Auction Empirics, Collusion and Bidding Rings, Part IV: Detection of Collusion in Auction Settings

John Asker*

February 16, 2015

1 Preliminaries

The objective in these lectures is to look a some of the more classic papers that are aimed at detecting bid rigging of some form or another (or at least those that I see being cited most often). It is not a complete list, but rather is designed to give you a sense of the types of approaches that are floating about in the literature.

There are three papers I want to examine, in varying degrees of detail:

- Porter and Zona (1993), Detection of Bid Rigging in Procurement Auctions, *J.P.E.* 101(3) 518
- Bajari and Yi (2003), Deciding between Competition and Collusion, *ReStat* 85(4) 971
- Athey, Levin, Seira (2011), Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions, *Quarterly Journal of Economics* 126, 207

The first two might reasonably be thought of as reduced form (at least as far as the detection parts). The last is an example of a structural approach.

Before delving into details of the papers, it's worth being upfront about my attitude to the detection literature: mostly, I feel people overstate it's potential usefulness. That is, it can be useful, but in somewhat more specific settings than folks normally claim. In this sense I am in full agreement with the following sentiment, from the introduction to Porter and Zona:

*Email: johnasker@econ.ucla.edu, www.johnasker.com

This paper proposes econometric test procedures that are designed to detect the presence of bid rigging in procurement auctions.[1] Our tests will be poor substitutes for a wiretap or a disclosure by a dissident ring member. However, our procedures may be preferable to the view that patterns of bid rotation, or relatively constant market shares, constitute irrefutable evidence of collusion. Rotating bids are consistent with competitive equilibria when there are decreasing returns to scale, such as when there are capacity constraints, as Zona (1986) demonstrates. Lang and Rosenthal (1991) show that the non-cooperative mixed-strategy equilibrium of a multiproject bidding game, in which firms simultaneously compete for several contracts, may entail negative correlation between a firm's bids, or an apparent bid rotation pattern. Similarly, comparisons of winning bids and engineers' estimates of costs, which attempt to measure economic returns, may be unreliable. Engineers' estimates may be unduly influenced by historical bid patterns and so may be an inflated measure of true costs.

In general, finding a single test procedure to detect bid rigging is an impossible goal. As in most tests for the exercise of market power, the idea is to identify differences between the observable implications of collusive and competitive behavior. The difficulty is that both competitive and collusive equilibria depend, to a great extent, on the economic environment, such as the auction rules and the nature of the good being traded. As Hendricks and Porter (1989) argue, collu-

So, with that preliminary comment, lets examine Porter and Zona

2 Porter and Zona (1993), Detection of Bid Rigging in Procurement Auctions, *J.P.E.* 101(3) 518

The paper has three types of section:

1. Institutional detail
2. model and resulting reduced form empirics
3. data analysis

Lets go through each in turn.

2.1 Institutional Setup

- NY DOT procures roadwork (specifically paving jobs)
- Details of what is needed on each job available from DOT for purchase
- 'Plan Buyers List' can be purchased from DOT - also distributed to subscribers. This lists all firms buying plans on each job offered
- FPSB auctions - lowest price wins
- After auction, DOT announces all bids and who won contract.
- $120 million in Nassau and Suffock county between 1979 and 1985 contracted by DOT. 186 Contracts in total. 116 paving contracts, 575 bids on these paving contracts. 75 contracts had 2 or more bids. A small number of bidders dominate the bidding, especially as projects get bigger, so that competition for jobs if really among a fairly small set of firms.
- 1 large firm is prosecuted on a project predating the data. The proscution occrs in 1984. 4 other firms are named in the indictment, but not prosected. These five firms are "CARTEL" firms. They account for 54% of bids on large projects.
- Hence the research question is something like "Was this an isolated event?"
- Why might systemic collusion be plausible in this market?
    1. competition is only on price
    2. DOT annouces a lot of info

3. DOT does not appear at all strategic (i.e. almost never rejects a winning bid)
4. Set of dominant firms appears small
5. suggestion of high barriers to entry. Usual type of capital equipment stroy but also this interesting story about the labor unions:

   > [5] According to an article published in *Newsday,* insiders say that leaders of the two most powerful construction unions on Long Island discouraged outside bidders by threatening future labor trouble. At least one of the unions covered the entire metropolitan area. Thus when word spread that particular firms had "the inside track on large public bids on the Island and the unions wanted it that way . . . very few people missed the message" (November 18, 1984, pp. 3, 30).

6. bidders active in local trade associations (also I vaguely recall that there was a mafia angle)
7. RFQ's are somewhat sequential

- So the suspected conduct is that they coordinate on a low bid and the others in the cartel submitt phantom bids. See below:

  > In addition to the evidence from the bidding data, information obtained from insiders suggests that a phantom bidding scheme was employed by these firms. These sources describe the process. "We all sat at the conference table . . . one of the contractors would have a list of upcoming contracts . . . they'd talk about the contract . . . how much money . . . who won the last one . . . who should get this one . . . . The contractors who were tagged to be the low bidders would work out their bid figures . . . . The rest of the contractors would then come up with higher bids" (*Newsday* [November 18, 1984],

2.2 Model

The model part is pretty simple, but also very neat. As usual they posit the standard FPSB bidding function

$$Pr(win|b) + (b-c)\frac{\partial Pr(win|b)}{\partial b} = 0$$

Then they posit that bidding as a function of observed covariates is (i indicates bidder, t is auction)

$$log(b_{it}) = \alpha_t + \beta X_{it} + \epsilon_{it}$$

where ϵ_{it} is interpreted as private information, with expectation zero and variance σ_t^2. Notice how close this is to how the structural auction literature tends to handle observed auction heterogeneity.[1]

Estimation is via GLS (i.e. conditional homoskedasticity (more precisely the spherical error assumption) is violated since the variance is auction specific). This ends model number 1.

The second model is what this paper is known for. The idea is to look at the ranking of bids and see if the ranking can be predicted and if the patterns appear stable across subsets of the data. The idea is that this should be so if bids reflects actual economic factors (e.g. distance to job), that is, if they are competitive. If bids are phantom, then this need not be true.

Let ϵ_{it} be distributed IID Type 1 extreme value distribution. This gives a logit structure. Hence

$$lnPr(b_{it} < b_{jt} \forall j \neq i) = \alpha_t + \beta Z_{it}$$

($Z_i t$ is an adjustment of the X's by the inferred variance of the ϵ_{it} from the GLS estimator above). The more familiar expression is:

$$Pr(b_{it} < b_{jt} \forall j \neq i) = \frac{e^{\beta Z_{it}}}{\sum_j e^{\beta Z_{jt}}}$$

This allows us to construct the likelihood of any sequence of ranks, just by multiplying the probability of the each bid being lower than the ones above it. Then for all actions, we multiply again. So (with a little abuse of notation) the likelihood for the data is

$$L(\beta) = \prod_t \prod_i \frac{e^{\beta Z_{it}}}{\sum_{j>i} e^{\beta Z_{jt}}}$$

Note that this can be estimated using any subset of the data. That is, independence is doing a bunch of work, but is very helpful (what if this were a environment where dynamics were important, but not well observed by the econometrician?)

The idea will be to compare estimates from subsets of data, and see if they are consistent. What is done, is to look at the model for the low bidders, and then compare it to the model for all other other rankings. The test is done using a likelihood ratio test due to Hausman and Ruud 1987.

2.3 Data Analysis

Variables are:

- BACKLOG: sum of dollars of jobs contracted for in last 3 months but not completed

[1] Also, note that unobserved auction heterogeneity is largely ignored, although a random effects estimator could have been used. That said, it would have made the rank order stuff (to follow) a lot harder to merge in.

- CAP: capacity measured as the max of BACKLOG observed in sample period for that firm
- UTIL: utilization rate which is BACKLOG/CAP
- UTILSQ: this is the square of UTIL
- ISLAND: = 1 if headquartered on long island
- NOBACK: =1 if never won a project

The first regression is the GLS of bid on stuff

TABLE 4

GLS ESTIMATES

	Data from All Firms (1)	Data from Competitive Firms (2)	Data from Cartel Firms (3)
Observations	476	319	157
Degrees of freedom	395	238	81
Wald statistic	21.9	494.7	28.4
UTIL	−.0053	−.0973	.1991
	(.2)	(2.8)	(1.2)
UTILSQ	.0358	.1720	−.1143
	(1.0)	(4.0)	(.8)
NOBACK	−.0010	−.0178	
	(.1)	(1.6)	
CAP	.1666	−1.2691	1.8225
	(1.8)	(10.4)	(4.6)
CAPSQ	−.4430	4.8519	−2.9029
	(2.1)	(13.0)	(4.4)
ISLAND	−.0288	−.0334	
	(.6)	(1.2)	

NOTE.—Absolute values of t-statistics are displayed in parentheses. Auction-specific constants were included but are not reported to save space. The Wald statistics pertain to a test of the joint significance of the reported coefficients. The coefficients of CAP and CAPSQ are scaled up by 10^4 and 10^8, respectively.

This is a bit opaque so Porter and Zona give economic magnitudes. When UTIL = 1 (running at full capacity) bids go up by 7.5% if a firm is competitive, being on L.I. gives a bid advantage of about 3%. Cartel firms look different and turn out to be statistically different.

The next two regressions examine the ranks.

TABLE 5

COMPETITIVE RANK BASED ESTIMATES

	All Ranks (1)	Low Ranks (2)	Higher Ranks (3)
Observations	244	75	169
Log likelihood	−291.4	−89.85	−199.4
UTIL	−.0070	.0161	−.0552
	(.1)	(.1)	(.3)
UTILSQ	.0986	.0534	.1596
	(.8)	(.3)	(1.0)
NOBACK	−.0283	.0089	−.0454
	(1.0)	(.2)	(1.3)
CAP	−1.888	−1.641	−2.100
	(3.8)	(2.4)	(3.0)
CAPSQ	6.869	6.517	7.020
	(3.9)	(2.6)	(2.9)
ISLAND	−.0182	−.0759	.1016
	(.3)	(.9)	(.9)

NOTE.—Absolute values of t-statistics are displayed in parentheses. The coefficients of CAP and CAPSQ are scaled up by 10^4 and 10^8, respectively.

TABLE 6

CARTEL RANK BASED ESTIMATES

	All Ranks (1)	Low Ranks (2)	Higher Ranks (3)
Observations	85	50	35
Log likelihood	−73.97	−44.58	−24.92
UTIL	.0429	.2107	.2310
	(.3)	(1.0)	(.6)
UTILSQ	−.0112	−.1128	−.4300
	(.1)	(.6)	(.9)
CAP	.4306	1.101	−2.537
	(.9)	(1.3)	(1.6)
CAPSQ	−.8473	−1.904	3.861
	(.9)	(1.2)	(1.4)

NOTE.—Absolute values of t-statistics are displayed in parentheses. The coefficients of CAP and CAPSQ are scaled up by 10^4 and 10^8, respectively.

The punchline is that in the cartel setting the lower rank model is statistically different from the higher ranks. This is not true in the competitive setting.

The main take away from this paper is the idea of using the ranking of bids as a diagnostic, subject to the following caveats from Rob and Doug:

> as methodological as well as descriptive. Unfortunately, if an antitrust authority or procurement agency were to publicly announce the adoption of our test procedure, it would be relatively easy for an effective cartel to tailor its phantom bids to disguise collusive behavior. For example, all cartel firms could scale their competitive bids up by the same percentage. The bid ranking would then coincide with cost rankings. If the cartel was not inclusive, differences between cartel and competitive bidding would be consistent with cost asymmetries between the two groups of firms. Presumably, a noninclusive

3 Bajari and Yi (2003), Deciding between Competition and Collusion, *ReStat* 85(4) 971

Like most of Pat Bajari's papers this article has a lot of ideas in it. That said, it lacks a little focus, due (I suspect) to the fact that Pat was looking to collect a bunch of ideas that didn't make it into other projects and park them somewhere. As such this is not the easiest paper to pull a punchline from.

What seems to have gained traction from this paper is the idea of testing for exchangeability of bids, which we will see is pretty similar to the Porter Zona idea above. Here is the set up:

- Consider an asymmetric IPV auction with N bidders indexed by i
- Let $G_i(b|z_1, z_2, ..., z_n)$ be the distribution of i's equilibrium bid, where z_j are observable characteristics of bidder j. Think of costs as something like $c_i = \alpha + \hat{\beta} z_i + \epsilon_i$ where the ϵ is private information
- Exchangeability says that if $\pi(x)$ is a one-to-one mapping of $\{1, 2, ...N\}$ onto itself, then

  $$G_i(b|z_1, z_2, ..., z_n) = G_{\pi(i)}(b|z_{\pi(1)}, z_{\pi(2)}, ..., z_{\pi(n)})$$

  That is, identity of the bidder should not matter for bid distributions beyond observables. That is, if bidder 1 has the observables of bidder 2 and vice versa, the $G(.)$ function captures everything that changes as a result of this permutation.
- Operationally, this means that if we estimate a reduced form bids function such that

  $$b_i = \gamma + \beta X + \beta_j Z_j + \beta_k Z_k + \varepsilon$$

  then we should expect $\beta_j = \beta_k$ under the null of exchangeability. So this is what Bajari and Ye test.

The data are form seal coat road construction projects in MN, ND and SD 1994 -1998 which equals 495 contracts. There are 11 main firms and a bunch of fringe firms. The variables they see are:

- $\text{BID}_{i,t}$: Amount bid by firm i on project t.
- EST_t: Engineer's cost estimate for project t.
- $\text{DIST}_{i,t}$: Distance between the location of the firm and the project.
- $\text{LDIST}_{i,t}$: $\log(\text{DIST}_{i,t} + 1.0)$.
- $\text{CAP}_{i,t}$: Used capacity measure of firm i on project t.
- $\text{MAXP}_{i,t}$: Maximum percentage free capacity of all firms on project t, excluding i.
- $\text{MDIST}_{i,t}$: Minimum of distances of all firms on project t, excluding i.

- $\text{LMDIST}_{i,t}$: $\log(\text{MDIST}_{i,t} + 1.0)$.
- $\text{CON}_{i,t}$: Proportion of work done (by dollar volume) by firm i in the state where project t is located prior to the auction.

And the summary stats are:

TABLE 5.—SUMMARY STATISTICS

Variable	No. Obs.	Mean	Std. Dev.	Min.	Max.
Winning bid	441	175,000	210,000	3893	1,732,500
Markup: [(winning bid) − estimate]/estimate	139	0.0031	0.1573	−0.3338	0.5421
Normalized bid: winning bid/estimate	139	1.0031	0.1573	0.6662	1.5421
Money on the table: (2nd bid) − (1st bid)	134	15,748	19,241	209	103,481
Normalized money on the table: [1st bid) − (2nd bid]/est.	134	0.0776	0.0888	0.0014	0.5099
Number of bidders	139	3.280	1.0357	1	6
Distance of winning firm	134	188.67	141.51	0	584.2
Distance of second lowest bidder	134	213.75	152.01	0	555
Capacity of winning bidder	131	0.3376	0.3160	0	0.9597
Capacity of second lowest bidder	131	0.4326	0.3435	0	1
All bids (normalized)	450	1.0819	0.1837	0.6662	1.8347
Distances (LDIST)	450	4.9315	1.1299	0.0000	6.4593
Capacities (CAP)	450	0.4172	0.3573	0.0000	1.0000
Maximal capacities among rivals (MAXP)	450	0.7915	0.3048	0.0000	1.0000
Minimal distance among rivals (LMDIST)	450	4.5679	1.3081	0.0000	9.2104
Job Concentration (CON)	450	0.5967	0.3601	0.0000	1.0000

The regression equations are:

$$\frac{\text{BID}_{i,t}}{\text{EST}_t} = \beta_{0,i} + \beta_{1,i}\,\text{LDIST}_{i,t} + \beta_{2,i}\,\text{CAP}_{i,t} + \beta_{3,i}\,\text{MAXP}_{i,t} + \beta_{4,i}\,\text{LMDIST}_{i,t} + \beta_5\,\text{CON}_{i,t} + \epsilon_{it}, \tag{28}$$

$$\frac{\text{BID}_{i,t}}{\text{EST}_t} = \alpha_0 + \alpha_1\,\text{LDIST}_{i,t} + \alpha_2\,\text{CAP}_{i,t} + \alpha_3\,\text{MAXP}_{i,t} + \alpha_4\,\text{LMDIST}_{i,t} + \alpha_5\,\text{CON}_{i,t} + \epsilon_{it}. \tag{29}$$

Note that auctions fixed effects are used here. It is worth thinking about whether this makes sense, given that the fixed effects will grow with the asymptotic. Depending on what is done with a regression like this, you need to be mindful of the incidental parameters problem.

The idea is to run an F-test testing the null of $\beta_j = \beta_k$ for the 11 main firms. The test statistic is

$$F = \frac{(SSR_c - SSR_u)/n}{SSR_u/(T-m)}$$

T is the number of Obs = 450, and m the number of regressors in the unconstrained model and n is the number of constraints. c is the constrained model (29) and u is the unconstrained model (28). The implementation looks at the whole set of data and also those pairs that compete against each other in at least 4 auctions.

TABLE 11.—EXCHANGEABILITY TEST

Firm Pair	n	m	F-Statistics	Upper Tail Area	Firm Pair	n	m	F-Statistics	Upper Tail Area
(1, 2)	4	194	1.2188	.3033	(4, 5)	4	194	1.0799	.3669
(2, 3)	4	194	2.1080	.0803	(4, 14)	4	194	0.9756	.4214
(2, 4)	4	194	1.0187	.3982	(5, 6)	4	194	1.2014	.3107
(2, 5)	4	194	3.9254	.0041	(5, 8)	4	194	1.2209	.3024
(2, 6)	4	194	0.7856	.5354	(5, 11)	4	194	0.2643	.9007
(2, 7)	4	194	2.3709	.0530	(5, 14)	4	194	2.3162	.0578
(2, 8)	4	194	0.6211	.6478	(5, 20)	4	194	1.2151	.3048
(2, 14)	4	194	2.1288	.0777	(6, 7)	4	194	2.2728	.0619
(2, 20)	4	194	1.6844	.1541	(6, 8)	4	194	0.1123	.9781
(3, 4)	4	194	1.8656	.1170	(7, 8)	4	194	2.0983	.0815
(3, 5)	4	194	1.5582	.1860	(14, 20)	4	194	1.1022	.3560
(3, 11)	4	194	1.1202	.3474	All pooled	40	158	1.4506	.0474

4 Athey, Levin, Seira (2011), Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions, *Quarterly Journal of Economics* 126, 207

This is a structural paper that examines the question of detection. It is worth noting from the outset that a closely related paper is Baldwin, Laura, Robert C. Marshall and Jean-Francois Richard (1997) Bidder Collusion at Forest Service Timber Sales, *Journal of Political Economy* 105(4) 657. To see the evolution in style in the structural approach it is worth reading the two papers side by side. That said, for time reasons I will just talk about the Athey et al paper.

The research questions: i) what is the revenue ranking of open and sealed auctions with asymmetry between bidders? and, ii) is there evidence of non-competitive bidding? (for my taste, these should have been two separate papers, as the composite is a bit unfocused, but that's hardly a reason to not like the paper)

4.1 The model

- consider a single tract as a unit of observation. Auction is either open outcry or FPSB. N potential bidders who much pay K to enter the auction. Learn their private value and the number of competing bidders when they enter. Number of participants conditional on entry is n

- private values are $v_i \sim F_i$

- there are two types of bidder: loggers and mills (mills are bigger and have manufacturing capacity). Indicate with subscript L and M respectively.

- The authors assume in some parts that F_M stochastically dominates F_L but not important for structural part

- equilibrium is a bidding function, b, and entry probability, p, for each type of bidder

- conditional on entry, bid functions are completely standard (e.g. open v = b and fpsb, v = b - markup)

- the entry part is a little more non-standard:

  - assume that bidders enter if the expected profit > 0, that is

$$\pi_i(p) - K > 0$$

where

$$\pi_i(p) = \sum_{n \in N} \text{profit given participants} \times \text{probability of participant mix}$$

The following gives uniqueness:

Proposition 1. Suppose that for all n_L, n_M, $\pi^s_M(n_L, n_M + 1) > \pi^s_L(n_L, n_M)$. Then there is a unique type-symmetric entry equilibrium for both auction formats. In equilibrium, either $p_L = 0$ or $p_M = 1$.

  - Note that in the structural estimation this is not assumed, but the condition on profits can be checked to see if appropriate to impose the uniqueness in the characterization. It's a neat bit of theory to cut through a bunch of thorny empirical modeling issues. The upshot is that in equilibrium, all mills enter and the loggers randomize.

- To model collusion, we assume that the mills collude and only do so in the open outcry auctions (this follows informal industry commetry).
- Further, the mills collude perfectly and frictionlessly (c.f. Asker 2010). This means that the highest value mill enters only. Without endogenous entry this would be fairly anodyne from an efficiency point of view. But with endogenous entry this will affect the returns to loggers and thus affect the loggers entry process.
- so maintained assumption is no collusion in FPSB auctions and that their may be in the open outcry auctions.

4.2 Data

The paper uses data from CA and ID/MT (the norther forests). Lets look through the summary stats. The reduced form part of the paper uses the fact that in the northern forests their was randomization into auction format for a subsample of the data. For the structural part, this is not important.

TABLE I (a)
SUMMARY STATISTICS FOR CALIFORNIA SALES

	Open auctions				Sealed auctions			
	Full sample		Selected		Full sample		Selected	
N	886		732		347		339	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Auction outcomes								
Winning bid ($/mbf)	61.82	43.84	63.27	44.91	69.45	46.64	69.42	46.63
Entrants	4.07	2.41	4.13	2.44	4.37	2.81	4.40	2.83
# Loggers entering	2.53	2.43	2.75	2.46	3.18	2.57	3.23	2.57
# Mills entering	1.53	1.70	1.38	1.68	1.19	1.74	1.17	1.74
Fraction loggers entering	0.59	0.40	0.64	0.39	0.75	0.34	0.76	0.33
Logger wins auction	0.54	0.50	0.59	0.49	0.67	0.47	0.68	0.47
Appraisal variables								
Volume of timber (hundred mbf)	32.00	43.06	23.02	34.30	16.70	30.01	14.65	25.50
Reserve price ($/mbf)	24.70	24.66	25.68	25.46	26.65	24.30	26.64	24.44
Selling value ($/mbf)	252.60	131.88	253.04	130.67	259.44	125.05	259.35	125.33
Road construction ($/mbf)	5.86	9.57	4.36	8.69	2.94	7.61	2.71	7.44
No road construction	0.59	0.49	0.68	0.47	0.78	0.42	0.79	0.41
Logging costs ($/mbf)	79.87	64.40	78.61	64.49	79.86	63.42	79.60	63.61
Manufacturing costs ($/mbf)	108.82	85.81	107.53	86.09	112.75	87.03	112.65	87.38
Sale characteristics								
Contract length (months)	23.02	17.93	22.19	16.35	16.78	14.72	15.94	13.38

TABLE I (a)
(CONTINUED)

	Open auctions				Sealed auctions			
	Full sample 886		Selected 732		Full sample 347		Selected 339	
N	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Species herfindal	0.61	0.28	0.61	0.28	0.59	0.27	0.59	0.27
Density of timber (mbf/acres)	7.83	7.01	7.85	7.20	8.91	8.21	8.97	8.26
Salvage sale	0.37	0.48	0.37	0.48	0.40	0.49	0.41	0.49
Scale sale	0.41	0.49	0.38	0.49	0.37	0.48	0.36	0.48
Quarter of sale	2.43	0.99	2.45	0.99	2.47	0.98	2.47	0.98
Year of sale	86.32	2.41	86.32	2.45	85.95	2.61	85.94	2.61
Housing starts	1557.53	255.73	1572.33	235.52	1542.04	274.69	1540.41	275.81
Potential competition								
Logging companies in county	44.41	20.88	43.17	21.32	41.48	22.00	41.73	22.10
Sawmills in county	8.67	4.32	8.45	4.35	7.67	4.26	7.56	4.15
Active loggers	20.14	9.71	19.91	9.71	19.24	8.90	19.47	8.86
Active mills	5.20	2.14	5.25	2.20	5.77	2.47	5.79	2.48

Notes. Data are from U.S. Forest Service sales held between 1982 and 1990 in Kootenai and Idaho Panhandle National Forests. Data is from the U.S. Forest Service, Timber Data Company, and U.S. Census Bureau. Timber is measured in thousand board feet, or mbf. The selected sample consists of sales with a propensity score, or estimated probability of being sealed bid, between 0.075 and 0.925. An Active Logger or Mill is a firm that has bid in the same forest district in the prior 12 months.

TABLE I (b)
SUMMARY STATISTICS FOR NORTHERN SALES

	Open auctions				Sealed auctions			
	Full sample		Selected		Full sample		Selected	
N	1290		325		774		382	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Auction outcomes								
Winning bid ($/mbf)	91.23	143.40	85.10	102.93	79.25	61.90	80.39	63.20
Entrants	4.08	2.35	3.72	2.24	3.74	2.57	4.27	2.76
# Loggers entering	1.33	1.58	1.95	1.89	2.81	2.22	3.01	2.35
# Mills entering	2.75	1.90	1.77	1.77	0.94	1.41	1.26	1.57
Fraction loggers entering	0.32	0.33	0.55	0.39	0.78	0.31	0.73	0.31
Logger wins auction	0.26	0.44	0.50	0.50	0.74	0.44	0.67	0.47
Appraisal variables								
Volume of timber (hundred mbf)	67.26	50.41	23.77	24.14	9.22	17.68	11.24	12.45
Reserve price ($/mbf)	36.79	36.72	37.31	49.64	36.07	32.48	33.26	32.28
Selling value ($/mbf)	270.83	99.48	233.41	143.15	246.92	251.63	239.72	119.74
Road construction ($/mbf)	9.67	12.58	4.39	9.76	1.09	4.31	1.58	5.23
No road construction	0.32	0.46	0.64	0.48	0.90	0.29	0.87	0.34
Logging costs ($/mbf)	108.28	44.78	88.42	58.59	86.56	56.74	97.73	58.49
Manufacturing costs ($/mbf)	122.74	42.85	100.02	61.11	99.62	62.89	104.93	60.28

TABLE I (b)
(CONTINUED)

	Open auctions				Sealed auctions			
	Full sample 1290		Selected 325		Full sample 774		Selected 382	
N	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Sale characteristics								
Contract length (months)	27.54	14.57	16.23	10.34	10.24	7.46	11.97	6.78
Species herfindal	0.56	0.23	0.59	0.25	0.60	0.24	0.61	0.25
Density of timber (mbf/acres)	11.54	14.75	11.93	16.94	18.38	220.16	11.37	15.83
Salvage sale	0.13	0.34	0.24	0.43	0.36	0.48	0.25	0.43
Scale sale	0.90	0.30	0.75	0.44	0.66	0.47	0.75	0.43
Quarter of sale	2.37	1.02	2.58	1.00	2.71	0.89	2.62	0.96
Year of sale	85.26	2.12	85.58	2.29	85.61	2.29	85.07	2.15
Housing starts	1593.72	254.02	1564.30	235.05	1559.08	247.12	1578.31	264.04
Potential competition								
Logging companies in county	21.76	18.30	20.75	18.66	20.00	17.39	21.09	19.00
Sawmills in county	6.28	6.05	5.77	5.08	6.00	6.13	6.70	7.45
Active loggers	10.32	7.08	10.11	6.86	10.48	5.98	11.29	6.62
Active mills	5.99	3.34	5.42	3.54	5.26	2.76	5.48	2.85

Notes. Data are from U.S. Forest Service sales held between 1982 and 1989 in California. Data is from the U.S. Forest Service, Timber Data Company, and U.S. Census Bureau. Timber is measured in thousand board feet, or mbf. The selected sample consists of sales with a propensity score, or estimated probability of being sealed bid, between 0.075 and 0.925. An Active Logger or Mill is a firm that has bid in the same forest in the prior 12 months.

4.3 Structural Analysis

The research questions can be expressed more tightly now:

1. can a calibrated version of the model capture departures from revenue equivalence
2. can we use the model to asses the extent of any departures from competitive bidding
3. what is the welfare impact of changing auction format

I will focus on the second question

The idea will be to use bid and entry data from FPSB auctions to estimate parameters for the model. Recall the maintained assumption that the FPSB auctions are competitive.

Following that, we use the model to predict the data. For FPSB auctions this is a measure of fit exercise. For open outcry auctions, this is a joint test of the modeling assumptions, including competitive bidding.

4.3.1 Estimation

Estimation proceeds in two steps: first use a parametric implementation of the GPV approach to get the value distributions. Second, use this to get profits to pop out the entry cost from the entry data.

I threshold question to consider is why the authors wanted to use a parametric model rather than the non-parametric modeling approach?

Let's look at the estimation of the value distributions first. Mostly, this is about getting a model of bids.

- The idea here is to posit a parametric model of the bid distribution, with unobserved auction heterogeneity, and then use the first order condition, in the spirit of GPV, to pop out the implied distribution on private information (values)
- recall that, conditional on entry, the number of bidders is known
- let bidders valuations be given by

$$F_L(.|X, u, N)$$

and

$$F_M(.|X, u, N)$$

with corresponding bid distributions being

$$G_L(.|X, u, N, n)$$

and

$$G_M(.|X, u, N, n)$$

X is observed auction heterogeneity and u is unobserved (to the econometrician) auction heterogeneity.

- Lastly we need a distribution on u, the unobserved (to the econometrician) auction heterogeneity.
- The parametric assumptions are that G_k's have a Weibull distribution and u has a gamma. The Weilbull-Gamma combo turns out to make it easy to write the likelihood...
- So:

$$G_k(b|X, u, N, n) = 1 - \exp\left(-u.\left(\frac{b}{\lambda_k(X, N, n)}\right)^{\rho_k(n)}\right)$$

where

$$\log \lambda_k(X, N, n) = X\beta_X + N\beta_N + n\beta_n + \beta_0$$

and

$$\log \rho_k(n) = n\gamma_n + \gamma_0$$

- lastly, u is distributed gamma with mean zero and variance θ.
- Estimation is via MLE. The log-likelihood is in the online appendix, which I reproduce here for reference:

The Likelihood Function

A useful property of Gamma-Weibull models is that the unobserved heterogeneity can be integrated out analytically. This leads to the following log-likelihood for auction t:

$$\begin{aligned}\ln L_t &= (n_{Lt}+n_{Mt})\ln\theta + \ln\Gamma\left(\frac{1}{\theta}+n_{Lt}+n_{Mt}\right) - \ln\Gamma\left(\frac{1}{\theta}\right) \\ &+ \sum_{i=1}^{n_{Lt}+n_{Mt}} \ln\left(p_{it}\lambda_{it}\left(\frac{b_{it}}{\lambda_{it}}\right)^{p_{it}-1}\right) + \left(\frac{1}{\theta}+n_{Lt}+n_{Mt}\right)\ln\left(1+\theta\sum_{i=1}^{n_{Lt}+n_{Mt}}\left(\frac{b_{it}}{\lambda_{it}}\right)^{p_{it}}\right).\end{aligned}$$

Here θ is the Gamma variance, $b_{1t}, ..., b_{(n_{Lt}+n_{Mt})t}$ are the observed bids in auction t, and λ_{it}, p_{it} are the Weibull parameters for bidder i in auction t. As defined in the text, these are functions of (X_t, N_t, n_t), the unknown parameter vectors β and γ, and bidder i's type — logger or mill.

- Once you have the estimated bid model, it is easy to use the FOC's a la GPV to pop out a value distribution (conditional on a value of u). Let's look at the results

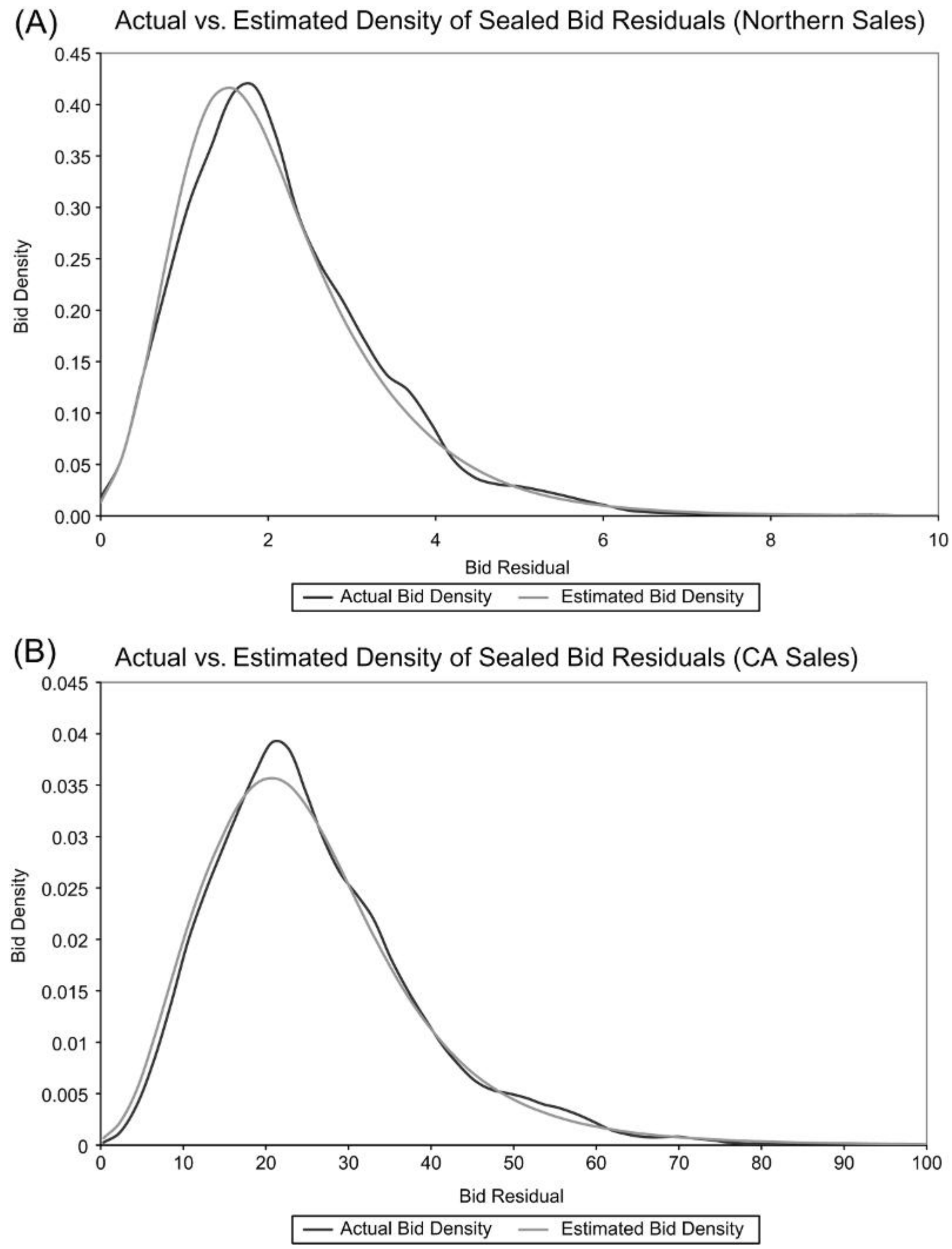

FIGURE I

Actual versus Estimated Density of Sealed Bid Residuals (Northern Sales and CA Sales)

Notes. Figures show estimated and actual distribution of bid residuals from sealed bids in Northern sales (A) and California sales (B). The bid residual for bid i in auction t is defined as $\varepsilon_{it} = b_{it} / \exp(X_t\beta_X + N_t\beta_N)$, using the estimated b's. The plotted distribution of bid residuals is smoothed using a kernel. Estimated bid residuals are those predicted by the Gamma–Weibull bid distribution model.

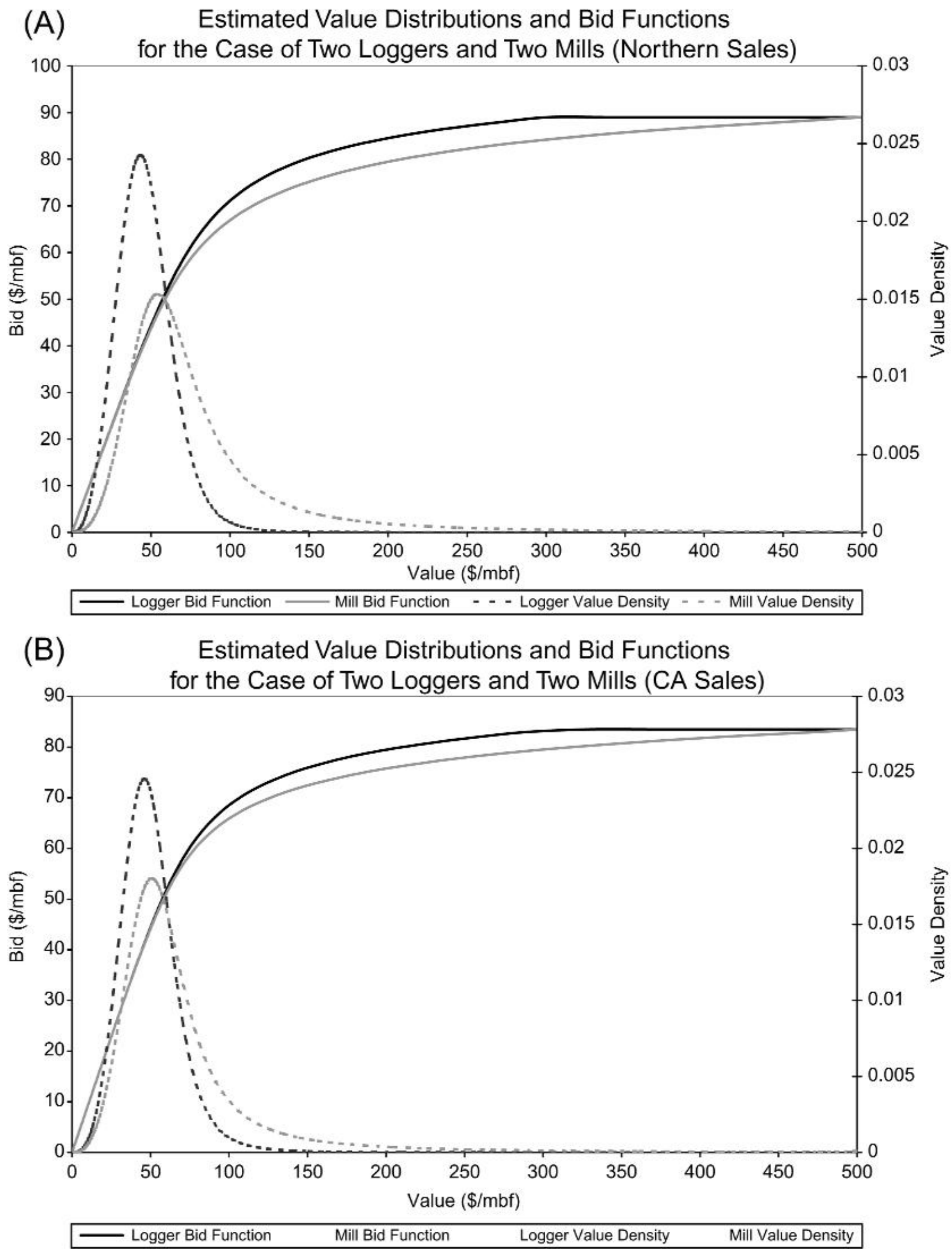

FIGURE II

Estimated Value Distributions and Bid Functions for the Case of Two Loggers and Two Mills (Northern Sales and CA Sales)

Notes. Figures show estimated value distributions and equilibrium bid functions for a sale in the Northern region (A) or California (B) with average covariates, and u = 1, assuming that two loggers and two mills participate in the auction.

Now we turn to the estimation of the entry model. The parameter to recover is the entry cost, K.

- From theory, all mills enter. So entry is estimated off the behavior of the loggers.
- The first thing we need is the set of potential entrants: this is taken from the set of all entrants from the previous year.
- Next we need to write down the entry condition:

$$\Pi(X,N) = \sum_{n \in N} \pi_L(X,N,n) Pr(n|X,N,n,i \in n) = K(X,N)$$

- From the bid fuction and value distribution estimation above, we can work out $\pi_L(X,N,n)$, so the only thing is to estimate the probability of the numbers of loggers turning up. WIth rich enough data this could be done merely by counting, here however, we have to take a slightly more parametric approach:

In particular,

(12)
$$\Pr[n_L|X,N,i \in n,s] = \binom{N_L - 1}{n_L - 1} p^s(X,N)^{n_L-1} \left(1 - p^s(X,N)\right)^{N_L - n_L}.$$

For estimation, we specify a parametric model:

$$p^s(X,N) = \frac{\exp\left(X\alpha_X + N\alpha_N\right)}{1 + \exp\left(X\alpha_X + N\alpha_N\right)}. \quad (13)$$

We estimate the parameter vector α by maximum likelihood using the observed logger entry into sealed bid auctions. These estimates are reported in Table IV.[25]

- Then, recover K for each auction (ie. X,N pair) by setting $\Pi(X,N) = K(X,N)$

So how does the model fit? and what can we learn about competition in the open auctions?

step, however, provides a demanding test of the theory. We use the model to predict the outcomes of the open auctions and compare these predictions to the data. Here we are asking the model

TABLE V
ACTUAL OUTCOMES VERSUS OUTCOMES PREDICTED BY MODEL

	N	(1) Actual	(2) Predicted (bidding only)		(3) Predicted (bidding + entry)	
A: Northern sales						
Sealed bid sales						
Avg. bid	1,492	59.6	58.2	(1.4)	57.4	(1.3)
Avg. logger bid	1,096	50.8	48.7	(1.4)	47.4	(1.4)
Avg. mill bid	396	83.8	84.7	(2.7)	85.2	(2.7)
Avg. sale price ($/mbf)	339	69.4	69.9	(1.4)	70.4	(1.6)
Avg. revenue ($000s)	339	111.4	108.1	(4)	109.9	(4.2)
% sales won by loggers	339	68.1	68.0	(0.90)	65.0	(0.01)
Avg. logger entry	339	3.23			3.23	(0.09)
						(0.1)
Open auction sales						
Avg. sale price (competition)	732	63.3	67.9	(1.8)	67.8	(2.1)
Avg. sale price (collusion)	732	63.3	44.2	(1.3)	44.1	(2.2)
Avg. revenue (competition)	732	144.7	152.7	(6.8)	154.8	(7.9)
Avg. revenue (collusion)	732	144.7	61.0	(2)	64.7	(5.0)
% sales won by loggers	732	59.0	56.0	(0.01)	54.4	(0.02)
Avg. logger entry	732	2.75			2.67	(0.17)
B: California sales						
Sealed bid sales						
Avg. bid	1,630	73.6	74.7	(2.3)	74.2	(2.3)
Avg. logger bid	1,150	64.0	63.6	(2.1)	62.3	(2.4)
Avg. mill bid	480	96.5	101.2	(3.5)	102.8	(3.8)
Avg. sale price ($/mbf)	382	80.4	83.8	(2.1)	84.4	(2.4)
Avg. revenue ($000s)	382	103.1	110.7	(3.8)	111.9	(4.0)
% sales won by loggers	382	66.8	66.4	(1.2)	62.6	(1.3)
Avg. logger entry	382	3.01			3.01	(0.07)
Open auction sales						
Avg. sale price (competition)	325	85.1	87.2	(2.7)	86.7	(3.1)
Avg. sale price (collusion)	325	85.1	46.1	(1.2)	51.0	(1.6)
Avg. revenue (competition)	325	227.0	244.7	(9.7)	242.4	(10.9)
Avg. revenue (collusion)	325	227.0	93.2	(2.6)	112.9	(5.6)
% sales won by loggers	325	50.5	48.2	(1.1)	43.6	(1.8)
Avg. logger entry	325	1.95			1.90	(0.13)

Notes. Column (1) shows average outcomes for sale sealed bid or open sales in the region. Column (2) shows predicted outcomes from the model for those same sales, conditional on the number of entering firms observed in the data. Column (3) shows predicted outcomes based on the equilibrium model of entry and bidding. All standard errors obtained by a parametric bootstrap.

The hypothesis that the actual price and the predicted price for Northern open auctions are the same is rejected. Whether this difference is economically important is open for debate. One thing the authors do is disaggregate this difference a little into auctions with different entry patterns. See below

TABLE VI
ACTUAL VERSUS PREDICTED SALE PRICES BY MILL PARTICIPATION

	N	(1) Actual	(2) Predicted (bidding only)	(3) Predicted (bidding + entry)
Zero mills				
Sealed bid sales	181	51.7	51.4	51.4
Open auction sales	321	49.8	50.5	47.1
One mill				
Sealed bid sales	70	66.8	64.6	66.9
Open auction sales	150	50.0	52.2	59.5
Two or more mills				
Sealed bid sales	88	108.1	112.1	112.2
Open auction sales	261	87.5	98.5	98.0

Notes. All numbers are for sales in the Northern region. Column (1) shows average sales prices for sales with zero, one, or two or more participating mills. Columns (2) and (3) show predicted prices for these sales based on the estimated model.

In thinking through this paper as a detection device, bear in mind that at heart what is going on is a completely specification test, and that the model is perhaps most extreme in the assumptions about the entry of the mills.