Deciding between Competition and Collusion

Author(s): Patrick Bajari and Lixin Ye

Source: *The Review of Economics and Statistics*, Nov., 2003, Vol. 85, No. 4 (Nov., 2003), pp. 971–989

Published by: The MIT Press

Stable URL: https://www.jstor.org/stable/3211820

DECIDING BETWEEN COMPETITION AND COLLUSION

Patrick Bajari and Lixin Ye*

Abstract—We develop an approach to identify and test for bid rigging in procurement auctions. First, we introduce a general auction model with asymmetric bidders. Second, we study the problem of identification in our model. We state a set of conditions that are both necessary and sufficient for a distribution of bids to be generated by a model with competitive bidding. Third, we discuss how to elicit a prior distribution over a firm's structural cost parameters from industry experts. Given this prior distribution, we use Bayes's theorem to compare competitive and collusive models of industry equilibrium. Finally, we apply our methodology to a data set of bidding by construction firms in the Midwest. The techniques we propose are not computationally demanding, use flexible functional forms, and can be programmed using most standard statistical packages.

I. Introduction

BID rigging is a serious problem in many procurement auctions. According to *Engineering News-Record,* criminal bid-rigging cases have recently been filed in New York City and Chicago for building public schools, bridge repair, interior remodeling, paving, and many other types of construction. A widely publicized instance of bid rigging occurred in the New York cement industry in the 1980s, where organized crime designed an elaborate scheme that inflated building costs, making the price of poured concrete the highest in the nation.[1] Developing effective and computationally simple tools that can be used by regulators to detect bid rigging might serve as a deterrent to future collusion, and thereby lower prices and enhance efficiency in some industries.

In this paper, we develop an approach to identify and test for bid rigging in procurement auctions. We begin by describing a model of competitive bidding for procurement contracts. A unique aspect of our model is that bidders are asymmetric, that is, ex ante, the costs of bidders may differ. Asymmetries are commonplace in procurement and may arise from the location of firms, capacity constraints, or familiarity with local rules and regulations.

Next, we study the problem of identification in the asymmetric auction model. We state a set of conditions that are both necessary and sufficient for a distribution of bids to be rationalized by our model of competitive bidding, and we discuss how these conditions can be tested. A first condition that can easily be tested is conditional independence, which implies that after controlling for all information about costs that are publicly observed by the firms, the bids must be independent. If collusion is occurring, however, we might expect to find correlated bids when cartel members submit "phantom" bids meant to give the appearance of competition. A second condition is exchangeability, which implies that costs alone should determine how firms bid and that the identities of a firm's competitors, holding information about costs constant, should not change how a firm bids. If collusion is occurring, however, we might find that cartel members do not bid against each other as aggressively as a control group of noncartel firms.[2]

Finally, we use the tools of statistical decision theory to decide between competitive and collusive models of industry equilibrium. We elicit from industry experts a prior distribution over the structural cost parameters that enter our models. Given this distribution, we use Bayes's theorem and the laws of conditional probability to decide between competitive and collusive models of industry equilibrium. We apply our methodology to a data set of bidding by construction firms. The techniques needed for the computations are not particularly complex and can be programmed using standard statistical packages.

In our application, we begin by testing for conditional independence and exchangeability. Although the vast majority of bidders appear to satisfy the conditions, two pairs, (firm 2, firm 4) and (firm 2, firm 5), fail at least one of these tests. The owners of all three of these firms have previously been sanctioned for bid rigging in this market. Our reduced form testing suggests that there are three models of interest: no collusion, (firm 2, firm 4) collude, and (firm 2, firm 5) collude. We then compare these three models by computing posterior probabilities for each model.

In empirical industrial organization, it is common to consult with industry experts during a research project in order to learn about the nature of demand, costs, and strategic interactions in the marketplace. However, information

[1] In his biography, Mafia informer Sammy (The Bull) Gravano, who claims to have orchestrated collusion in the construction industry for the Gambino crime family, is reported to have said, "If one of them [contractors] gets a contract for, say, thirteen million, the next thing you know, after he knows he's got it, he jacks up the whole thing before it's over to a sixteen- or seventeen-million-dollar job. Now he's increased the cost thirty-three percent. So our [the Mafia's] greed is compounded by the greed of them so-called legitimate guys [contractors]" (Maas, 1997, p. 271).

[2] As we discuss below, a clever cartel could submit collusive bids that satisfy both of these conditions. We are unaware of any empirical example of collusion where this has occurred. However, we do discuss empirical examples of collusion where conditional independence and exchangeability fail.

garnered from industry sources is seldom *formally* incorporated into the estimation. A unique feature of our analysis is that we elicit the beliefs of industry experts about the model parameters and incorporate these beliefs into estimation and testing.

Our two industry experts are leading owner-operators in the seal coat industry with over 50 years of combined experience. Because bidders typically receive quotes from the same subcontractors and materials suppliers and because all bids are publicly available after the contract is awarded, a bidder will have a lot of information about the markups of competing bidders. In our data set, over 100 firms bid, but only 7 of these firms have a market share that exceeds 5%. If our industry experts are not savvy in managing their own costs and understanding the costs of their competitors, it is unlikely that they will remain in business for very long.

Our estimates show that the markups implied by the competitive model are much closer to the industry experts' prior beliefs than the two collusive models, and thus the competitive model has a posterior probability close to 1. We do not take this, or any of our tests, as definitive proof of competition or collusion. However, the collusive models generate markups much higher than industry experts find plausible. We argue that this is a meaningful piece of evidence against the collusive models.

We do not advocate exclusively estimating models by imposing priors from external sources such as industry experts, consulting engineers, or internal cost records of firms. However, we believe that augmenting the data with a prior distribution elicited from industry sources can be helpful when other identifying information is not available, the data are of poor quality, or the data are incomplete. Moreover, using a prior distribution elicited from industry experts will often improve the small-sample properties of our estimators and our hypothesis tests. This is particularly important in antitrust problems where both the quantity and quality of the data may be limited, but where the market regulators must make a decision regardless of the limitations of the data. Finally, we found that eliciting a prior and discussing the market with the industry experts helped us to pick up on many subtleties of strategic behavior in this industry.

There are a number of recent empirical papers on the subject of bid rigging. The first set describe the observed bidding patterns of cartels and compare cartel with noncartel bidding behavior. Porter and Zona (1993, 1999) and Pesendorfer (2000) analyze data sets where it is known that bid rigging has taken place. These papers find the following empirical regularities: First, cartel members tend to bid less aggressively than non-cartel-members. Second, the bids of cartel members tend to be more correlated with each other than with the bids of non-cartel-members. Third, collusion tends to increase prices over those of a noncollusive control group.

The second set of empirical papers, such as Porter and Zona (1993) and Baldwin, Marshall, and Richard (1997), propose econometric tests designed to detect collusive bidding. Baldwin et al. (1997) nest competition and collusion within a single model to test for collusion. Their model is applicable for oral or second-price auctions with private values. Porter and Zona (1993) propose a procedure where two models of bidding are estimated. The first model is a logistic regression on the identity of the lowest bidder. The second model is an ordered logit regression on the ranking of all the bidders. Under the null hypothesis of no collusion, the parameter values from the two models should be equal. Our analysis sheds some new light on the analysis of Porter and Zona (1993, 1999) and Baldwin et al. (1997) by demonstrating how observable differences across firms, such as location and capacity, play a key role in the identification of collusion.

Our paper is also related to the literature on structural estimation of the first-price auction model. This literature began with Paarsch (1992). Most closely related to our research is that of Guerre, Perrigne, and Vuong (2000), who estimate the first-price auction model nonparametrically. Also related is that of Bajari (1997), Campo, Perrigne, and Vuong (2001), Flambard and Perrigne (2001), Hong and Shum (1999), and Pesendorfer and Jofre-Bonet (2000), who structurally estimate asymmetric models of the first-price auction. Finally, our paper is related to a literature in statistics on the elicitation of prior beliefs from experts and its use in estimation and testing. See, for instance, Cooke (1991), Garthwait and Dickey (1988, 1992), Kadane and Wolfson (1998), Kadane et al. (1980), and O'Hagan (1998).

We believe that the tests we propose are a useful diagnostic for detecting suspicious bidding behavior. No method for detecting collusion is foolproof. Also, the economist must exercise careful judgement to determine whether his hypothesis tests are economically significant, not just statistically significant. Despite the limitations inherent in any test for collusion, we believe that our methods can be used as a first step in determining whether suspicious bidding might have occurred and whether further investigation is warranted.

II. The Model

We consider a procurement auction model in which N firms compete for a contract to build a single and indivisible public works project. The firms have independent private cost estimates. Firm i knows its own cost estimate (c_i), but not the cost estimates of other firms (c_{-i}). We assume that c_i is drawn from a distribution with cumulative distribution function $F_i(\cdot)$ and density function $f_i(\cdot)$. Both $F_i(\cdot)$ and $f_i(\cdot)$ are common knowledge among all firms before bidding starts. For technical convenience, we assume that c_i has the

same support $[\underline{c}, \bar{c}]$ for all i.[3] As in other procurement lettings, a first-price sealed-bid auction is conducted: firms submit sealed bids, and the lowest bidder wins the contract at a price equal to the bid submitted by that bidder.

We assume that firms are risk-neutral with respect to their monetary income. Firm i's strategy is a function $B_i(\cdot) : [\underline{c}, \bar{c}] \to \mathbf{R}_+$. Suppose there exists an increasing equilibrium such that $B_i(\cdot)$ is strictly increasing and differentiable on the support of c_i for all i; then its inverse bid function, $\phi_i(\cdot)$, is also strictly increasing and differentiable on the support of the bids. Suppose that all the competing firms follow strategies B_{-i}; then if firm i bids b_i, its probability of winning is $\Pr(c_j > \phi_j(b_i))$ for all $j \neq i$). Firm i's expected profit can thus be written as

$$\pi_i(b_i, c_i; B_{-i}) = (b_i - c_i)Q_i(b_i), \qquad (1)$$

where

$$Q_i(b_i) = \prod_{j \neq i} [1 - F_j(\phi_j(b_i))] \qquad (2)$$

is the probability that firm i wins the contract. As we can see from equation (1), firm i's expected profit is a markup times the probability that firm i is the lowest bidder.

Since each firm's profit function depends only on its own private information, this is a procurement auction model with private values (costs). In the seal coat industry that we are going to analyze, because labor and material costs are mainly firm-specific, we believe that the assumption of private values (costs) is appropriate.[4]

Unlike most of the auction models in literature, our model allows for asymmetric bidders. In procurement settings, the symmetric-bidder assumption implies that private cost estimates are independently and identically distributed. This may not be the case in the seal coat industry as well as many other procurement settings, because asymmetries can easily arise from different locations, different capacity constraints, different technologies, and different levels of managerial efficiencies across firms.

A. Dynamic Bidding

Under certain assumptions, our model also allows for nontrivial dynamics. We look for a stationary equilibrium such that each firm's strategy only depends on the current state of the industry. Let s denote a vector of state variables that can be observed by all participants in the auction. This vector may include, for example, capacities and locations of all the participants and market prices of key materials. At

state s, let $V_i^W(s)$ be firm i's continuation value when it wins, and $V_{i,j}^L(s)$ be its continuation value when some other firm, j, wins. In what follows we assume that $V_{i,j}^L(s) = V_i^L(s)$, that is, the continuation value for a firm when it fails to win does not depend on the identity of the firm that wins.[5] We employ this simplification to ensure that our model has a unique equilibrium, which is essential for how we test for collusion in the next sections. For an analysis of dynamic bidding that does not employ this assumption, see Pesendorfer and Jofre-Bonet (2000).

With this assumption, firm i's expected profit at state s by bidding b_i while all its competing firms follow an increasing equilibrium $B_{-i}(\cdot)$ can be written as

$$\pi_i(b_i, c_i; B_{-i}) = [b_i - c_i + V_i^W(s)]Q_i(b_i) + V_i^L(s)$$
$$\times [1 - Q_i(b_i)] \qquad (3)$$

$$= [b_i - c_i + V_i^W(s) - V_i^L(s)]Q_i(b_i)$$
$$+ V_i^L(s), \qquad (4)$$

where $Q_i(b_i)$ is given by (2). Since $V_i^L(s)$ is a constant, we can write the firm's maximization problem as follows:

$$\max_{b_i} [b_i - c_i + V_i^W(s) - V_i^L(s)]Q_i(b_i). \qquad (5)$$

As in equation (1), a firm's objective function is the markup times the probability of winning, the only difference being that a firm's cost estimate now reflects the option value of having free capacity available for future projects. Therefore, the framework developed for the static model above can be applied to the analysis of dynamic bidding with capacity constraints.[6]

B. Collusion

Many elaborate schemes for collusion have been found in industry. For example, bid rotation schemes like "phases of the moon" are followed by cartel members to allocate projects, side payments are made within a cartel to divide spoils, and geographic territories are established as parts of cartel arrangements.

It turns out that a collusive bidding model where the cartel behaves efficiently can also be analyzed within the framework developed above. An efficient cartel operates as follows. First the cartel members communicate before an

---

[3] If this assumption is violated, we may end up with *nonessential* equilibria as discovered in Griesmer, Levitan, and Shubik (1967). The same support assumption may sound restrictive, but if we can add an arbitrarily small perturbation to any given distribution of the costs, then we can always start with the assumption of common supports.

[4] In Armantier, Florens, and Richard (1997) and Porter and Zona (1993), the private-value model is also assumed in procurement auction analyses.

[5] $V_{i,j}^L(s)$ in general depends on the identity of the winner. However, our data set suggests that our assumption may not be too restrictive. If a firm's continuation value depended on which competitor won, it would bid differently depending on the capacity of its competitors. However, in the analysis of our data set, while firm i's own capacity was significant, the capacity of other firms failed to be significant. This finding is consistent with our assumption that firm i is indifferent to which firm wins the contract.

[6] When costs reflect option values, the distributions of cost estimates may not have the same supports across firms. However, as discussed earlier, as long as we can add arbitrarily small perturbation to the distributions, assuming common supports is without loss of generality.

auction is conducted and compare their cost estimates. Then the member with the lowest cost estimate submits a serious bid, and all the other members either refrain from bidding or submit high bids (phony bids). Let $C \subseteq \{1, 2, \ldots, N\}$ denote the cartel, and c_c the cost of the cartel; then

$$c_c = \min_{j \in C} c_j.$$

Because c_c is an order statistic, its distribution can be easily computed. Following the framework developed above, if other bidders are aware of the identity of the cartel, then it is straightforward to adapt our previous analysis to this case with efficient collusion. If the cartel members bid against each other with sufficient frequency and if side payments are possible, it can be shown using standard folk-theorem arguments that efficient collusion can be sustained through repeated play.

III. Properties of Equilibrium

For the asymmetric auction model described in the previous section, the equilibrium properties have been well established in the theoretical literature. Following the literature, we maintain two regularity conditions regarding the information structure:

- *Assumption 1.* For all i, the distribution of costs $F_i(\cdot)$ has support $[\underline{c}, \bar{c}]$. The probability density function $f_i(\cdot)$ is continuously differentiable (in c_i).
- *Assumption 2.* For all i, $f_i(\cdot)$ is bounded away from zero on $[\underline{c}, \bar{c}]$.

We assume that firms follow a Bayes-Nash equilibrium in pure strategies. An equilibrium in pure strategies is a collection of functions $B_1, \ldots, B_N$ such that $B_i(c_i)$ maximizes $\pi(b_i, c_i; B_{-i})$ in b_i for all i and c_i in its support. Suppose there exists an increasing equilibrium such that each firm (i) bids according to a strictly increasing function $B_i(c_i)$ [denote the inverse bid function as $\phi_i(b_i)$]. Then the first-order condition for such an equilibrium is

$$\frac{\partial}{\partial b_i} \pi_i(b_i, c_i; B_{-i}) = (b_i - c_i)Q_i'(b_i) + Q_i(b_i) = 0, \quad (6)$$

where $Q_i(b_i)$ is given by (2).

After simplification, the first-order condition for the equilibrium can be written as follows:

$$c_i = b_i - \frac{1}{\sum_{j \neq i} \dfrac{f_j(\phi_j(b_i))\phi_j'(b_i)}{1 - F_j(\phi_j(b_i))}}, \quad i = 1, \ldots, N. \quad (7)$$

Alternatively, we can rearrange terms to obtain the following system of N ordinary differential equations:

$$\phi_i'(b_i) = \frac{1 - F_i(\phi_i(b_i))}{(N-1)f_i(\phi_i(b_i))}$$
$$\times \left[\frac{-(N-2)}{b_i - \phi_i(b_i)} + \sum_{j \neq i} \frac{1}{b_i - \phi_j(b_i)} \right], \quad (8)$$

$$i = 1, \ldots, N.$$

Under the above two regularity assumptions, Lebrun (1999) and Maskin and Riley (2000a) demonstrate that a set of equilibrium bid functions exist and these functions are strictly increasing and differentiable:

Theorem 1 (Lebrun, 1996; Maskin & Riley, 2000b). Under assumptions 1 and 2, there exists an equilibrium in pure strategies. Furthermore, the equilibrium bid function for each bidder is strictly monotone and differentiable.

The equilibrium can also be characterized as the solution to a system of N differential equations with $2N$ boundary conditions:

Theorem 2 (Lebrun, 1999; Maskin & Riley, 2000a). Under assumptions 1 and 2, the equilibrium inverse bid functions can be characterized as the solutions to the system of N differential equations (8) with the following boundary conditions:

(i) For all i, $\phi_i(\bar{c}) = \bar{c}$.
(ii) For all i, $\phi_i(\beta) = \underline{c}$ for some constant β.

Finally, under the same set of assumptions, it is shown that the equilibrium is unique:

Theorem 3 (Maskin & Riley, 1996; Bajari, 1997, 2001; Lebrun (2000)). Under assumptions 1 and 2, the equilibrium (in pure strategies) is unique.

IV. Identification

In this section we identify a set of conditions about a distribution of bids that are implied from the model of competitive bidding introduced in section 2. We also show that under this set of conditions, we can construct a unique set of latent cost distributions that rationalizes the distribution of bids in equilibrium under competitive bidding.

Assume that each firm's cost distribution can be parametrized by θ, a vector of parameters, and z_i, a set of covariates that is unique to firm i. We write the cumulative distribution function of firm i's cost in the form $F(c_i | z_i, \theta)$. As an example, suppose that a firm's cost estimate can be written as follows:

$$c_i = \alpha + \beta d_i + \epsilon_i, \quad i = 1, \ldots, N. \quad (9)$$

In equation (9), α is a constant that captures the common factors affecting all firms identically, such as how many

miles of highway must be seal-coated; d_i is the distance from firm i's location to the project site; and β reflects firm i's unit transportation costs. The idiosyncratic shock ϵ_i is private information to firm i. In this example we assume that ϵ_i is distributed as Normal$(0, \sigma^2)$; then the distribution of costs depends on the vector of parameters $\theta = (\alpha, \beta, \sigma^2)$, which are common to all firms. Let $z_i = d_i$; then $z = (z_1, \ldots, z_N)$ is a set of covariates that is observable to firms. In what follows, for notational simplicity we shall suppress the dependence of private information on the parameters θ, which we shall assume to remain fixed across contracts. In Section 5.3, we shall extend our results to allow for firm-specific parameters.

Let $G_i(b; z)$ be the cumulative distribution of firm i's bids, and $g_i(b; z)$ be the associated probability density function. Note that the distribution of bids depends on the entire vector $z = (z_1, \ldots, z_N)$. Using the theorems from the previous section, it is straightforward to show that the following conditions must hold in equilibrium.

A1. Conditional on z, firm i's bid and firm j's bid are independently distributed.

A2. The support of each distribution $G_i(b; z)$ is identical for each i.

Conditional on z, each firm's signal, c_i, is independently distributed, and since bids are a function of c_i, this implies that A1 must hold in equilibrium. Although this condition must hold when bidding is competitive, it may fail when bidding is collusive. In Porter and Zona (1993, 1999), for instance, cartel members submitted phantom bids meant to give the appearance of competition. However, since the cartel members colluded, their bids were correlated conditional on z.[7] The condition A2 must hold by the characterization theorem (theorem 2) stated in section III.

A third condition that must hold in equilibrium is that the distribution of bids must be *exchangeable* in z_i. Let π be a permutation, that is, a one-to-one mapping from the set $\{1, \ldots, N\}$ onto itself. The definition of exchangeability is that for any permutation π and any index i the following equality must hold:

$$G_i(b; z_1, z_2, z_3, \ldots, z_N) = G_{\pi(i)}(b; z_{\pi(1)}, z_{\pi(2)}, z_{\pi(3)}, \ldots, z_{\pi(N)}).$$ (10)

Equation (10) implies that if the cost distributions for the bidders are permuted by π, then the empirical distribution of bids must also be permuted by π. For instance, if we permute the values of z_1 and z_2, holding all else fixed, exchangeability implies that the distribution of bids submitted by firms 1 and 2, $G_1(b)$ and $G_2(b)$, also permute. If

collusion occurs, then bidding need not be exchangeable if cartel members do not bid aggressively against each other as compared to a control group of noncartel firms. It is straightforward to show that exchangeability will hold in our model, because the equilibrium is unique and therefore permuting the cost distributions permutes the equilibrium bid functions in a symmetric fashion.

A3. The equilibrium distribution of bids is exchangeable. That is, for all permutations π and any index i, one has $G_i(b; z_1, z_2, z_3, \ldots, z_N) = G_{\pi(i)}(b; z_{\pi(1)}, z_{\pi(2)}, z_{\pi(3)}, \ldots, z_{\pi(N)})$.

The fourth condition that must hold in equilibrium is equivalent to the monotonicity of firm i's bid function in c_i. Note that from (7), we can rewrite the first-order conditions as follows:

$$c_i = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{f(\phi_j(b; z)|z_j)\phi_j'(b; z)}{1 - F(\phi_j(b; z)|z_j)}}.$$ (11)

Since equilibrium bid functions are strictly monotone, it follows using a simple change-of-variables argument that $G_i(b; z)$ and $g_i(b; z)$ must satisfy

$$G_i(b; z) = F(\phi_i(b; z)|z_i),$$ (12)

$$g_i(b; z) = f(\phi_i(b; z)|z_i)\phi_i'(b; z).$$ (13)

Substituting equations (12) and (13) into equation (11), the first-order conditions for equilibrium can be expressed as

$$\phi_i(b, z) = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{g_i(b; z)}{1 - G_i(b; z)}}.$$ (14)

In equilibrium, the bid functions must be strictly monotone. An equivalent condition to the monotonicity of the bid functions is the monotonicity of the function $\xi_i(b; z)$ in b, where $\xi_i(b; z)$ is defined as

$$\xi_i(b; z) = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{g_i(b; z)}{1 - G_i(b; z)}}.$$ (15)

A4. For all i and b in the support of the $G_i(b; z)$, the function $\xi_i(b, z)$ is strictly monotone.

Finally, from our characterization theorem, the following boundary conditions should also hold:

A5. $\xi_i(\bar{b}; z) = \bar{c}, \xi_i(\underline{b}; z) = \underline{c}$ for $i = 1, \ldots, N$.

We formalize the above observations into the following theorem.

---

[7] Why do cartel members submit phantom bids if this behavior can possibly be detected? We conjecture that the failure of a firm to bid at all when it has a favorable cost distribution might be a more obvious sign of collusion to procurement officials.

Theorem 4. Suppose that the distribution of bids $G_i(b; z)$, $i = 1, \ldots, N$, is generated from a Bayes-Nash equilibrium. Then conditions A1–A5 must hold.

Next, we show that if conditions A1–A5 hold, then it will be possible to construct a distribution of costs $F(b|z_i)$ that uniquely rationalizes the observed bids $G_i(b; z)$ as an equilibrium. In other words, conditions A1–A5 are not only necessary for an equilibrium, but also sufficient.

Theorem 5. Suppose that the distribution of bids $G_i(b; z)$ satisfies conditions A1–A5. Then it is possible to construct a unique set of distributions $F_i(c|z_i)$ such that $G_i(b; z)$ is the equilibrium distribution of bids when the costs are distributed $F_i(c|z_i)$.

Proof. See appendix.

V. Testing

In this section, we discuss how conditions A1–A5 can be used to test whether a given distribution of bids is consistent with our model of competitive bidding. If the bids $b = (b_1, \ldots, b_N)$ are conditionally independent, then it should be the case that

$$G(b_1, \ldots, b_N; z) = \prod_{i=1}^{N} G_i(b_i; z), \quad (16)$$

where $G(b_1, \ldots, b_N; z)$ is the joint distribution of bids. With unlimited data, to determine whether or not the bids are conditionally independent, the econometrician would ideally estimate $G(b_1, \ldots, b_N; z)$ and $G_i(b_i; z)$ nonparametrically and then test (16). Similarly, to determine whether or not the distribution of bids is exchangeable, the econometrician could test (10) after nonparametrically estimating the distribution of bids. Alternatively, the economist can estimate $G_i(b_i; z)$ using regression and following Porter and Zona (1993, 1999), and test conditional independence by testing whether the residuals to the bid function are independent.

Regression-based methods can be used to test exchangeability as well. Exchangeability implies that, for any given permutation π, equation (10) holds. As a simple example, suppose that our model of costs is given by (9) and that firms bid on projects at varying locations, so that the distance $\text{DIST}_{i,t}$ of firm i from various projects t changes. Let there be $t = 1, \ldots, T$ auctions, and let $z_i = (\beta_1, \text{DIST}_{i,t}, \sigma)$. Suppose that we estimate the following regression:

$$b_{i,t} = \beta_0 + \beta_{1,i} \text{DIST}_{i,t} + \sum_{j \neq i} \eta_{i,j} \text{DIST}_{j,t} + \epsilon_{i,t}. \quad (17)$$

If the distribution of bids is exchangeable, then it must be the case that as T becomes infinite,[8]

$$\text{for all } i \neq j, \quad \beta_{1,i} = \beta_{1,j}; \quad (18)$$

$$\text{for all } j, k, \quad j \neq i, \ k \neq i, \quad \eta_{i,j} = \eta_{i,k}; \quad (19)$$

$$\text{for all } i \neq j \text{ and for all } s, \quad \eta_{i,s} = \eta_{j,s}. \quad (20)$$

A. Distinguishing Competition and Collusion

If no variation in z is observed in the data, it will typically not be possible to determine whether collusion has occurred. This can be proved using an approach similar to theorem 5. When there is no variation in z, the first-order condition (14) implies that there is a one-to-one mapping between i's bid b_i and i's private information c_i. Therefore, given a distribution of bids, $G(b_1, \ldots, b_N)$, which does not vary with any observable variables z, we can rationalize the observed bids with a competitive model by assuming that when i bids b_i, then i's private information c_i satisfies

$$c_i = b_i - \frac{1}{\displaystyle\sum_{j \neq i} \frac{g_i(b; z)}{1 - G_i(b; z)}}. \quad (21)$$

Alternatively, we can rationalize the observed bids as the outcome of collusive bidding by a cartel $C \subseteq \{1, \ldots, N\}$ of firms (as in section 2.2). Let G_c denote the distribution of the lowest bid submitted by the firms in the set C. Then if a noncartel firm i bids b_i, let the firm's private information satisfy

$$c_i = b_i - \frac{1}{\dfrac{g_c(b; z)}{1 - G_c(b; z)} + \displaystyle\sum_{j \neq i, \ j \notin C} \frac{g_i(b; z)}{1 - G_i(b; z)}}, \quad (22)$$

and if the lowest cartel bid is b_c, let the cartel's private information c_c satisfy

$$c_c = b_c - \frac{1}{\displaystyle\sum_{j \neq i, \ j \notin C} \frac{g_i(b; z)}{1 - G_i(b; z)}}. \quad (23)$$

Using equation (21), it is possible to rationalize the data as the outcome of noncollusive bidding. If, instead, we use equations (22) and (23), we can rationalize the observed bids as the outcome of bidding by a cartel C. The intuition

---

[8] It is straightforward to show that if the econometrician projects T randomly drawn bids from the two distributions $G_i(b_i; z)$ and $G_j(b_j; z)$ onto the same linear subspace using least squares, then the projections should be the same as T tends to infinity under the null hypothesis of exchangeability. Note that this test is valid even if the linear regression (17) is a misspecified model of the bid function.

FIGURE 1.—
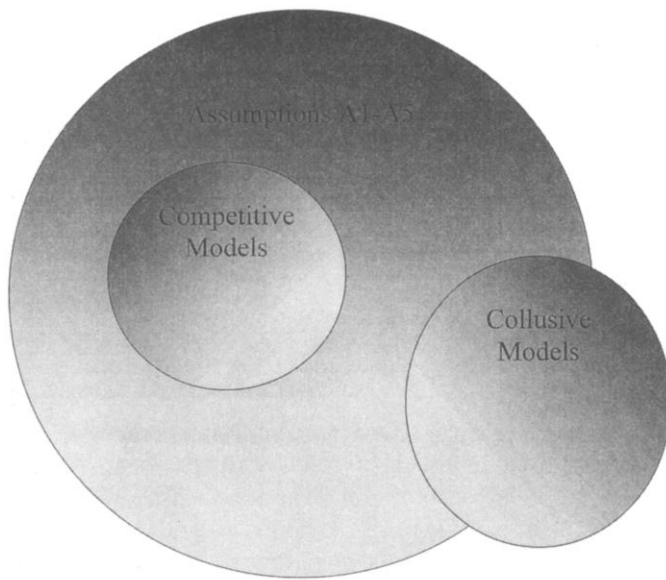
behind this result is that the first-order condition just identifies a bidder's private information c_i.

Theorem 6. If there is no variation in z and A1–A2 and A4–A5 hold, then competition is observationally equivalent to a cartel C that is not all inclusive.

Even if there is variation in z, it may still not be possible to empirically distinguish collusion from competition. A sophisticated cartel that includes all N firms may be able to construct a mechanism for collusion that satisfies conditions A1–A5. For instance, suppose that the cartel operates by first having each firm compute its competitive bid and then submit a bid 1.1 times its competitive bid. It is straightforward to show that conditions A1–A5 are satisfied if the cartel colludes in this fashion. In figure 1, we present a diagram that summarizes the relationship between A1–A5 and the hypothesis of competition and collusion.

As we can see from figure 1, it is in some sense never possible to reject the hypothesis of collusion by observing only z and the distribution of bids. It is always possible to construct a collusive model that satisfies A1–A5. However, if we see that A1–A5 are violated, then we know that the observed distribution of bids could not arise from a competitive model. In previous empirical studies of cartel behavior such as Porter and Zona (1993, 1999), the cartels did violate assumptions A1–A5.

B. Determining Which Firms Are in the Cartel

Even if a cartel is present in the industry, the bid functions of firms that do not collude must still satisfy a certain type of exchangeability. Suppose that the cost structure for all firms in the industry is generated as in equation (9); then the value of z_i for all the firms that do not collude must be $z_i = (\beta_1, \text{DIST}_{i,t}, \sigma)$. If the first m firms do collude, then the cost

distribution for the cartel, if it colludes efficiently as in the model of section II, will have cost parameters $z_c = (\beta_1, \text{DIST}_{1,t}, \ldots, \text{DIST}_{m,t}, \sigma)$. The bid functions of the noncolluding firms must be exchangeable in z_i holding z_c fixed. Also, they clearly must not be correlated; for, holding z_c and the z_i's fixed, the noncartel firms have not coordinated their bids. This is important in our empirical analysis, because it offers a criterion that will hold for firms that do not collude but will fail for firms that do collude.

Theorem 7. Let $\text{NC} \subseteq \{1, \ldots, N\}$ be the set of firms that do not collude. Suppose that the colluding firms jointly maximize profits as in Section 2.2. Then, conditional on z_c, the bid functions of the firms in NC are exchangeable in the z_i's and conditionally independent given z_c and the z_i's.

The theorem above demonstrates that even if collusion is occurring, those bidders that are not in the cartel still have conditionally independent and exchangeable bids. If cartel members are not sophisticated and use strategies that fail to be conditionally independent or exchangeable, we should in principle be able to identify the members of the cartel.

C. Unobserved Heterogeneity

Another challenge that the economist may face in practice is that he may not observe all relevant characteristics of the firm or of the project. Standard approaches to control for this in a panel data context are to use fixed effects for the project and fixed effects for the firms.

Suppose that the cost draws for firm i have the following form:

$$c_i = d_i + d_t + \tilde{c}_i, \tag{24}$$

where

$$\tilde{c}_i \sim F(z_i), \tag{25}$$

where d_i and d_t are constants that are publicly observable to all of the firms in the auction. In the specification above, we allow for there to be a firm fixed effect and a project fixed effect in the costs.

Suppose that when $d_i + d_t = 0$, a firm with private information bids $\mathbf{b}_i(\tilde{c}_i)$. When $d_i + d_t \neq 0$, it can easily be shown that the new equilibrium is for firm i to bid $\mathbf{b}_i(\tilde{c}_i) + d_i + d_t$. If we project our bids onto a linear econometric model with firm and project dummies, then as the number of bids becomes sufficiently large, exchangeability conditions such as (18)–(20) must continue to hold in z_i.[9]

VI. Competitive Bidding for Seal Coat Contracts

As an application of the methods introduced in the previous sections, we test for collusion using a unique data

---

[9] Of course, estimating these parameters precisely with common sample sizes is not always possible.

set of bidding by construction firms for highway repair contracts. The data set was purchased from Construction Market Data (CMD), which sells information to general contractors about upcoming construction projects. The data set contains detailed bidding information for almost all the public and private road construction projects conducted in Minnesota, North Dakota, and South Dakota during the years 1994–1998. Nearly 18,000 procurement contracts are contained in this data set.[10]

In our empirical study we focus on contracts for seal coating, which is a highway maintenance process designed to extend the life of a road. By adding oil and aggregate (sand, crushed rock, gravel, or pea rock) to the surface of a road, seal coating gives the road a new surface to wear. The added oil soaks into the underlying pavement, which also slows the development of cracks in the highway. Compared to resurfacing a highway, seal coating is a low-cost alternative.[11]

A. Contract Award Procedures

All public-sector seal coat contracts are awarded through an open competitive bidding process. In seal coat projects, contractors do not submit a single bid; rather they submit a vector of bids. This is known as a unit price contract and has the following form:

Contract item 1	Estimated quantity for item 1	Unit price for item 1
Contract item 2	Estimated quantity for item 2	Unit price for item 2
Contract item 3	Estimated quantity for item 3	Unit price for item 3
⋮	⋮	⋮

The contract items might include mobilization, gallons of oil, and tons of aggregate. Both the contract items and the estimated quantities are established by the owner of the contract (typically a city government or state DOT), and the

[10] Based on our conversations with general contractors, the Department of Transportation (DOT), and CMD, we believe that almost all construction projects exceeding $10,000 are documented in this data set. Some of the data fields were not complete, so we phoned hundreds of county and city governments throughout the midwest to fill in those missing fields. For each project, the data contains detailed information about the project location, the bid submission deadline, the identities of the bidders, the locations of the bidders, the engineer's cost estimate, and the bonding requirements.

[11] There are two main reasons why we decided to focus on the seal-coat industry: first, it is relatively easy to measure the total work done by any contractor in a seal-coating process; second, the technology involved in seal coating is simple compared to other types of highway construction. A typical crew for a seal coat company consists of two workers on the chip spreader, one distributor operator, four roller operators, four flag persons, one worker to drive a pilot car, one worker to drive the broom, and one worker to set temporary pavement markings. On a typical project there can be between 5 and 15 trucks hauling aggregate to the project site and a loader operator to fill the trucks with aggregate.

According to one company in the industry, who primarily works in the Dakotas, a typical crew (excluding trucks) costs $1,500 per day in labor and $1,000 per day in the implicit rental price for machinery. A crew can typically expect to seal coat 7 to 15 miles of highway per day, depending on conditions. The cost of trucking, according to the firm, is $35 per hour (including the driver's wage).

TABLE 1.—IDENTITIES OF MAIN FIRMS

Firm ID	Name of the Company	Firm ID	Name of the Company
1	Allied Blacktop Co.	11	Asphalt Surfacing Co.
2	Astech	12	Bechtold Paving
3	Bituminous Paving Inc.	14	Border States Paving Inc.
4	Lindteigen Constr. Co. Inc.	17	Mayo Constr. Co. Inc.
5	McLaughlin & Schulz Inc.	20	Northern Improvement
6	Morris Sealcoat & Trucking Inc.	21	Camas Minndak Inc.
7	Pearson Bros Inc.	22	Central Specialty
8	Caldwell Asphalt Co.	23	Flickertail Paving & Supply
9	Hills Materials Co.	25	Topkote Inc.

unit prices are chosen by the contractor.[12,13] If the contract is awarded, it must be awarded to the lowest responsible bidder. Public officials have the right to reject all bids, but this occurs infrequently in practice.[14]

B. Reduced Form Bid Functions

We observe all public-sector seal coat contracts awarded from January 1994 through October 1998 in our data set. There are four types of owners: city, county, state, and federal. Most of the contracts are owned by city or state governments. Among all jobs, 230 (46.5%) are owned by cities, 195 (39.3%) are owned by states, 68 (13.7%) are owned by counties, and only 2 owned by the federal government. The total value of contracts awarded in our data set is $92.8 million. The identities of the firms and their market shares are summarized in tables 1 and 2.

The size of contracts varies greatly. In table 3, we provide summary statistics for the winning bid (BID1), the second lowest bid (BID2), and the difference between the two (BID12). Of the 495 contracts in our data set, 7 contracts were awarded for more than $1 million, 256 contracts were awarded for less than $1 million but more than $100 thousand, and 232 contracts were awarded for less than $100 thousand. A total of 98 firms bid on at least one of these 495 contracts, with 43 firms never winning a contract for the period reported. Note that BID12 is on average

[12] In our analysis, we abstract away from the fact that the firms make a vector of bids instead of a single bid. As in Athey and Levin (2001), it would be possible to model bidding as a two-stage procedure. In the first stage, bidders choose a total bid, and in the second stage they choose unit prices optimally.

[13] The contractor is compensated according to the quantities that are actually used on the job. DOT personnel monitor the firm while the work occurs and are responsible for verifying measurements of quantities of material put in place. If actual quantities are 20% less or more than the estimated quantities, the price will be renegotiated according to procedures described in the contract.

[14] Firms also have strong financial incentives to honor their contractual obligations if they are the low bidders. Contractors usually must submit a bid bond of 5% to 10% of their total bid, guaranteeing that they will not withdraw their bid after the public reading of all bids. After the contract is awarded, the low bidder must submit a performance bond and a pay bond to guarantee the completion of the contract and that all subcontractors will be paid. For a more complete discussion of contract procedures see Minnesota Department of Transportation (1995), Bartholomew (1998), Clough and Sears (1994), and Hinze (1993).

TABLE 2.—BIDDING ACTIVITIES OF MAIN FIRMS

Firm ID	No. of Wins	Avg. Bid	% Mkt. Share	No. Bids Submitted	% Participation
1	92	82,790	8.2	145	29.3
2	102	191,953	21.1	331	66.9
3	20	363,565	7.8	69	14.0
4	35	241,872	9.1	114	23.0
5	29	283,323	8.9	170	34.3
6	40	77,423	3.3	84	17.0
7	45	62,085	3.0	121	24.4
8	16	87,231	1.5	134	27.1
9	10	237,408	2.6	14	2.8
11	4	328,224	1.4	28	5.7
12	3	317,788	1.0	8	1.6
14	4	754,019	3.2	25	5.1
17	5	1,018,578	5.5	8	1.6
20	13	355,455	5.0	38	7.7
21	2	903,918	1.9	5	1.0
22	2	903,953	2.0	8	1.6
23	2	439,619	1.0	4	0.8
25	3	382,012	1.2	13	2.6
Total	427		87.7		

Average bid is the average of all bids that a particular firm submitted, no. bids submitted is the total number of bids that the firm submitted, and % participation is the fraction of seal coat contracts the firm bid for.

TABLE 3.—FIRST AND SECOND LOWEST BIDS

	Observations	Mean	Std. Dev.	Min.	Max.
BID1	466	191,355	227,427	3,893	1,772,168
BID2	466	207,079	244,897	4,679	1,959,928
BID21	466	15,724	29,918	33	352,174

TABLE 4.—BID CONCENTRATION

Number of bids	1	2	3	4	5	6	7
Number of contracts	29	87	190	118	44	22	5

15,724, suggesting that bidders leave "money on the table" due to asymmetric information.

Table 4 summarizes the distribution of the number of bids per contract. The modal number of bids in this industry is 3, so taking account of market power will clearly be important.

The owner of the largest firm in the market, Astech (firm 2), received a one-year prison sentence for bid rigging in the mid-1980s. The owners of two other firms, McLaughlin & Schulz Inc. (firm 5) and Allied Paving (firm 1), were also fined for bid rigging with Astech in the seal coat industry. The owners of all three firms were, at one time, banned from bidding for public-sector seal coat contracts.

One important control variable for our analysis will be the engineer's estimate. This is a cost estimate formed either by civil engineers employed by the government or by consulting engineering firms. The engineer's estimate is supposed to represent a "fair market value" for completion of the project.[15] We found that estimates were available for 139 out of the 441 projects in the data set. Table 5 shows that the engineer's estimate is a useful control for project costs. The normalized winning bid (winning bid divided by the estimate) is almost exactly 1 and has a standard deviation of 0.1573.

Another generated variable is distance, which we construct using information about both the location of the firms and the location of the project.[16] For jobs covering several

locations, we use the midpoints of the jobs to do the calculation. Table 6 summarizes the distances of firms based on the rank of their bid, that is, DIST1 is the distance of the low bidder, DIST2 is the distance of the second lowest bidder, and so on. Firms with shorter distances from project locations are more likely to win the job, because they will have lower transportation costs.

Based on the winning bids and bidding dates, we construct a new variable CAP, which is meant to measure each firm's capacity utilization level. A firm's capacity at a particular bidding time is defined as the ratio of the firm's used capacity (measured by the firm's total of winning bids up to that time) to the firm's total of winning bids in the entire season.[17]

A final determinant of firm i's success in winning contracts is familiarity with local regulators and local material suppliers. We summarize the concentration of selected firms by state in table 7. Our results suggest that the majority of the firms in our data set work primarily in one state. For instance, firm 3 is located near the boundaries of Minnesota, North Dakota, and South Dakota. Yet it does over 70% of its dollar volume of seal coating in South Dakota. Firm 6 is located near the Minnesota–South Dakota border, yet it has won no contracts in South Dakota.

Next, we estimate a set of reduced-form bid functions to measure the relationship between a number of variables and the firms' observed bidding behavior. The variables we will use in these regression are as follows:

- $BID_{i,t}$: Amount bid by firm i on project t.
- EST_t: Engineer's cost estimate for project t.
- $DIST_{i,t}$: Distance between the location of the firm and the project.
- $LDIST_{i,t}$: $\log (DIST_{i,t} + 1.0)$.
- $CAP_{i,t}$: Used capacity measure of firm i on project t.
- $MAXP_{i,t}$: Maximum percentage free capacity of all firms on project t, excluding i.
- $MDIST_{i,t}$: Minimum of distances of all firms on project t, excluding i.

[15] In conversations with engineers at Minnesota's, North Dakota's, and South Dakota's Departments of Transportation, the engineers stated that they formed the estimates by gathering information on materials prices, prevailing wage rates, and other relevant cost information.

[16] The calculation is facilitated by using Yahoo's map searching engine http://maps.yahoo.com/py/ddResults.py. Using city and state's names as input for both locations, the engine gives distances automatically. Doing this manually would be too time-consuming, so we wrote an "electronic spider" to do the job.

[17] The season during which seal coating can take place lasts from late May to mid-September; in our definition, the entire season starts on September 1 and ends on August 31 of the following calendar year. This measure of capacity was computed using the entire database of bidding information, even though in our econometric analysis we will focus on a subset of these projects.

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

- LMDIST$_{i,t}$: log (MDIST$_{i,t}$ + 1.0).
- 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.

Summary statistics for these variables are reported in table 5.

We assume that firm i's cost estimate for project t satisfies the following structural relationship:

$$\frac{c_{i,t}}{\text{EST}_t} = c(\text{DIST}_{i,t}, \text{CAP}_{i,t}, \text{CON}_{i,t}, \omega_i, \delta_t, \epsilon_{it}). \qquad (26)$$

Equation (26) implies that firm i's cost in auction t can be written as a function of its distance to the project, its backlog, the previous experience that firm i has in this market (which we proxy for using CON$_{i,t}$), a firm i productivity shock ω_i, an auction-t-specific effect δ_t, and ϵ_{it}, an idiosyncratic shock to firm i that reflects private information

TABLE 6.—DISTANCES (IN MILES)

	Mean	Min	Max		Mean	Min	Max
DIST1	122.3	0	584.2	DIST5	160.3	13	555.2
DIST2	151.9	0	585.2	DIST6	177.9	63	484.4
DIST3	177.9	0	637.6	DIST7	91	44	128.9
DIST4	166.4	11.2	608.6				

TABLE 7.—CONCENTRATION OF FIRM ACTIVITY BY STATE

	Concentration		
Firm	MN	ND	SD
1	1	0	0
2	0.2781	0.7218	0
3	0	0.2377	0.7623
4	0	1	0
5	0.1246	0.5338	0.3414
6	0.8195	0.1804	0
7	0.9572	0.0427	0
8	0.7290	0.2709	0
11	0	0	1
14	0	1	0
20	0	1	0

it will have about its own costs. The results of section II demonstrate that under certain simplifying assumptions about dynamic competition, a dynamic model with capacity-constrained bidders is equivalent to a static model where a firm's cost is $c_i + V_{iL}(s) - V_{iW}(s)$: a sum of current project costs c_i plus a term $V_{iL}(s) - V_{iW}(s)$ that captures the option value of keeping free capacity. In practice, the measure of backlog CAP$_{i,t}$ will be a good proxy for $V_{iL}(s) - V_{iW}(s)$. Mapping the structural cost function back to the framework of section IV implies that $z_i = (\text{DIST}_{i,t}, \text{CAP}_{i,t}, \text{CON}_{i,t}, \omega_i, \delta_t)$.

Firm i's bid function should depend on the entire parameter vector $z = (z_1, \ldots, z_N)$. However, given the limited number of data points in our sample, it will not be possible to model the bid functions in a completely flexible fashion, because z is a vector with many elements. We choose to include a firm's own distance, capacity, and concentration. From our conversations with firms that actually bid in these auctions, we believe that the most important characteristics of the other firms to include in the reduced-form bid function are the location of the closest competitor and the backlog of the competitor that has the most free capacity. Also, we computed and simulated the equilibrium of the asymmetric auction model, using the techniques developed by Bajari (2001). These simulations also suggest a similar specification is appropriate. To control for δ_t, we use fixed effects for the auction, and to control for ω_i, we use firm fixed effects for the largest 11 firms in the market. We are able to identify both our auction fixed effect and firm fixed effects because we do not use fixed effects for all of the firms. This implies that firms that are not the 11 largest have an identical productivity shock ω_i, which is probably not an unrealistic assumption in this industry.

Since there are 138 auctions, 11 main firms, and one pooled group of nonmain firms in our restricted data set, we have 137 auction dummies and 11 firm dummies.[18] The set of regressors thus contains a constant (C), 148 dummy variables, own distance (DIST$_{i,t}$), own capacity (CAP$_{i,t}$), maximal free capacity

[18] One auction with an abnormal bid was removed from our data set.

among competitors ($\text{MAXP}_{i,t}$), minimal distance among competitors ($\text{MINDIST}_{i,t}$), and the job concentration variable ($\text{CON}_{i,t}$). To take care of the heteroskedasticity problem, we take the ratio of the bid and the value (the engineer's estimate) as the dependent variable ($\text{BID}_{i,t}/\text{EST}_t$):

$$\frac{\text{BID}_{i,t}}{\text{EST}_t} = \beta_0 + \beta_1 \text{LDIST}_{i,t} + \beta_2 \text{CAP}_{i,t} + \beta_3 \text{MAXP}_{i,t}$$
$$+ \beta_4 \text{LMDIST}_{i,t} + \beta_5 \text{CON}_{i,t} + \epsilon_{it}. \qquad (27)$$

The results from the regression, estimated using ordinary least squares, are displayed in table 8.

The results from our reduced-form bid function are consistent with basic economic intuition. Firm i's bid is an increasing function of firm i's distance from the project site and firm i's capacity utilization. As firm i's distance increases, so does i's cost. So the positive coefficient on own distance is consistent with the model of competitive bidding in section II. Note that the coefficient on $\text{CAP}_{i,t}$ is also positive and significant. As firm i's backlog increases, all else held constant, the option value of free capacity will increase because once i becomes completely capacity-constrained, firm i will no longer have a chance to bid on future projects. The coefficient on $\text{CON}_{i,t}$ is negative, indicating that if firm i has more prior experience in the state, firm i will tend to bid more aggressively.

Our reduced-form bid function also produces results that are consistent with the strategic interactions implied by the asymmetric auction model. As the distance of firm $j \neq i$ increases or as the capacity utilization of firm $j \neq i$ increases, competition will soften and firm i will raise its bid. However, the reaction to $\text{MAXP}_{i,t}$ is not significant at conventional levels.

C. Testing Conditional Independence

In this section, we test the conditional independence assumption A1 in section IV. We use a reduced-form bid function as in the previous subsection; however, we will

TABLE 8.—REDUCED-FORM BID FUNCTION

Variable	OLS
C (constant)	0.6809
	(5.95)
$\text{LDIST}_{i,t}$ (own distance)	0.0404
	(3.45)
$\text{CAP}_{i,t}$ (own used capacity)	0.1677
	(8.51)
$\text{MAXP}_{i,t}$ (maximal free capacity among rivals)	0.0255
	(0.713)
$\text{LMDIST}_{i,t}$ (minimal distance among rivals)	0.0240
	(1.81)
$\text{CON}_{i,t}$ (job concentration)	−0.0590
	(−1.866)
Sample size	450
R^2	0.8480

The regression also includes a fixed effect for each project t and one for each of the largest 11 firms in the market.

allow the model to be more flexible. If firm i is one of the largest 11 firms in the industry, we use equation (28) with firm-varying coefficients as its bid function. If firm i is not one of the largest 11 firms in the industry, we use equation (29) to model its bid function. We pool equations (28) and (29) in the estimation and include auction fixed effects:[19]

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

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

Suppose the coefficient of correlation between the residual to firm i's bid function and firm j's bid function, $\epsilon_{i,t}$ and $\epsilon_{j,t}$, is ρ_{ij}. The test of conditional independence is then equivalent to the test of the following null hypothesis:

$$H_0 : \rho_{ij} = 0. \qquad (30)$$

We first report the number of pairwise simultaneous bids and the correlation coefficients (computed when the number of simultaneous bids is no less than 4) in table 9. Simultaneous bids are reported in the lower part of the matrix, and the correlation coefficients in the upper part. We use the Fisher test to test the hypothesis (30). Suppose the correlation coefficient between two firms' bids is ρ. Let r be the correlation coefficient calculated from sample data (as reported in table 9); then the Fisher Z transformation is given by

$$Z = \frac{1}{2} \ln \frac{1 + r}{1 - r}. \qquad (31)$$

Let n be the number of samples; then the distribution of Z is approximately normal with

$$\mu_Z = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} \quad \text{and} \quad \sigma_Z = \frac{1}{\sqrt{n - 3}}. \qquad (32)$$

Hence $z = (Z - \mu_Z)\sqrt{n - 3}$ has approximately the standard normal distribution. In our case, under the null hypothesis, $\rho = 0$, $\mu_Z = 0$. The test statistic is $Z\sqrt{n - 3}$ for each pair of firms whenever $n > 3$. The results are reported in table 10.

Among all 23 pairs which have at least four simultaneous bids, the null hypothesis cannot be rejected except for four pairs of firms at 5% significance level. These four pairs are (firm 1, firm 2), (firm 2, firm 4), (firm 5, firm 14), and (firm 6, firm 7). However, of these pairs, only the pair (firm 2,

[19] Note that in equation (28) we force the coefficient on concentration to be equal for all firms in our sample. This is because there was not sufficient variation in concentration within a single firm to identify a firm-specific parameter for this coefficient.

TABLE 9.—SIMULTANEOUS BIDS AND COEFFICIENTS OF CORRELATION

Firm	1	2	3	4	5	6	7	8	11	14	20
1		-.744									
2	15		-.5897	-.5247	-.1512	.1330	-.3010	.0909		.4260	.1304
3	0	9		-.6374	.2439				-.2345		
4	0	67	4		-.1910					-.3197	
5	0	76	8	63		-.3365		.5742	.8854	-.6963	.3588
6	1	17	3	3	8		-.7850	.2327			
7	2	9	3	0	3	7		-.2711			
8	2	12	3	2	5	12	6				
11	1	2	7	0	4	0	0	0			
14	0	9	0	8	10	0	0	0	0		.5768
20	0	5	1	2	5	1	1	1	0	6	

TABLE 10.—FISHER TEST FOR CONDITIONAL INDEPENDENCE

Firms	n	r	z	Firms	n	r	z
(1, 2)	15	-.744	-3.3234	(4, 5)	63	-.1910	-1.4979
(2, 3)	9	-.5897	-1.6588	(4, 14)	8	-.3197	-0.7408
(2, 4)	67	-.5247	-4.6624	(5, 6)	8	-.3365	-0.7829
(2, 5)	76	-.1512	-1.3018	(5, 8)	5	.5742	0.9246
(2, 6)	17	.1330	0.5006	(5, 11)	4	.8854	1.4002
(2, 7)	9	-.3010	-0.7609	(5, 14)	10	-.6963	-2.2756
(2, 8)	12	.0909	0.2734	(5, 20)	5	.3588	0.5310
(2, 14)	9	.4260	1.1145	(6, 7)	7	-.7850	-2.1165
(2, 20)	5	.1304	0.1855	(6, 8)	12	.2327	0.7111
(3, 4)	4	-.6374	-0.7538	(7, 8)	6	-.2711	-0.4816
(3, 5)	8	.2439	0.5566	(14, 20)	6	.5768	1.1391
(3, 11)	7	-.2345	-0.4779				

firm 4) bid against each other more than a handful of times. The pairs (firm 1, firm 2), (firm 5, firm 14), and (firm 6, firm 7) bid against each other on average no more than two or three times a year in the data set.

D. Test for Exchangeability

In this section we use our regression model (28) and (29) to test whether the empirical distribution of bids is exchangeable. Exchangeability implies that capacities and distances should enter the firm's bidvalue function in a "symmetric" way. Formally, in the reduced-form bid function, let β_{i1}, β_{i2}, β_{i3}, β_{i4} be the coefficients of LDIST1, CAP1, MAXP, LMDIST for firm i, one of the largest 11 firms. Then exchangeability is equivalent to the following hypothesis:

$$H_0 : \beta_{ik} = \beta_{jk} \text{ for all } i, j, i \neq j, \text{ and for all } k = 1, \ldots, 4. \tag{33}$$

We use the F-test to test for exchangeability. Let SSR_U and SSR_C be the sums of squared errors in the unconstrained and constrained models, respectively. Also let T be the number of observations ($T = 450$ in our data set), m be the number of regressors, and n be the number of constraints implied by H_0. Then the statistic

$$F = \frac{(\mathrm{SSR}_C - \mathrm{SSR}_U)/n}{\mathrm{SSR}_U/(T - m)} \tag{34}$$

has an F-distribution with parameters $(n, T - m)$ under the null hypothesis. Note that the F-test is also a variation of the quasi-likelihood-ratio (QLR) test on nonlinear two- and three-stage least squares.

We conduct two tests of exchangeability in this subsection. The first set is to test exchangeability for the whole market, that is, the constrained regression that pools all the 11 main firms together. The second set is to test the exchangeability on a pairwise basis, that is, the constrained regression pools two of the main firms together at each test (hence the number of constraints is 4). That is, we test whether empirically exchangeability holds on a pairwise basis for the firms in our data set. We perform this set of tests for each pair of firms with at least four simultaneous bids. Table 11 summarizes the test results. It shows that for

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

almost all the tests, we just fail to reject the null hypothesis at 5% significance level. In fact, we only reject the null when we pool all the 11 main firms and when we pool firm 2 and firm 5.

E. Discussion

The results of our tests of exchangeability and conditional independence imply that there are five pairs of firms that exhibit bidding patterns that are not consistent with our characterization of competitive bidding. The four pairs (firm 1, firm 2), (firm 2, firm 4), (firm 5, firm 14), and (firm 6, firm 7) fail the conditional independence test, and the pair (firm 2, firm 5) fails the exchangeability test. However, of these pairs, only the pairs (firm 2, firm 4) and (firm 2, firm 5) bid against each other more than a handful of times. The other three pairs—(firm 1, firm 2), (firm 5, firm 14), and (firm 6, firm 7)—bid against each other on average no more than two or three times a year in the data set. Also, according to industry participants, these firms function in different submarkets and would have no reason to view each other as principal competitors. Therefore, we take the view that the firms we should be most concerned about colluding are (firm 2, firm 5) and (firm 2, firm 4).

Overall, bidding in this industry appears to conform to the axioms A1–A5. This observation is important to policymakers since there is a history of bid rigging in the seal coat industry. Several of the largest firms in the industry were colluding in the early to mid-1980's and paid damages for bid rigging. Our analysis suggests that currently most bidding behavior in the industry is consistent with our model of competitive bidding.

We are aware of the following three limitations to our approach. First, our tests for conditional independence and exchangeability depend on the functional form for the reduced-form bid functions. In our analysis, we used a number of different functional forms for the reduced-form bid function, and the results in the tests for conditional independence and exchangeability were robust across these alternative specifications. Firm i's bid function depends on the vector z, which has a large number of elements. Given that we have only 138 auctions in our data set, we can never be certain that our independence and exchangeability results were not influenced by a poor choice of functional form.

Second, our results might be incorrect if there are omitted variables. If there are elements of z that the firms see but that are not present in our data set, our regression coefficients will be biased. In our analysis, we use fixed effects for each contract and for the largest 11 firms. Therefore, we should be most worried about omitted variables that are elements of z, but not collinear with the firm or contract fixed effects. This could happen, for instance, when there are three firms among which firm 1 and firm 2 always use a quote from a particular subcontractor when computing their cost estimates while firm 3 does not. If this quote is publicly observed, it will then induce positive correlation between

the residuals to the bid functions of firms 1 and 2.[20] A similar critique could be made to our test of exchangeability. Omitted cost variables could lead us to falsely conclude that firm 2 and firm 5 fail to have an exchangeable distribution of bids.[21]

Third, if a sophisticated cartel is operating in this market, then, as we mentioned in section V, the cartel could satisfy assumptions A1–A5 by generating phony bids in a clever fashion. Therefore, from our tests, we shall not be able to identify whether those firms who passed the conditional independence and exchangeability tests are competitive or are smart colluders. In recent empirical papers that document cartel behavior, such as Porter and Zona (1993, 1999) and Pesendorfer (2000), the authors know from court records and investigations the identity of the cartel. In all these papers, both exchangeability and conditional independence fail. To the best of our knowledge, there is no documented case of cartel bidding where the cartel intentionally submitted phony bids that were still both conditionally independent and exchangeable.

Therefore, although collusion is certainly one reason why the tests of conditional independence and exchangeability fail, it is certainly not the only reason. The restrictions of the competitive bidding model are quite stringent and there are a variety of reasons, other than collusion, for our failure to accept the competitive model. Therefore, we shall interpret firm pairs (2, 4) and (2, 5) as merely a candidate set of cartels.

In interpreting the results of tests of exchangeability and conditional independence, it is important for the economist to exercise judgement about whether the results are economically significant, not just statistically significant. For instance, if the residuals to the bid functions have a correlation coefficient of 0.8, we should be more concerned that the competitive model is not working than if the coefficient is 0.2. As we discussed in section V C, the forms of unobserved heterogeneity that we can allow for in our reduced-form tests are quite limited and must enter in a

---

[20] However, in our test for conditional independence, we found that the residuals between firms 2 and 4 are negatively correlated. If this is due to omitted cost variables, then the omitted variables must induce negative correlation between the costs of these two firms. So far, we have not been able to come up with a scenario that would generate this type of cost shock. However, if firm 2 and firm 4 engaged in a scheme of submitting phony bids, this might induce a negative correlation in the residuals, since phony bidding implies that when firm 2 bids high firm 4 must bid low.

[21] In deciding on the correct covariates, we worked closely with contractors in the industry. In our first set of estimates, we did not include $CON_{i,t}$, and we noticed a large number of failures of conditional independence and exchangeability. We then spoke with contractors and explained, in layman's terms, the pattern of correlation among the residuals. The contractors immediately suggested that we had failed to control for the importance of being familiar with state-level regulations, which suggested constructing the variable $CON_{i,t}$. The tests for exchangeability and conditional independence therefore helped us to learn about the correct cost variables that need to be included in our regressions. The contractors that we spoke with were not able to suggest an omitted variable that could explain the observed bidding patterns of the pairs (firm 2, firm 5) and (firm 2, firm 4).

linear fashion. If our model of unobserved heterogeneity is not sufficiently rich, then a naive interpretation of hypothesis tests may lead to falsely accusing firms of collusion. The economist should then consult with industry experts to make sure that these failures of the competitive model are not due to his ignorance of the industry cost structure.

VII. Structural Econometric Models

In this section, we describe how to estimate and decide between three alternative structural models of industry equilibrium. We shall refer to the three models as $M1$, $M2$, and $M3$. In the first model all firms determine their bids competitively as in the model outlined in sections II and III. In the second model, firms 2 and 4 collude in an efficient manner. Before the bidding begins, firm 2 and firm 4 compute cost estimates $c_{2,t}$ and $c_{4,t}$ respectively for project t. The cartel operates efficiently, so that only the firm with the lowest cost submits a bid. That is, the cost to the cartel is $c_{c,t} = \min \{c_{2,t}, c_{4,t}\}$. The other cartel member either refrains from bidding or submits a phantom bid. The cost distribution for the cartel is the minimum of $c_{2,t}$ and $c_{4,t}$. This distribution has a probability density function $f_2(c_{c,t}|z, \theta)$ $[1 - F_4(c_{c,t}|z, \theta)] + f_4(c_{c,t}|z, \theta)$ $[1 - F_2(c_{c,t}|z, \theta)]$. In model $M2$, we can treat the cartel as a single firm that has cost $c_{c,t}$. In computing expected utility, it is sufficient for noncartel firms to concern themselves only with the low bid among all of the cartel members. Therefore, as we mentioned in section IIB, efficient cartels are a special case of our asymmetric auction model. The third model, $M3$, is analogous to $M2$, with firm 2 and firm 5 being the cartel members.

We believe that the assumption of an efficient cartel is reasonable for the seal coat industry in the midwest. According to industry officials and insiders we have spoken with, before the early 1980s, collusion occurred (and was prosecuted) in the industry. Firms frequently made side payments to each other. Many side payments were made directly in cash, but some were through the use of falsified invoices. For instance, firm A would rent equipment from firm B on paper. However, the equipment would not actually be rented. The payment that firm A made to firm B for this phantom equipment rental would serve as a side payment.[22,23]

Let $g_i(b; \theta, z, M)$ be the pdf of firm i's equilibrium bids, and $G_i(b; \theta, z, M)$ be the cdf in the model $M \in \{M1, M2,$

[22] There are many possible models of collusion. For instance, the economist could include the fear of detection in his model of collusion. Ultimately, the choice of which model of collusion is appropriate must be based on an analysis of conditions within a particular industry.

[23] Another assumption of the model is that collusion is common knowledge among industry participants. We believe that this is a reasonable assumption in the seal coat industry and many other industries. Some contractors we spoke with claimed that they entered the industry in the early 1980s because they knew it was collusive and hence profitable for them to enter. Similarly, in the New York construction industry as described by Porter and Zona (1993), the industry participants seemed to be quite aware of the collusion that was taking place.

$M3\}$. In the competitive model $M1$, each firm makes a best response to the distribution of the bids of all other firms. In the collusive model $M2$ or $M3$, each noncartel firm needs to make a best response only to the minimum bid of the cartel and to the bids of all the other noncartel firms.

If we assume that the firms in the industry are profit-maximizing, then it can be shown that, as in Guerre et al. (2000), by rewriting (7), firm i's private cost c_i^M in model M must satisfy

$$
c_i^M = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{g_j(b; \theta, z, M)}{1 - G_j(b; \theta, z, M)}}. \tag{35}
$$

This relationship must hold for models $M = M1$, $M2$, and $M3$.

Let $\hat{g}_i(b; \theta, z, M)$ and $\hat{G}_i(b; \theta, z, M)$ denote the estimated values of $g_i(b; \theta, z, M)$ and $G_i(b; \theta, z, M)$. Let $c_{i,t}^{M,\mathrm{EST}}$ denote an estimate of the cost of firm i in project t, and let $\mathbf{c}^{M,\mathrm{EST}}$ denote the estimated vector of the $c_{i,t}^{M,\mathrm{EST}}$. Following Guerre et al. (2000), we estimate firm i's private cost by evaluating equation (35) using $\hat{g}_i(b; \theta, z, M)$ and $\hat{G}_i(b; \theta, z, M)$:

$$
c_{i,t}^{M,\mathrm{EST}} = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{\hat{g}_j(b; \theta, z, M)}{1 - \hat{G}_j(b; \theta, z, M)}}. \tag{36}
$$

In our problem, the distribution $g_i(b; \theta, z, M)$ depends on a large number of parameters and covariates. Obviously, kernel estimation of $g_i(b; \theta, z, M)$ is not practical in our application. We begin by estimating equations (28) and (29). As we mentioned previously, in the case of $M2$ we estimate the cartel's bid as the minimum of the bids of firm 2 and firm 4, and in the case of $M3$, the cartel's bid is the minimum of the bids of firm 2 and firm 5. In all three models, we shall assume that the distribution of firm i's bid satisfies

$$
b_{i,t} = (\text{OLS fitted value of } b_{i,t}) + \epsilon_{i,t},
$$

where $\epsilon_{i,t}$ is modeled as a mixture of four normal distributions. We use the mixture-of-normals specification because it allows us to flexibly model the distribution of $b_{i,t}$ and to compute $\hat{g}_i(b; \theta, z, M)$ and $\hat{G}_i(b; \theta, z, M)$ in a simple manner. To estimate the mixture of normal distributions we used maximum likelihood estimation. The estimated distributions of the residuals in $M1$, $M2$, and $M3$ appear to closely match the empirical distributions of the residuals.

In general, the vector of costs $\mathbf{c}^{M,\mathrm{EST}}$ cannot be estimated exactly. The econometrician will have uncertainty about the value of $\mathbf{c}^{M,\mathrm{EST}}$. A joint density of $\mathbf{c}^{M,\mathrm{EST}}$ can be derived using standard econometric methods, and we shall refer to this density as $p_{\hat{c}}(\mathbf{c}^{M,\mathrm{EST}})$.

VIII. Testing For Collusion: A Bayesian Framework

In the previous section, we demonstrated how to estimate the joint distribution of the latent costs $p_{\hat{c}}(c^{M,\text{EST}})$ for three alternative models of industry equilibrium. In this section, we will suggest a Bayesian framework for computing a posterior probability for each of our three models.

Each of our three models of industry equilibrium implies a different distribution over structural cost parameters. Let $p_c(c_{i,t}|z_t, \theta_i, M)$ for $M \in \{M1, M2, M3\}$ denote the probability density function for the cost distribution of firm i in auction t conditional on covariate z_t, parameters θ_i, and model M. For instance, in the collusive model $M2$, the cost distribution for the cartel is $f_2(c_{c,t}|z, \theta)[1 - F_4(c_{c,t}|z, \theta)] + f_4(c_{c,t}|z, \theta)[1 - F_2(c_{c,t}|z, \theta)]$. In model $M3$, the probability density function of the cost distribution of the cartel is $f_2(c_{c,t}|z, \theta)[1 - F_5(c_{c,t}|z, \theta)] + f_5(c_{c,t}|z, \theta)[1 - F_2(c_{c,t}|z, \theta)]$. Using the fact that bids are independent across auctions, it is straightforward to compute $p(c^{M,\text{EST}}|z_t, \theta, M)$, the probability density function for all the costs in all the auctions:

$$p_{c^{M,\text{EST}}}(c^{M,\text{EST}}|z, \theta, M) = \prod_{t=1}^{T} \prod_i p_c(c_{i,t}|z_t, \theta_i, M). \qquad (37)$$

In Bayesian statistics, it is necessary to specify a prior distribution over structural parameters. As we shall explain below, we elicit the prior distribution of the parameters, $p_\theta(\theta)$, from industry experts. Given the likelihood function for each model (37), we define the marginalized likelihood of model M as follows:

$$\text{ML}_M = \iint p_{c^{M,\text{EST}}}(c^{M,\text{EST}}|z, M, \theta)p_\theta(\theta)p_{\hat{c}}(c_T^{M,\text{EST}}) \qquad (38)$$

That is, the marginalized likelihood is the expected value of the likelihood function after marginalizing out the distribution of θ and the distribution of $c^{M,\text{EST}}$.

In a Bayesian framework, we must specify a prior probability for each model. Let $p_M(M_i)$ denote the prior belief (probability) that $M = M_i$, and let ML_i denote the value of likelihood function conditional on $M = M_i$. Then according to Bayes's theorem, we can compute the posterior probability of $M = M_i$, which we denote by $p^*(M_i)$, as follows:

$$p^*(M_i) = \frac{p_M(M_i) \cdot \text{ML}_i}{p_M(M_1) \cdot \text{ML}_1 + p_M(M_2) \cdot \text{ML}_2 + p_M(M_3) \cdot \text{ML}_3} \qquad (39)$$

In the following section, we describe how we elicit the prior distribution over structural cost parameters $p(\theta)$ from industry experts. Given this distribution and a (flexible) functional form for the likelihood $p_{c^{M,\text{EST}}}(c^{M,\text{EST}}|z, \theta, M)$, to compute the posterior probability of each model, we only need to evaluate the integral (38) and apply Bayes's theorem as in equation (39).

A. Eliciting Prior Beliefs

In this section, we specify a flexible functional form for the distribution of $p_{c^{M,\text{EST}}}(c_{i,t}|z_t, \theta_i, M)$, and we demonstrate how we elicit a prior distribution of beliefs $p_\theta(\theta)$ from industry experts. We will model the cost (normalized by the estimate) as follows:

$$\frac{c_{i,t}}{\text{EST}_t} = \frac{\bar{c}_{i,t}(\theta, z_t)}{\text{EST}_t} + \eta_{it}. \qquad (40)$$

In equation (40), the normalized cost is a function of a fitted value $\bar{c}_{i,t}(\theta, z_t)$ which depends on the structural parameters θ, a set of auction t covariates z_t, and an idiosyncratic error term η_{it}. In order to model the error distribution flexibly, we allow it to be a mixture of four normal distributions, that is,

$$\eta_{it} \sim \pi_1 N(\mu_1, \sigma_1) + \pi_2 N(\mu_2, \sigma_2) + \pi_3 N(\mu_3, \sigma_3) + \pi_4 N(\mu_4, \sigma_4). \qquad (41)$$

What remains to be specified is how to use the expert information to compute $\bar{c}_{i,t}(\theta, z_t)$, the fitted value of the cost, and how to determine the vectors of parameters $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$, $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$, and $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$.

To construct our prior distribution p_θ, we elicited the views of two experienced bidders in the market on the distribution of markups. Using this prior over markups, we induce a prior over the structural cost parameters θ. We believe that these experts (both of which are among our 11 large firms) should have fairly reliable information about markups in the industry. At state DOT bid lettings, contractors typically congregate at a single hotel. The Associated General Contractors and the DOT circulate a list of plan holders for each projects, that is, the set of firms that have asked to receive the contract and the bidding documents from the DOT. Also, a list of contractors' room numbers is circulated. The materials suppliers and the subcontractors will then either telephone the contractors or knock on the contractors' doors and deliver quotes to contractors who are plan holders for a particular project. As a result, the contractors have a fairly good idea about their own costs and the costs of their competitors for materials. Also, labor costs and bonding costs are fairly well known throughout the industry, because contractors purchase these services in the same market.

Another factor that allows contractors to learn about their competitors' markups is the public reading of bids. Shortly after the deadline for submitting bids, all of the bids are posted on DOT Web pages, and the contractors can receive faxed copies of the bids. Contractors typically study the unit prices bid by competing contractors with great interest. There are a large number of small, fringe firms in the construction industry, because entry costs are quite low. Contractors who are unable to learn about costs and manage their own costs effectively will simply not survive in this

market. As a result, we believe that successful bidders will have very useful information about the overall distribution of markups.[24]

Let $m_{i,t}$ be the markup of player i in auction t, and let $p_m(m_{i,t})$ be the prior density function. After extensive discussions with the experts, we elicited their beliefs about the 25th, 50th, 75th, and 99th percentiles of the distribution of markups of auctions in our sample. The views of the experts were quite close, and we average them in the equations below:

$$25^{th} \text{ percentile} = 3\%,$$

$$50^{th} \text{ percentile} = 5\%,$$

$$75^{th} \text{ percentile} = 7\%, \tag{42}$$

$$99^{th} \text{ percentile} = 15\%.$$

We assume that markups were independently and identically distributed; then (42) implies the following distribution:

$$p_m(m_{i,t}) = \begin{cases} 8.3333 & \text{if } 0 \quad < m_{i,t} < 0.03, \\ 12.5 & \text{if } 0.03 < m_{i,t} < 0.05, \\ 12.5 & \text{if } 0.05 < m_{i,t} < 0.07, \\ 3.125 & \text{if } 0.07 < m_{i,t} < 0.15. \end{cases} \tag{43}$$

Given the prior over markups, we now describe how to induce a prior distribution over θ, the structural cost parameters. If the markup on project t is $m_{i,t}$ and the observed bid in project t is $b_{i,t}$, it follows that the cost $c_{i,t}$ must satisfy

$$c_{i,t} = (1 - m_{i,t})b_{i,t}. \tag{44}$$

Next, given the probability density function $p_m(m_{i,t})$ in equation (43), draw a random vector of markups $\mathbf{m}_T$ for all projects $t = 1, \ldots, T$ and all bidders i who bid on the tth project. Then using equation (44) we can compute a set of latent costs $c_{i,t}$. We then estimate the following equations using OLS to find a fitted value $\bar{c}_{i,t}(\theta, z_t)$ for all i and t:

$$\frac{c_{i,t}}{EST_t} = \beta_{0,i} + \beta_{1,i} \text{ LDIST}_{i,t} + \beta_{2,i} \text{ CAP}_{i,t} + \beta_{3,i} \text{ LMDIST}_{i,t}$$
$$+ \beta_4 \text{ CON}_{i,t} + \epsilon_{it}, \tag{45}$$

[24] Other sources that might have information about markups are consulting engineering firms and bonding companies. Consulting engineers prepare cost estimates for projects and need to keep abreast of current market price for labor and major inputs in order to construct cost estimates reliably. Bonding companies are liable for the completion of the project if the contractor goes bankrupt. The bonding companies, as a result, frequently ask contractors to turn in daily profit-and-loss statements for each project that they bond and therefore have fairly detailed knowledge of contractor profits in the industry. (If a contractor is found to be very dishonest with a bonding company, it will be unlikely that he will be able to get bonds in the future and hence be able to bid for public contracts.) We chose to use information elicited from the contractors because we thought that they were probably the most reliable among the three sources.

$$\frac{c_{i,t}}{EST_t} = \alpha_0 + \beta_1 \text{ LDIST}_{i,t} + \alpha_2 \text{ CAP}_{i,t} + \alpha_3 \text{ LMDIST}_{i,t}$$
$$+ \alpha_4 \text{ CON}_{i,t} + \epsilon_{it}. \tag{46}$$

We include firm- and auction-specific fixed effects in order to capture unobserved heterogeneity across projects and firms. Just as in our estimation of the bid functions in equations (28) and (29), we use equation (45) for the main firms, allowing the coefficients to be firm-specific, and equation (46) for the nonmain firms. From equations (45) and (46) we can also find an estimated value $\hat{\sigma}$ for the standard deviation of ϵ_{it}. We must next compute the value of $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)$, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$. We shall assume that the parameters of the mixture of normals are drawn uniformly from all of those distributions satisfying

$$\pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3 + \pi_4\mu_4 = 0,$$

$$\pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_3\sigma_3^2 + \pi_4\sigma_4^2 = \hat{\sigma}^2,$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1, \tag{47}$$

$$-0.2 \le \mu_i \le 0.2,$$

$$0.01 \le \sigma_i \le 0.2.$$

We assume that $-0.2 \le \mu_i \le 0.2$ and $0.01 \le \sigma \le 0.2$, so that the supports of the μ_i and σ_i are compact. We believe that this assumption is appropriate because we have found that the fitted residuals from equations (45) and (46) fall outside the interval $(-0.2, 0.2)$ very infrequently. We also found that $\hat{\sigma}$ is seldom above 0.2 or below 0.01; therefore the restriction that $0.01 \le \sigma_i \le 0.2$ is of small consequence.

To summarize, given a random vector of markups $\mathbf{m}_T$, we compute a vector of fitted costs $\bar{c}_{i,t}$ using equations (45) and (46), and we draw the values of $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)$, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ using equation (47). The likelihood for a vector of costs for the competitive model $M1$ conditional on θ is then

$$p_{c^{M,EST}}\left(\frac{c_{i,t}}{EST_t}\Big| z, \theta, M1\right) = \pi_1 N\left(\frac{c_{i,t} - \bar{c}_{i,t}(\theta, z_t)}{EST_t}, \mu_1, \sigma_1\right)$$
$$+ \pi_2 N\left(\frac{c_{i,t} - \bar{c}_{i,t}(\theta, z_t)}{EST_t}, \mu_2, \sigma_2\right)$$
$$+ \pi_3 N\left(\frac{c_{i,t} - \bar{c}_{i,t}(\theta, z_t)}{EST_t}, \mu_3, \sigma_3\right)$$
$$+ \pi_4 N\left(\frac{c_{i,t} - \bar{c}_{i,t}(\theta, z_t)}{EST_t}, \mu_4, \sigma_4\right). \tag{48}$$

The likelihood for the vector of costs from models $M2$ and $M3$ is computed similarly, except we assume that the cost for the cartel is the order statistic of its members' costs.

B. Putting It All Together

We now summarize all of the steps needed to compute a simulated value of the likelihood function for model M:

1. Draw a random vector $c_T^{M,\text{EST}}$ over the distribution of cost estimates from the distribution $p_{\hat{c}}(c_T^{M,\text{EST}})$ that was estimated in section VII.
2. Draw a random vector $\mathbf{m}_T$ from the distribution (43).
3. Given $\mathbf{m}_T$, compute fitted values $\bar{c}_{i,t}$ for all auctions t and all firms i.
4. Draw the parameters of the mixture of normal distribution π, μ, σ according to the distribution in equation (47).
5. Compute the value of the likelihood function, $p_{c^{M,\text{EST}}}$ $(c_{i,t}/\text{EST}_t | z, \theta, M1)$ as in equation (48).

Using standard simulation methods, it is possible to evaluate (38).

Using equation (39), we can compute the posterior beliefs for the three competing models. The posterior beliefs can be useful for decision-making. For instance, these beliefs will be used to determine whether to investigate and prosecute the suspected cartel. Suppose that the antitrust authority possesses a loss function $u(a, M)$, where $M \in \{M1, M2, M3\}$, and a is the action the authority must take, such as to prosecute or not. The authority's expected loss for an action a, conditional on the data, is

$$p^*(M1)u(a, M1) + p^*(M2)u(a, M2)$$
$$+ p^*(M3)u(a, M3). \tag{49}$$

Given the posterior probabilities over models $M1$, $M2$, and $M3$, it is then straightforward for the antitrust official to choose a loss-minimizing action.

A difference between using equation (49) and classical methods for model selection is that equation (39) explicitly takes into account the preferences of the decision-maker. For instance, an antitrust authority may only want to investigate when the posterior probability of collusion is quite large.

TABLE 12.—DISTRIBUTION OF MARKUPS UNDER ALTERNATIVE MODELS

Percentile	$M1$	$M2$	$M3$
10	0.01229	0.01273	0.0114
20	0.01597	0.01818	0.0182
30	0.02077	0.02422	0.0256
40	0.02536	0.03201	0.0343
50	0.03329	0.04126	0.0447
60	0.04227	0.05434	0.0584
70	0.05692	0.0754	0.0930
80	0.1000	0.1621	0.1756
90	0.2381	0.3354	0.5826

TABLE 13.—PERCENTAGE OF COST ESTIMATES FOUND TO BE NEGATIVE

Model	Percentage Negative
$M1$	3%
$M2$	5%
$M3$	7%

IX. Results

Using equation (36), we can compute an implied distribution of markups under the three alternative models, which we summarize in table 12. It is clear from table 12 that markups are consistently higher under the collusive models $M2$ and $M3$. The difference between these models becomes particularly striking near the 80th and 90th percentiles. In model $M2$, 20% of the projects have a markup in excess of 15%, and 10% have markups in excess of 33%. In model $M3$, 20% of the projects have a markup in excess of 17%, and 10% have markups in excess of 58%. Models $M2$ and $M3$ generate a much higher frequency of markups greater than 15% than the prior beliefs of our industry experts.

Another difference between models $M1$, $M2$, and $M3$ is the percentages of costs that are found to be negative, which we summarize in Table 13. This occurs when the right-hand side of (36) is negative at the estimated distribution of bids. In the seal coat industry, contractors frequently read the plan and specification strategically. They search for items where the estimated quantities will not agree with the actual quantities. When this occurs, the final compensation to the contractor will exceed the total bid, as described in section VI A. If contractors anticipate large adjustments to the final compensation, either from discrepancies in the estimate or actual quantities or from change orders, they will shade their bid downwards. This behavior could account for low bids that cannot be rationalized by positive costs. See Athey and Levin (2001) for a description of similar behavior in timber auctions.

We believe that it is inappropriate for us to censor the observations in which costs are negative. We interpret the fact that model $M3$ is unable to assign positive costs to 7% of the bids as evidence in favor of model $M1$ over $M3$. However, we do need to modify our likelihood function slightly to allow for negative costs. In computing the marginalized likelihoods, we modified the model of the previous section so that, with 95% probability, the likelihood function was constructed as in section VII. With 5% probability, however, we assume that $c_{i,t}/\text{EST}_t$ is uniformly distributed between 0.0 and -10. This simple modification of the likelihood function allows costs to be negative.

Using the procedure described in the previous section, we compute the marginalized likelihoods associated with each of the three competing models. These computations were done using approximately 5 hours of CPU time on a Sun Ultra 60 workstation. We summarize our results in table 14.[25] As we can see from table 14, the competitive model $M1$

---

[25] All confidence intervals are based on 500,000 simulated values of the likelihood function.

TABLE 14.—MARGINALIZED LIKELIHOODS AND MODEL POSTERIOR PROBABILITIES

Model	Expected Value of Log Likelihood	Log of Marginalized Likelihood	Posterior Probability
$M1$	62.169 (61.916, 62.423)	163.4 (161.89, 163.9)	1.00
$M2$	−320.41 (−320.58, 320.24)	−185.58 (−186.8, 184.75)	0.0
$M3$	−243.15 (−250.75, 235.56)	−186.6 (−187.81, 184.55)	0.0

strongly dominates the other two alternatives. This is because, as we saw in tables 12 and 13, $M1$ agrees more closely with the beliefs of our industry experts about markups in the industry. A large fraction of the markups generated in the collusive models are implausible a priori, given the beliefs of our industry experts. Also, the collusive models generate a much larger fraction of costs that are negative. Since the marginalized likelihood of the competitive model is much larger than those of the collusive models, we can be confident that our results will be robust to small perturbations of our prior distribution and to other aspects of our specification.

As an aside, we note that it is straightforward to simulate the posterior distribution of the structural parameters using Gibbs sampling. If we impose the prior beliefs elicited from our industry experts, our specification is just a linear model with a flexible error term. This is an alternative to the methods proposed by Guerre et al. (2000). In small samples, such as in our data set, there are often not enough data to implement nonparametric procedures. We would expect a Bayesian approach to have much better small-sample properties, because it utilizes information from industry experts and does not depend on asymptotic approximations.[26]

X. Conclusion

Most current research in industrial organization focuses on a single industry. Economists often speak informally to industry experts and read publications written by them in the course of doing research. Absorbing the views of industry experts is viewed by most empirical industrial organization economists as an essential step in conducting thorough research.

However, almost no formal use is made of expert opinion in estimation. A contribution of this paper is to make formal use of the prior beliefs of industry experts in order to decide between nonnested models of industry equilibrium. We believe that using the prior beliefs of our industry experts is a powerful tool that is potentially useful in many circumstances. In fact, our view is that if an economist's results strongly disagree with the prior beliefs of several industry

experts, it would be wise for the economist to double-check her work. In the seal coat industry, there are nearly 100 firms who bid at some point in our data set. However, only seven of the firms have a market share that exceeds 5%. In such a competitive environment, it is hard to understand how a firm can have wildly inaccurate views about markups and still manage to be one of the seven leading firms.

The views of industry experts are particularly useful when the economist only has a small sample. Economists are often forced to make decisions between alternative models or to make predictions with only a handful of data points. We believe that in many cases, using the a priori beliefs of experts will result in improved small-sample properties. It is possible to extend the approach we have developed to other problems in industrial organization. If it is possible to elicit a prior distribution over parameters or induce a prior distribution over parameters (in our case by using beliefs about markups), the tools of Bayesian statistics can be used to form a posterior distribution over parameters that take into account the beliefs of industry experts.

To summarize, in this research we have analyzed a model that allows for asymmetric bidders and nontrivial dynamics, which are important strategic considerations in many procurements. We stated a set of conditions that are both necessary and sufficient for a distribution of bids to arise from competitive bidding in the asymmetric auction model. Two of these conditions, conditional independence and exchangeability, can be tested in a straightforward fashion. We also argued that deciding between competitive and collusive bidding can be naturally formulated as a statistical decision problem. We developed techniques for structurally estimating asymmetric auction models and for computing the posterior probability of competitive and collusive models. Although no empirical techniques for detecting collusion are likely to be flawless, we believe that the tests we propose, taken together, can be a useful first step in detecting suspicious bidding patterns.

---

[26] Another way to increase the efficiency of our estimates would be to directly evaluate the likelihood function in a parametric auction model, as done by Paarsch (1992), who uses maximum likelihood, or Bajari and Hortacsu (2002), who use Bayesian methods. Bajari (2001) describes how to compute the equilibrium of the asymmetric auction model. Hirano and Porter (2002) demonstrate that Bayesian estimation of parametric auction models is asymptotically efficient, whereas alternative classical estimators are not. This method, however, has the disadvantage of being more computationally complex.

REFERENCES

Armantier, Olivier, Jean-Pierre Florens, and Jean-Francois Richard, "Empirical Game Theoretic Models: Constrained Equilibrium and Simulation," University of Pittsburgh working paper (1997).

Athey, Susan, and Jonathan Levin, "Information and Competition in U.S. Forest Service Timber Auctions," *Journal of Political Economy* 109:2 (2001), 375–417.

Bajari, Patrick, "The First-Price Auction with Asymmetric Bidders: Theory and Applications," University of Minnesota Ph.D. thesis (1997).

—— "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory* 18 (2001), 187–205.

Bajari, Patrick, and Ali Hortacsu, "The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions," Stanford University working paper (2002).

Baldwin, Laura, Robert Marshall, and Jean-Francois Richard, "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy* 105:4 (1997), 657–699.

Bartholowmew, S. H., *Construction Contracting: Business and Legal Principles* (Prentice-Hall, 1998).

Campo, Sandra, Isabelle Perrigne, and Quong Vuong, "Asymmetry in First-Price Auctions with Affiliated Private Values," *Journal of Applied Econometrics* 18:2 (2003), 179–207.

Clough, R., and G. Sears, *Construction Contracting* (New York: Wiley, 1994).

Cooke, R. M., *Experts in Uncertainty: Opinion and Subjective Probability in Science* (New York: Oxford University Press, 1991).

Flambard, V., and I. Perrigne, "Asymmetry in Procurement Auctions: Some Evidence from Snow Removal Contracts," University of Southern California working paper (2001).

Garthwait, P. H., and J. M. Dickey, "Quantifying Expert Opinion in Linear Regression Problems," *Journal of the Royal Statistical Society, B* 50 (1988), 462–474.

—— "Elicitation of Prior Distributions for Variable Selection Problems in Regression," *Annals of Statistics* 20 (1992), 1697–1719.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica* 68:3 (2000), 525–574.

Griesmer, James, Richard Levitan, and Martin Shubik, "Towards a Study of Bidding Processes, Part Four: Games with Unknown Costs," *Naval Research Quarterly* 14 (1967), 415–433.

Hinze, J., *Construction Contracts. McGraw-Hill Series in Construction Engineering and Product Management* (Irwin/McGraw-Hill, 1993).

Hirano, Keisuke, and Jack Porter, "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," Harvard University working paper (2002).

Hong, Han, and Matthew Shum, "Structural Estimation of First-Price Auction Models," Princeton University working paper (1999).

Kadane, J. B., J. M. Dickey, R. L. Winkler, W. S. Smith, and S. C. Peters, "Interactive Elicitation of Opinion for a Normal Linear Model," *Journal of the American Statistical Association* 75 (1980), 845–854.

Kadane, J. B., and L. J. Wolfson, "Experiences in Elicitation," *The Statistician* 47:Part 1 (1998), 3–19.

Lebrun, Bernard, "Existence of an Equilibrium in First Price Auctions," *Economic Theory* 7:3 (1996), 421–443.

—— "First-Price Auction in the Asymmetric N Bidder Case," *International Economic Review* 40:1 (1999), 125–142.

—— "Uniqueness of Equilibrium in the N Bidder Asymmetric Auction," Université Laval mimeograph (2000).

Maas, P., *Underboss: Sammy the Bull Gravano's Story of Life in the Mafia* (Mass Market Paperback, 1997).

Maskin, Eric, and John Riley, "Asymmetric Auctions," *Review of Economic Studies* 67:3 (2000a), 413–438.

—— "Equilibrium in Sealed High Bid Auctions," *Review of Economic Studies* 67:3 (2000b), 493–454.

Minnesota Department of Transportation, *Standard Specifications for Construction* (1995) (Minneapolis: State Department of Transportation, 1995).

O'Hagan, A., "Eliciting Expert Beliefs in Substantial Practical Applications," *The Statistician* 47:Part 1 (1998), 21–35.

Paarsch, Harry, "Deciding between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics* 51:1–2 (1992), 191–215.

Pesendorfer, Martin, "A Study of Collusion in First-Price Auctions," *Review of Economic Studies* 67:3 (2000), 381–411.

Pesendorfer, Martin, and Mireia Jofre-Bonet, "Bidding Behavior in a Repeated Procurement Auction," Yale University mimeo (2000).

Porter, Robert and Douglas Zona, "Detection of Bid-Rigging in Procurement Auctions," *Journal of Political Economy* 101 (1993), 518–538.

—— "Ohio School Milk Markets: An Analysis of Bidding," *Rand Journal of Economics* 30:2 (1999), 263–288.

APPENDIX

PROOF OF THEOREM 5. To construct the distribution of costs that rationalizes the distribution of bids, note that by A4 the function $\xi_i(b, z)$ is strictly increasing, and thus we can define the distribution of costs $F(c|z_i)$ as follows:

$$F(c|z_i) = \Pr(\xi_i(b, z) \leq c) = G_i(\xi_i^{-1}(c, z); z). \tag{50}$$

By A2, A5, and (15), the cumulative distribution functions $F(c|z_i)$ all have the same support. By the existence theorem and uniqueness theorem of the previous section, a unique equilibrium exists when firms $i = 1, \ldots, N$ have costs independently distributed according to the construction in equation (50). Let $\phi_i(b, z)$ denote the equilibrium bidding strategies. By our uniqueness theorem there is one and only one set of inverse bid functions that satisfies the differential equations

$$\phi_i(b, z) = b - \frac{1}{\displaystyle\sum_{j \neq i} \frac{f(\phi_j(b; z)|z_j)\phi_j'(b; z)}{1 - F(\phi_j(b; z)|z_j)}} \tag{51}$$

and the boundary conditions $\phi_i(\bar{b}, z) = \bar{c}$, $\phi_i(\underline{b}, z) = \underline{c}$. By our construction (50) it follows that

$$\begin{aligned} F(\xi_i(b, z)|z_i) &= G_i(b; z), \\ f(\xi_i(b, z)|z_i) &= g_i(b; z)\xi_i'(b, z). \end{aligned} \tag{52}$$

Using (52), it is easily verified that $\phi_i(b, z) = \xi_i(b, z)$ solves (51). By A5, $\xi_i(b, z)$ also satisfies the boundary conditions. This shows that $\xi_i(\cdot, z)$ is the unique equilibrium strategy when the distributions of the bidders' private costs are given by $F(\cdot|z_i)$ as constructed in (50). In other words, the constructed cost distributions (50) give rise to the given distributions of bids in our model with asymmetric bidders.

Finally, such a latent cost distribution must be unique, because it must satisfy (50) where $\xi_i(\cdot, z)$ is the unique equilibrium strategy. Q.E.D.