., L$*, each characterized by a vector* $X_k \in \mathbb{R}^p$ *of auction-specific covariates and a number* $I_k = I_{X_k} \geq 2$ *of bidders.*
The sellers in these auctions do not ex-ante reveal their reservation prices, if any.
The random vectors $X_k^+ = (X_k', I_k)', k = 1, 2, ..., L,$ *are independently and identically distributed as* $X^+ = (X', I_X)'$.
The bids $B_{j,k}, j = 1, 2, ..., I_k,$ *in auction* k *are observed and are distributed as*

$$b_0(V_{j,k}|X_k^+) = V_{k,j} - \Gamma_0(V_{k,j}|X)^{1-I_k} \int_0^{V_{k,j}} \Gamma_0(y|X_k)^{I_k-1} \mathrm{d}y,$$

where for each k *the values* $V_{j,k}$, $j = 1, 2, ..., I_k$*, are independent random drawings from the conditional value distribution* $\Gamma_0(v|X_k)$.

The conditional value distribution

To incorporate auction-specific heterogeneity in the conditional value distribution we need to put some structure on the conditional value distribution $\Gamma_0(v|X)$.
We will do that by assuming a log-linear model for the values.

Assumption 2. *Each value* $V_{i,k}$ *of bidder* i *in auction* k *is generated by*

$$\ln(V_{i.k}) = \beta_0' X_k + \ln(W_{i,k}),$$

where within and across auctions the $W_{i,k}$*'s are drawn independently from an unknown absolutely continuous distribution* $F_0(w)$ *with continuous density* $f_0(w)$ *and support* $(0, \infty)$. *Moreover,*

$$E[W_{i,k}] = \int_0^\infty w f_0(w) \mathrm{d}w < \infty$$

In the auction literature it is usually assumed that the value-distribution has bounded support.
This bounded value condition assures that the expected revenue of the seller is finite.
However, the latter also holds, conditional on $X^+ = (X', I_X)'$, under the moment condition

$$\int_0^\infty w f_0(w) \mathrm{d}w < \infty.$$

The conditional bid function

The conditional distribution function of V given X now takes the form

$$\Gamma_0(v|X) = F_0(v.\exp(-\beta_0' X)).$$

Therefore, the actual conditional bid function can be written as

$$\begin{aligned} b_0\left(v|X^+\right) &= v - \frac{\exp(\beta_0' X)}{F_0(v.\exp(-\beta_0' X))^{I_X-1}} \int_0^{v\exp(-\beta_0' X)} F_0(y)^{I_X-1}\mathrm{d}y \\ &= \frac{(I_X-1)\exp(\beta_0' X)}{F_0(v.\exp(-\beta_0' X))^{I_X-1}} \int_0^{v\exp(-\beta_0' X)} yF_0(y)^{I_X-2} f_0(y)\mathrm{d}y \\ &= \frac{(I_X-1)\exp(\beta_0' X)}{F_0(v.\exp(-\beta_0' X))^{I_X-1}} \int_0^{F_0(v\exp(-\beta_0' X))} z^{I_X-2} F_0^{-1}(z)\mathrm{d}z \end{aligned}$$

where F_0^{-1} is the inverse of F_0 and $X^+ = (X', I_X)'$.

As is well-known, a random drawing V from the conditional value distribution

$$\Gamma_0(v|X) = F_0(v.\exp(-\beta_0' X)).$$

can be generated by solving

$$F_0(V.\exp(-\beta_0' X)) = U,$$

hence

$$V = \exp(\beta_0' X) F_0^{-1}(U),$$

where U is a random drawing from the uniform distribution on $(0, 1)$.

Therefore, a typical bid B in this auction is distributed as

$$\begin{aligned} b_0\left(V|X^+\right) &= \frac{(I_X-1)\exp(\beta_0'X)}{F_0(V.\exp(-\beta_0'X))^{I_X-1}}\int_0^{F_0(V\exp(-\beta_0'X))} z^{I_X-2}F_0^{-1}(z)\mathrm{d}z \\ &= \frac{\exp(\beta_0'X)}{U^{I_X-1}}\int_0^U (I_X-1)z^{I_X-2}F_0^{-1}(z)\mathrm{d}z \\ &= \exp(\beta_0'X)\int_0^1 (I_X-1)z^{I_X-2}F_0^{-1}(z.U)\mathrm{d}z \\ &= \bar{b}_0(U|X^+), \text{ say,} \end{aligned}$$

where for $u \in [0,1]$,

$$\bar{b}_0(u|X^+) = \exp(\beta_0'X)\int_0^1 (I_X-1)z^{I_X-2}F_0^{-1}(z.u)\mathrm{d}z.$$

Clearly, the new bid function

$$\bar{b}_0(u|X^+) = \exp(\beta_0' X) \int_0^1 (I_X - 1) z^{I_X - 2} F_0^{-1}(z.u)\mathrm{d}z$$

is strictly monotonic increasing in u, with upper bound

$$\begin{aligned}
\bar{b}_0\left(1|X^+\right) &= (I_X - 1)\exp(\beta_0' X) \int_0^1 z^{I_X - 2} F_0^{-1}(z)\mathrm{d}z \\
&\leq (I_X - 1)\exp(\beta_0' X) \int_0^1 F_0^{-1}(z)\mathrm{d}z \\
&= (I_X - 1)\exp(\beta_0' X) \int_0^\infty w f_0(w)\mathrm{d}w,
\end{aligned}$$

where the last equality follows from

$$\infty > \int_0^\infty w f_0(w)\mathrm{d}w = \int_0^\infty F_0^{-1}(F_0(w))\mathrm{d}F_0(w) = \int_0^1 F_0^{-1}(z)\mathrm{d}z.$$

Thus, the bids $B_{j,k}, j = 1, 2, ..., I_k$, in auctions $k = 1, 2, ..., L$ are distributed as

$$\bar{b}_0(U_{j,k}|X_k^+) = \exp(\beta_0' X_k) \int_0^1 (I_k - 1) z^{I_k - 2} F_0^{-1}(z.U_{j,k}) \mathrm{d}z$$

where the $U_{j,k}$'s are independent random drawings from the uniform distribution on $(0, 1)$.

Semi-nonparametric identification

The question I will now address is: Are β_0 and F_0 identified? If not, then there exists a β and an absolutely continuous distribution function F with support $(0, \infty)$ and alternative bid function

$$\overline{b}(u|X^+) = \exp(\beta' X) \int_0^1 (I_X - 1) z^{I_X - 2} F^{-1}(z.u) \mathrm{d}z$$

such that for a pair U_1, U_2 of uniformly $(0, 1)$ distributed random variables,

$$\overline{b}_0(U_1|X^+) \sim \overline{b}(U_2|X^+).$$

This is only possible if

$$\overline{b}_0(u|X^+) = \overline{b}(u|X^+) \text{ a.s. for all } u \in (0, 1),$$

It can now be shown that if

Assumption 3. $\mathrm{Var}(X)$ *is finite and nonsingular,*

then $\beta = \beta_0$ and $F \equiv F_0$.

Semi-parametric estimation of the log-linear value model

I will now set forth additional conditions for the construction of a consistent estimator $\widetilde{\beta}_L$ of the parameter vector β_0 in the log-linear value model

$$\ln(V_{i.k}) = \beta_0' X_k + \ln(W_{i,k}),$$

such that for $L \to \infty$,

$$\sqrt{L}\left(\widetilde{\beta}_L - \beta_0\right) \xrightarrow{\text{d}} N_p(0, \Sigma),$$

without the need to know the distribution function F_0.

Recall that the bids $B_{i,k}$, $i = 1, 2, ..., I_k$, in auction k are distributed as

$$\overline{b}_0\left(U_{i,k}|X_k^+\right) = \exp(\beta_0' X_k) \int_0^1 (I_k - 1) z^{I_k - 2} F_0^{-1}(U_{i,k}.z) \mathrm{d}z$$

where the $U_{i,k}$'s are random drawings from the uniform $(0, 1)$ distribution.

Hence

$$b_{i,k} \stackrel{\text{def.}}{=} \ln(B_{i,k}) \sim \beta_0' X_k + Z_{i,k}$$

where

$$Z_{i,k} = \ln\left(\int_0^1 (I_k - 1) z^{I_k - 2} F_0^{-1}(U_{i,k}.z) \mathrm{d}z\right).$$

Now suppose that

Assumption 4. $\int_0^\infty (\ln(w))^2 f_0(w) \mathrm{d}w < \infty,$

so that the errors $\ln(W_{i,k})$ in the log-linear value model

$$\ln(V_{i.k}) = \beta_0' X_k + \ln(W_{i,k}),$$

have finite second moments.

This is a standard condition for least squares estimation of log-linear regression models.

Then it can be shown that under Assumptions 1-4,

$$E[Z_{i,k}^2] < \infty.$$

Next, suppose that

Assumption 5. *The marginal distribution of the number of bidders, I_X, has finite support, i.e., or some fixed $M \in \mathbb{N}\backslash\{1\}$, $\Pr[I_X = K_m] = p_m > 0, \sum_{m=1}^{M} p_m = 1$, where $2 \leq K_1 < K_2 < ... < K_M$.*

The trick is split up the data in sub-samples according to the values of I_k, as follows.
Denote for $K \geq 2$,

$$\Omega(u|K) = \ln\left((K-1)\int_0^1 z^{K-2}F_0^{-1}(u.z)\mathrm{d}z\right),$$

so that the error term $Z_{i,k}$ in the log-linear model

$$b_{i,k} = \ln(B_{i,k}) = \beta_0' X_k + Z_{i,k}$$

takes the form

$$Z_{i,k} = \Omega(U_{i,k}|I_k).$$

Then it follows that, with $\mathbf{1}(.)$ the indicator function,

$$(b_{i,k} - \beta' X_k) \mathbf{1}(I_k = K_m)$$
$$= (\beta_0 - \beta)' X_k \mathbf{1}(I_k = K_m) + \Omega(U_{i,k}|K_m) \mathbf{1}(I_k = K_m),$$
$$i = 1, 2, ..., K_m.$$

Next, denote for $m = 1, 2, 3, ..., M$,

$$\widehat{p}_{m,L} = \frac{1}{L} \sum_{j=1}^{L} \mathbf{1}(I_j = K_m),$$

$$\overline{X}_{m,L} = \frac{\frac{1}{L} \sum_{j=1}^{L} X_j . \mathbf{1}(I_j = K_m)}{\widehat{p}_{m,L}},$$

$$\overline{b}_{i,m,L} = \frac{\frac{1}{L} \sum_{j=1}^{L} b_{i,j} . \mathbf{1}(I_j = K_m)}{\widehat{p}_{m,L}}.$$

The statistic $\overline{b}_{i,m,L}$ can be written as

$$\overline{b}_{i,m,L} = \beta_0' \overline{X}_{m,L} + \frac{\frac{1}{L} \sum_{j=1}^{L} \Omega(U_{i,j}|K_m) . \mathbf{1}(I_j = K_m)}{\widehat{p}_{m,L}}.$$

It can now be shown that

$$\sum_{m=1}^{M}\frac{1}{L}\sum_{k=1}^{L}(X_k-\overline{X}_{m,L})\frac{1}{K_m}\sum_{i=1}^{K_m}(b_{i,k}-\overline{b}_{i,m,L})\mathbf{1}\left(I_k=K_m\right)$$
$$-\sum_{m=1}^{M}\left(\frac{1}{L}\sum_{k=1}^{L}(X_k-\overline{X}_{m,L})(X_k-\overline{X}_{m,L})'\mathbf{1}\left(I_k=K_m\right)\right)\beta$$
$$=\sum_{m=1}^{M}\left(\frac{1}{L}\sum_{k=1}^{L}(X_k-\overline{X}_{m,L})(X_k-\overline{X}_{m,L})'\mathbf{1}\left(I_k=K_m\right)\right)(\beta_0-\beta)$$
$$+\sum_{m=1}^{M}\frac{1}{L}\sum_{k=1}^{L}\frac{1}{K_m}\sum_{i=1}^{K_m}\left(\Omega(U_{i,k}|K_m)-\int_0^1\Omega(u|K_m)\mathrm{d}u\right)$$
$$\times(X_k-\overline{X}_{m,L})\mathbf{1}\left(I_k=K_m\right).$$

Consequently, denoting

$$
\begin{aligned}
\widetilde{\Sigma}_{1,L} &= \sum_{m=1}^{M} \frac{1}{L} \sum_{k=1}^{L} (X_k - \overline{X}_{m,L})(X_k - \overline{X}_{m,L})' \mathbf{1}\left(I_k = K_m\right), \\
\widetilde{\eta}_L &= \sum_{m=1}^{M} \frac{1}{L} \sum_{k=1}^{L} (X_k - \overline{X}_{m,L}) \frac{1}{K_m} \sum_{i=1}^{K_m} \left(b_{i,k} - \overline{b}_{i,m,L}\right) \mathbf{1}\left(I_k = K_m\right), \\
\widetilde{\rho}_L &= \frac{1}{L} \sum_{k=1}^{L} \sum_{m=1}^{M} \frac{1}{K_m} \sum_{i=1}^{K_m} \left(\Omega(U_{i,k} | K_m) - \int_0^1 \Omega(u|K_m) \mathrm{d}u \right) \\
&\quad \times (X_k - \overline{X}_{m,L}) \mathbf{1}\left(I_k = K_m\right),
\end{aligned}
$$

we have

$$
\widetilde{\eta}_L - \widetilde{\Sigma}_{1,L}\beta = \widetilde{\Sigma}_{1,L}(\beta_0 - \beta) + \widetilde{\rho}_L.
$$

Hence,

$$\widetilde{\Sigma}_{1,L}^{-1}\widetilde{\eta}_L = \beta_0 + \widetilde{\Sigma}_{1,L}^{-1}\widetilde{\rho}_L,$$

provided that $\widetilde{\Sigma}_{1,L}$ is non-singular.
It can be shown that

$$\widetilde{\Sigma}_{1,L} \xrightarrow{\text{p}} \sum_{m=1}^{M} \text{Var}\left(X|I_X = K_m\right) \times p_m = \Sigma_1,$$

hence Σ_1 is nonsingular if

Assumption 6. *For at least one* m, $\text{Var}(X|I_X = K_m)$ *is nonsingular*.

Thus, under this condition $\widetilde{\Sigma}_{1,L}$ nonsingular with probability converging to 1 as $L \to \infty$.

Moreover, it follows easily from

$$\widetilde{\rho}_L = \frac{1}{L}\sum_{k=1}^{L}\sum_{m=1}^{M}\frac{1}{K_m}\sum_{i=1}^{K_m}\left(\Omega(U_{i,k}|K_m) - \int_0^1 \Omega(u|K_m)\mathrm{d}u\right)$$
$$\times(X_k - \overline{X}_{m,L})\mathbf{1}\left(I_k = K_m\right)$$

that

$$\widetilde{\rho}_L \xrightarrow{\text{p}} 0,$$

hence the estimator

$$\widetilde{\beta}_L = \widetilde{\Sigma}_{1,L}^{-1}\widetilde{\eta}_L = \beta_0 + \widetilde{\Sigma}_{1,L}^{-1}\widetilde{\rho}_L$$

is consistent, i.e.,

$$\widetilde{\beta}_L = \widetilde{\Sigma}_{1,L}^{-1}\widetilde{\eta}_L = \beta_0 + \widetilde{\Sigma}_{1,L}^{-1}\widetilde{\rho}_L \xrightarrow{\text{p}} \beta_0$$

as $L \to \infty$.

Furthermore, it can be shown that

$$
\begin{aligned}
\sqrt{L}\widetilde{\rho}_L &= \frac{1}{\sqrt{L}} \sum_{k=1}^{L} \sum_{m=1}^{M} \frac{1}{K_m} \sum_{i=1}^{K_m} \left(\Omega(U_{i,k}|K_m) - \int_0^1 \Omega(u|K_m)\mathrm{d}u \right) \\
&\quad \times (X_k - \overline{X}_{m,L}) \mathbf{1}\left(I_k = K_m\right) \\
&= \frac{1}{\sqrt{L}} \sum_{k=1}^{L} \sum_{m=1}^{M} \frac{1}{K_m} \sum_{i=1}^{K_m} \left(\Omega(U_{i,k}|K_m) - \int_0^1 \Omega(u|K_m)\mathrm{d}u \right) \\
&\quad \times \left(X_k - E\left[X|I_X = K_m\right]\right) \mathbf{1}\left(I_k = K_m\right) + O_p(1/\sqrt{L}) \\
&= \frac{1}{\sqrt{L}} \sum_{k=1}^{L} Z_k + O_p(1/\sqrt{L}), \text{ say,}
\end{aligned}
$$

where

$$
\begin{aligned}
Z_k &= \sum_{m=1}^{M} \frac{1}{K_m} \sum_{i=1}^{K_m} \left(\Omega(U_{i,k}|K_m) - \int_0^1 \Omega(u|K_m)\mathrm{d}u \right) \\
&\quad \times \left(X_k - E\left[X|I_X = K_m\right]\right) \mathbf{1}\left(I_k = K_m\right).
\end{aligned}
$$

These Z_k's are i.i.d., with $E[Z_k] = 0, \ E[Z_k Z_k'] = \Sigma_2,$

where

$$\Sigma_2 = \sum_{m=1}^{M} \frac{p_m}{K_m} \left(\int_0^1 (\Omega(u|K_m))^2 \, \mathrm{d}u - \left(\int_0^1 \Omega(u|K_m)\mathrm{d}u \right)^2 \right) \times \mathrm{Var}\left(X|I_X = K_m\right),$$

which is also nonsingular under Assumption 6.

Consequently, by the standard central limit theorem,

$$\sqrt{L}\widetilde{\rho}_L = \frac{1}{\sqrt{L}} \sum_{k=1}^{L} Z_k + O_p(1/\sqrt{L}) \xrightarrow{\mathrm{d}} N_p\left(0, \Sigma_2\right)$$

and thus,

$$\sqrt{L}\left(\widetilde{\beta}_L - \beta_0\right) = \widetilde{\Sigma}_{1,L}^{-1}\sqrt{L}\widetilde{\rho}_L \xrightarrow{\mathrm{d}} N_p\left(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right).$$

A consistent estimator of the asymptotic variance matrix

In order to use the result

$$\sqrt{L}\left(\widetilde{\beta}_L - \beta_0\right) \xrightarrow{\text{d}} N_p\left(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right).$$

for inference on β_0 we need a consistent estimator of $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$. Note that we already have a consistent estimator of Σ_1, namely

$$\widetilde{\Sigma}_{1,L} = \sum_{m=1}^{M}\frac{1}{L}\sum_{k=1}^{L}(X_k - \overline{X}_{m,L})(X_k - \overline{X}_{m,L})'\mathbf{1}\left(I_k = K_m\right),$$

and since Σ_1 is nonsingular we have

$$\widetilde{\Sigma}_{1,L}^{-1} \xrightarrow{\text{p}} \Sigma_1^{-1}.$$

Thus, it remains to derive a consistent estimator of

$$\Sigma_2 = \sum_{m=1}^{M}\frac{p_m}{K_m}\left(\int_0^1 \left(\Omega(u|K_m)\right)^2 \text{d}u - \left(\int_0^1 \Omega(u|K_m)\text{d}u\right)^2\right) \times \text{Var}\left(X|I_X = K_m\right),$$

as follows.

Denote

$$\widetilde{\Gamma}_{L,m} = \frac{1}{L}\sum_{k=1}^{L}(X_k - \overline{X}_{m,L})(X_k - \overline{X}_{m,L})'\mathbf{1}\left(I_k = K_m\right).$$

It can be shown that

$$\widetilde{\Gamma}_{L,m} = \mathrm{Var}\left(X|I_X = K_m\right) \times p_m + o_p(1),$$

where the $o_p(1)$ term is uniform in m.

Thus,

$$\Sigma_2 = \sum_{m=1}^{M}\frac{1}{K_m}\left(\int_0^1 \left(\Omega(u|K_m)\right)^2 \mathrm{d}u - \left(\int_0^1 \Omega(u|K_m)\mathrm{d}u\right)^2\right) \times\widetilde{\Gamma}_{L,m} + o_p(1)$$

So it remains to derive a consistent estimator $\widetilde{\sigma}^2_{L,m}$ of

$$\sigma^2_m = \int_0^1 \left(\Omega(u|K_m)\right)^2 \mathrm{d}u - \left(\int_0^1 \Omega(u|K_m)\mathrm{d}u\right)^2$$

as then $\widetilde{\Sigma}_{2,L} = \sum_{m=1}^{M}\frac{1}{K_m}\widetilde{\sigma}^2_{L,m}\widetilde{\Gamma}_{L,m} \xrightarrow{\mathrm{p}} \Sigma_2$.

Without loss of generality we may assume that in each auction the bids are sorted in random order.

Then

$(b_{1,k} - b_{2,k}) \mathbf{1} (I_k = K_m) = (\Omega(U_{1,k}|K_m) - \Omega(U_{2,k}|K_m)) \mathbf{1} (I_k = K_m)$

where the $U_{1,k}$'s and $U_{2,k}$'s are independent uniformly $(0, 1)$ distributed.

Hence by the law of large numbers,

$$\begin{aligned}
&\frac{1}{L}\sum_{k=1}^{L} (b_{1,k} - b_{2,k})^2 \mathbf{1} (I_k = K_m) \\
&\quad = \frac{1}{L}\sum_{k=1}^{L} (\Omega(U_{1,k}|K_m) - \Omega(U_{2,k}|K_m))^2 \mathbf{1} (I_k = K_m) \\
&\quad = E\left[(\Omega(U_{1,k}|K_m) - \Omega(U_{2,k}|K_m))^2\right] p_m + o_p(1) \\
&\quad = 2\left(\int_0^1 (\Omega(u|K_m))^2 \mathrm{d}u - \left(\int_0^1 \Omega(u|K_m)\mathrm{d}u\right)^2\right) p_m + o_p(1) \\
&\quad = 2\sigma_m^2 p_m + o_p(1).
\end{aligned}$$

Therefore,

$$\widetilde{\sigma}^2_{L,m} = \frac{\frac{1}{L}\sum_{k=1}^{L}\left(b_{1,k} - b_{2,k}\right)^2 \mathbf{1}\left(I_k = K_m\right)}{2.\widehat{p}_{m,L}} \xrightarrow{\text{p}} \sigma^2_m,$$

hence,

$$\widetilde{\Sigma}_{2,L} = \sum_{m=1}^{M} \frac{1}{K_m} \widetilde{\sigma}^2_{L,m} \widetilde{\Gamma}_{L,m} \xrightarrow{\text{p}} \Sigma_2,$$

where

$$\widetilde{\Gamma}_{L,m} = \frac{1}{L} \sum_{k=1}^{L} (X_k - \overline{X}_{m,L})(X_k - \overline{X}_{m,L})' \mathbf{1}\left(I_k = K_m\right).$$

A numerical experiment

In order to see how well this estimation procedure works, we have generated two sets of auctions, for $L = 100$ and $L = 500$, as follows.

- In each auction k the vector
$$X_k = (X_{1,k}, X_{2,k}, X_{3,k}, X_{4,k}, X_{5,k})'$$
of auction-specific covariates have been generated by drawing each $X_{i,k}$ from the standard normal distribution, conditional on $|X_{i,k}| < 1$.

- The corresponding number of bidders I_k has been generated as $I_k = 2 + R(X_{1,k})$, where $R(X_{1,k})$ is a random drawing from the Binomial-Logit$(3, p(X_{1,k}))$ distribution, with $p(x) = 1/(1 + \exp(-x))$. Thus, M and K_m in Assumption 5 are
$$M = 4,\ K_m = 1 + m,\ \ m = 1, 2, 3, 4.$$

- The parameter vector involved has been chosen as

$$\beta_0 = (1, 0, 0, 0, 0)'.$$

- The distribution function F_0 in Assumption 2 has been chosen as $F_0(w) = 1 - \exp(-w^2)$, with density $f_0(w) = 2w\exp(-w^2)$ and inverse

$$F_0^{-1}(u) = \sqrt{\ln\left(1/(1-u)\right)}.$$

In this case $\int_0^\infty w f_0(w)\mathrm{d}w = \sqrt{\pi}/2$, and it is not hard to verify that $\int_0^\infty (\ln(w))^2 f_0(w)\mathrm{d}w < \infty$.

- The bids $B_{i,k}$, $i = 1, 2, ..., I_k$, in auction k are generated as

$$B_{i,k} = \exp(X_{1,k}) \int_0^1 (I_k - 1) z^{I_k - 2} F_0^{-1}(U_{i,k}.z)\mathrm{d}z,$$

where the $U_{i,k}$'s are independent random drawings from the uniform $(0, 1)$ distribution, and the integrals involved have been computed numerically, via a Gaussian quadrature approach.

The estimation results are presented in Table 1.

Table 1. Estimation results for β_0

	$L = 100$		$L = 500$	
β_0	$\widetilde{\beta}_L$	*t-value*	$\widetilde{\beta}_L$	*t-value*
1	1.054020941	15.925	1.004394570	33.158
0	0.006968898	0.103	−0.000557115	−0.020
0	0.113922071	1.831	0.012743293	0.465
0	−0.004020778	−0.058	0.020242694	0.746
0	0.039059912	0.602	−0.027952404	−1.039

Clearly, the estimator $\widetilde{\beta}_L$ does a good job in this particular case.

Semi-nonparametric modeling of the bid function

Every absolutely continuous distribution function $F(w)$ on $[0, \infty)$ with continuous density $f(w)$ satisfying $f(w) > 0$ on $(0, \infty)$ can be written as

$$F(w) = H(G(w)),$$

where

- $G(w)$ is an **a priori** chosen absolutely continuous distribution function on $[0, \infty)$ with continuous density $g(w)$ satisfying $g(w) > 0$ on $(0, \infty)$ and inverse $G^{-1}(u)$, $u \in [0, 1]$,
- $H(u) = F(G^{-1}(u))$ is an absolutely continuous distribution function on $[0, 1]$ with density

$$h(u) = f(G^{-1}(u)) \frac{\mathrm{d}G^{-1}(u)}{\mathrm{d}u} = \frac{f(G^{-1}(u))}{g(G^{-1}(u))},$$

which is continuous and positive valued on $(0, 1)$.

Consequently, the density f of F can be written as

$$f(w) = h(G(w))g(w),\ w > 0,$$

and the inverse of F can be written as

$$F^{-1}(u) = G^{-1}(H^{-1}(u)),\ u \in [0, 1].$$

Then the true bid function can be written as

$$\begin{aligned}\bar{b}_0(u|X^+) &= \exp(\beta_0' X) \int_0^1 (I_X - 1) z^{I_X - 2} F_0^{-1}(z.u) \mathrm{d}z \\ &= \exp(\beta_0' X) \int_0^1 (I_X - 1) z^{I_X - 2} G^{-1}(H_0^{-1}(z.u)) \mathrm{d}z\end{aligned}$$

where $H_0(u) = F_0(G^{-1}(u))$.

For example, let G be standard exponential distribution function $G(w) = 1 - \exp(-w)$, which has inverse

$$G^{-1}(u) = \ln(1/(1 - u)).$$

Then

$$\bar{b}_0(u|X^+) = \exp(\beta_0' X) \int_0^1 (I_X - 1) z^{I_X - 2} \ln(1/(1 - H_0^{-1}(z.u))) \mathrm{d}z.$$

The reason for the transformations

$$F_0(w) = H_0(G(w)),\ f_0(w) = h_0(G(w))g(w)$$

is that there exists convenient series approximations of h_0 and H_0.

In particular, denote

$$\begin{aligned}
\delta_{0,m} &= \frac{\int_0^1 \sqrt{h_0(u)}\sqrt{2}\cos(m\pi u)\mathrm{d}u}{\int_0^1 \sqrt{h_0(u)}\mathrm{d}u},\ m \in N, \\
&\left(\text{which satisfy } \sum_{m=1}^{\infty} \delta_{0,m}^2 < \infty\right) \\
\boldsymbol{\delta}^0 &= \{\delta_{0,m}\}_{m=1}^{\infty}, \\
h(u|\pi_n\boldsymbol{\delta}^0) &= \frac{\left(1 + \sum_{m=1}^{n} \delta_{0,m}\sqrt{2}\cos(m\pi u)\right)^2}{1 + \sum_{m=1}^{n} \delta_{0,m}^2},
\end{aligned}$$

where π_n is the truncation operator, i.e.,

π_n applied to $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$ as $\pi_n\boldsymbol{\delta}$ replaces all the δ_m's for $m > n$ by zeros.

Then

$$\lim_{n\to\infty}\int_0^1 \left|h(u|\pi_n\boldsymbol{\delta}^0) - h_0(u)\right| \mathrm{d}u = 0.$$

Moreover, the c.d.f. $H(u|\pi_n\boldsymbol{\delta}^0) = \int_0^u h(z|\pi_n\boldsymbol{\delta}^0)\mathrm{d}z$ satisfies

$$\lim_{n\to\infty}\sup_{0\le u\le 1}\left|H(u|\pi_n\boldsymbol{\delta}^0) - H_0(u)\right| = 0,$$

hence

$$H_0(u) \equiv H(u|\boldsymbol{\delta}^0),$$

where $H(u|\pi_n\boldsymbol{\delta}^0)$ has the closed form expression

$$\begin{aligned} H(u|\pi_n\boldsymbol{\delta}^0) &= u \\ &+ \frac{1}{1+\sum_{m=1}^n \delta_{0,m}^2}\left[2\sqrt{2}\sum_{k=1}^n \delta_{0,k}\frac{\sin\left(k\pi u\right)}{k\pi} + \sum_{m=1}^n \delta_{0,m}^2\frac{\sin\left(2m\pi u\right)}{2m\pi}\right. \\ &\qquad + 2\sum_{k=2}^n\sum_{m=1}^{k-1}\delta_{0,k}\delta_{0,m}\frac{\sin\left(\left(k+m\right)\pi u\right)}{\left(k+m\right)\pi} \\ &\qquad \left. + 2\sum_{k=2}^n\sum_{m=1}^{k-1}\delta_{0,k}\delta_{0,m}\frac{\sin\left(\left(k-m\right)\pi u\right)}{\left(k-m\right)\pi}\right]. \end{aligned}$$

The true bid function can now be written as

$$\bar{b}_0\left(u|X^+,\beta_0,\boldsymbol{\delta}^0\right)=\exp(\beta_0'X).\Lambda(u|I_X,\boldsymbol{\delta}^0),$$

with SNP version

$$\bar{b}_0\left(u|X^+,\beta,\pi_n\boldsymbol{\delta}\right)=\exp(\beta'X)\Lambda(u|I_X,\pi_n\boldsymbol{\delta}),$$

where for $\boldsymbol{\delta}=\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2<\infty$,

$$\Lambda(u|K,\boldsymbol{\delta})=\int_0^1(K-1)z^{K-2}G^{-1}(H^{-1}(u.z|\boldsymbol{\delta}))\mathrm{d}z,\ K\geq 2,$$

with $H^{-1}(u|\boldsymbol{\delta})$ the inverse of $H(u|\boldsymbol{\delta})$.

Integrated squared simulated Laplace moments

For the case $\beta_0 = \beta = 0$ and I_X is a.s. constant, $\boldsymbol{\delta}^0$ can be consistently estimated by the integrated squared simulated moments sieve estimation approach in Bierens and Song (2012). In the present case this approach can be adapted as follows. For each bid $B_{i,k}$, $i = 1, 2, ..., I_k$, denote

$$\begin{aligned} Y_{i,k} &= \exp(-\widetilde{\beta}'_L X_k) B_{i,k} \\ &= \exp\left(-\left(\widetilde{\beta}_L - \beta_0\right)' X_k\right) \exp(-\beta'_0 X_k) B_{i,k} \end{aligned}$$

and recall that

$$\exp(-\beta'_0 X_k) B_{i,k} \sim \Lambda(U^0_{i,k} | I_k, \boldsymbol{\delta}^0).$$

with

$$\Lambda(u|K, \boldsymbol{\delta}) = \int_0^1 (K-1) z^{K-2} G^{-1}(H^{-1}(u.z|\boldsymbol{\delta})) \mathrm{d}z$$

for $K \geq 2$, where the $U^0_{i,k}$'s are random drawings from the uniform $(0, 1)$ distributed.

Next, denote

$$\widetilde{Y}_{i,k}(\boldsymbol{\delta}) = \Lambda(\widetilde{U}_{i,k}|I_k, \boldsymbol{\delta}),$$

where the $\widetilde{U}_{i,k}$'s are drawn independently from the uniform $(0,1)$ distribution.

Recall that for $K \geq 2$,

$$\Lambda(u|K, \boldsymbol{\delta}) = \int_0^1 (K-1)z^{K-2}G^{-1}(H^{-1}(u.z|\boldsymbol{\delta}))\mathrm{d}z$$

Moreover, denote for $m = 1, 2, 3, ..., M$ and $t \in (0, \infty)$,

$$\varphi_{m,L}(t) = \frac{\frac{1}{L}\sum_{k=1}^{L}\left(\frac{1}{K_m}\sum_{i=1}^{K_m}\exp\left(-t.Y_{i,k}\right)\right)\mathbf{1}\left(I_k = K_m\right)}{\frac{1}{L}\sum_{k=1}^{L}\mathbf{1}\left(I_k = K_m\right)},$$

$$\widehat{\varphi}_{m,L}(t|\boldsymbol{\delta}) = \frac{\frac{1}{L}\sum_{k=1}^{L}\left(\frac{1}{K_m}\sum_{i=1}^{K_m}\exp\left(-t.\widetilde{Y}_{i,k}(\boldsymbol{\delta})\right)\right)\mathbf{1}\left(I_k = K_m\right)}{\frac{1}{L}\sum_{k=1}^{L}\mathbf{1}\left(I_k = K_m\right)}.$$

Assuming that

Assumption 7. *The auction-specific covariates X_k have bounded support,*

it is a standard exercise to verify that under Assumptions 1-7,

$$p\lim_{L\to\infty}\sup_{0\leq t\leq\tau}\left|\varphi_{m,L}(t)-\varphi_m(t|\boldsymbol{\delta}^0)\right|=0, \text{ where}$$

$$\varphi_m(t|\boldsymbol{\delta}^0)=\int_0^1 \exp\left(-t.\Lambda(u|K_m,\boldsymbol{\delta}^0)\right)\mathrm{d}u,$$

for any $\tau\in(0,\infty)$, whereas under Assumptions 1-6,

$$p\lim_{L\to\infty}\sup_{0\leq t\leq\tau}\left|\widehat{\varphi}_{m,L}(t|\boldsymbol{\delta})-\varphi_m(t|\boldsymbol{\delta})\right|=0, \text{ where}$$

$$\varphi_m(t|\boldsymbol{\delta})=\int_0^1 \exp\left(-t.\Lambda(u|K_m,\boldsymbol{\delta})\right)\mathrm{d}u,$$

pointwise in $\boldsymbol{\delta}$.

Note that both $\varphi_m(t|\boldsymbol{\delta}^0)$ and $\varphi_m(t|\boldsymbol{\delta})$ are Laplace transforms of non-negative random variables.

As is well-known, two Laplace transforms are equal on $(0, \infty)$ if and only if their distributions are equal.
More generally, the following result holds.

Lemma. *Let V_1 and V_2 be nonnegative random variables with Laplace transforms $\mathcal{L}_1(t) = E[\exp(-t.V_1)]$ and $\mathcal{L}_2(t) = E[\exp(-t.V_2)]$, respectively, where $t \geq 0$. Suppose that on an open interval $T \subset (0, \infty)$, $\mathcal{L}_1(t) = \mathcal{L}_2(t)$ for all $t \in T$. Then V_1 and V_2 have the same distribution.*

Choosing $T = (0, 1)$, this lemma implies that if

$$\varphi_m(t|\boldsymbol{\delta}^0) = \varphi_m(t|\boldsymbol{\delta}) \text{ for all } t \in (0, 1),$$

then for two mutually independent uniformly $(0, 1)$ distributed random variables U_1 and U_2,

$$\Lambda(U_1|K_m, \boldsymbol{\delta}^0) \sim \Lambda(U_2|K_m, \boldsymbol{\delta}),$$

which in its turn implies that $\boldsymbol{\delta} = \boldsymbol{\delta}^0$.

These results suggests to estimate $\boldsymbol{\delta}^0$ by minimizing the integrated squared simulated Laplace moments (ISSLM) objective function

$$\widehat{Q}_L(\boldsymbol{\delta}) = \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(\widehat{\varphi}_{m,L}(t|\boldsymbol{\delta}) - \varphi_{m,L}(t)\right)^2 \mathrm{d}t$$

via a sieve estimation approach, similar to Bierens and Song (2012).

Note that

$$p\lim_{L\to\infty} \widehat{Q}_L(\boldsymbol{\delta}) = Q(\boldsymbol{\delta}),$$

pointwise in $\boldsymbol{\delta}$, where

$$Q(\boldsymbol{\delta}) = \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(\varphi_m(t|\boldsymbol{\delta}) - \varphi_m(t|\boldsymbol{\delta}^0)\right)^2 \mathrm{d}t,$$

so that $Q(\boldsymbol{\delta})$ takes a (unique) minimum zero at $\boldsymbol{\delta} = \boldsymbol{\delta}^0$.

The infinite-dimensional parameter space involved is

$$\Delta = \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}.$$

A natural norm on Δ is the pseudo-Euclidean norm

$$||\boldsymbol{\delta}|| = ||\{\delta_m\}_{m=1}^{\infty}|| = \sqrt{\sum_{m=1}^{\infty} \delta_m^2}$$

with associated metric $||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||$, so that Δ becomes a metric space.

However,

$$\widehat{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\delta} \in \Delta} \widehat{Q}_L(\boldsymbol{\delta})$$

is not uniquely defined, and in general none of these solutions are consistent.

The solution to this problem is sieve estimation, as follows.

Sieve estimation and consistency

For $n \in \mathbb{N}$, let

$$\Delta_n = \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{n} \delta_m^2 \leq B_n,\ \delta_m = 0 \text{ for } m > n \right\}$$

where B_n is an a priori chosen monotonic increasing sequence of positive constants converging (slowly) to ∞ for $n \to \infty$.

Note that Δ_n is compact because it is isomorph to the hyperball $||x||^2 \leq B_n$ in $\mathbb{R}^n$ which is closed and bounded and therefore compact.

Moreover, $\Delta_n \subset \Delta_{n+1}$, and $\Delta = \overline{\cup_{n=1}^{\infty} \Delta_n}$, where the bar denotes the closure.

The sequence Δ_n is called the "sieve", and each Δ_n is called a sieve space.

Now the sieve estimator of $\boldsymbol{\delta}^0$ is

$$\widehat{\boldsymbol{\delta}}_{n_L} = \arg \min_{\boldsymbol{\delta} \in \Delta_{n_L}} \widehat{Q}_L(\boldsymbol{\delta}),$$

where n_L is an arbitrary subsequence of the sample size L satisfying

$$\lim_{L \to \infty} n_L = \infty, \ \lim_{L \to \infty} n_L / L = 0.$$

Then it can be shown that

Theorem. *Under Assumptions 1-7 the sieve estimator* $\widehat{\boldsymbol{\delta}}_{n_L}$ *is consistent*:

$$p \lim_{L \to \infty} ||\widehat{\boldsymbol{\delta}}_{n_L} - \boldsymbol{\delta}^0|| = 0.$$

Remark. This result is typical for our SNP auction model, and may not be applicable to other SNP models, because its proof employs the facts that

$$0 \leq \widehat{Q}_L(\boldsymbol{\delta}) \leq 1, \ 0 \leq Q(\boldsymbol{\delta}) \leq 1, \ Q(\boldsymbol{\delta}^0) = 0,$$

which in general do not hold for other SNP models.

Finally, recall that

$$f_0(w) = h_0(G(w))g(w)$$

and

$$\lim_{n\to\infty} \int_0^1 \left| h(u|\pi_n \boldsymbol{\delta}^0) - h_0(u) \right| \mathrm{d}u = 0.$$

It can be shown that $p \lim_{L\to\infty} ||\widehat{\boldsymbol{\delta}}_{n_L} - \boldsymbol{\delta}^0|| = 0$ implies

$$p \lim_{L\to\infty} \int_0^1 \left| h(u|\widehat{\boldsymbol{\delta}}_{n_L}) - h(u|\pi_{n_L} \boldsymbol{\delta}^0) \right| \mathrm{d}u = 0.$$

Hence, denoting

$$\widehat{f}_L(w) = h(G(w)|\widehat{\boldsymbol{\delta}}_{n_L})g(w),\ \widehat{F}_L(w) = H(G(w)|\widehat{\boldsymbol{\delta}}_{n_L})$$

we have

$$\int_0^\infty \left| \widehat{f}_L(w) - f_0(w) \right| \mathrm{d}w = \int_0^1 \left| h(u|\widehat{\boldsymbol{\delta}}_{n_L}) - h_0(u) \right| \mathrm{d}u \xrightarrow{\text{p}} 0,$$

$$\sup_{w>0} \left| \widehat{F}_L(w) - F_0(w) \right| \xrightarrow{\text{p}} 0$$

as $L \to \infty$.

Concluding remarks

The paper involved is incomplete.

- My coauthor, Hosin Song, is in the process of implementing the sieve estimation approach, so that we can add some numerical examples to the paper, in the form of graphs, similar to our previous paper Bierens and Song (2012). The first preliminary results look promising.
- Also, we are planning to add an empirical application, provided that we can find a suitable data set.