Online Appendix

A.1 Proofs and theoretical discussions

Simulation of policy implications: synergy versus affiliation Table 2 shows that when $N = 2$, the observed probability that the same bidder wins both tracts in a pair is 74%. I use simulations with a simple model to explore some implications of failing to distinguish synergy and affiliation across auctions when analyzing this data. Consider a sequence of two second-price auctions, where bidders' marginal value distributions in both auctions are $U[0,1]$. Suppose I assume winning both tracts results in total value $v_1 + v_2 + \alpha$, where α is a constant, but I disallow affiliation by assuming independence of values across auctions. To rationalize a 74% probability of the same bidder winning both tracts, α must be roughly 0.3, indicating a substantial amount of synergy. Simulations show that this model would recommend sequential auctions over a bundled second-price auction for 5% higher revenue.

Meanwhile, the observed statistic could also be rationalized by a model with affiliation but no synergy. For instance, the marginal distributions of v_1 and v_2 could be related by a Gaussian copula with dependence parameter of roughly 0.7. If this was the true model, the former model would have grossly mismeasured synergy as an object of interest. Moreover, the formerly recommended policy of sequential auctions would actually result in a revenue loss relative to bundled auctions, of roughly 5%. The two explanations lead to opposing policy recommendations regarding revenue. In fact, they also lead to opposing recommendations regarding which option would improve efficiency, or generate the greatest value for auction winners.

Deriving the first-order condition in section 3.2 A bidder will bid the b that maximizes the expected profit $\pi(v_1, b)$. Taking the derivative of $\pi(v_1, b)$ with respect to b and setting it equal to zero gives

$$\begin{aligned} 0 &= -G^{N-1}(b) + (N-1)G^{N-2}(b)g(b) \int_{v_2=\underline{v}}^{\bar{v}} \Bigg\{ v_1 - b \\ &\quad + \int_{u=\underline{v}}^{s(v_1,v_2)} (s(v_1,v_2) - u) dH_1(u|b) - \int_{u=\underline{v}}^{v_2} (v_2 - u) dH_2(u|b) \Bigg\} dF_2(v_2|v_1). \end{aligned} \tag{10}$$

Rearranging this gives

$$\frac{G(b)}{(N-1)g(b)} = \int_{v_2=\underline{v}}^{\bar{v}}\{v_1 - b + \int_{u=\underline{v}}^{s(v_1,v_2)}(s(v_1,v_2)-u)dH_1(u|b) - \int_{u=\underline{v}}^{v_2}(v_2-u)dH_2(u|b)\}dF_2(v_2|v_1).$$

Some algebra using integration by parts shows that

$$\int_{u=\underline{v}}^{s(v_1,v_2)}(s(v_1,v_2)-u)dH_1(u|b) = \int_{u=\underline{v}}^{s(v_1,v_2)} H_1(u|b)du$$

$$\int_{u=\underline{v}}^{v_2}(v_2-u)dH_2(u|b) = \int_{u=\underline{v}}^{v_2} H_2(u|b)du.$$

So the first-order condition can be simplified to

$$b = v_1 + \int_{v_2=\underline{v}}^{\bar{v}}\{\int_{u=\underline{v}}^{s(v_1,v_2)} H_1(u|b)du - \int_{u=\underline{v}}^{v_2} H_2(u|b)du\}dF_2(v_2|v_1) - \frac{G(b)}{(N-1)g(b)}.$$

Proof of Proposition 1 First, I show that if $v_1' > v_1$, $b \in BR(v_1)$ (best response set), and $b' \in BR(v_1')$, then it must be that $b' \geq b$. Suppose not; suppose $b' < b$. By definition of best response, $b \in BR(v_1)$ means $\pi(v_1,b) - \pi(v_1,b') \geq 0$ and $b' \in BR(v_1')$ means $0 \geq \pi(v_1',b) - \pi(v_1',b')$. Defining $\kappa(v_1) \equiv \pi(v_1,b) - \pi(v_1,b')$, this means $\kappa(v_1) \geq \kappa(v_1')$. Writing out $\kappa(v_1)$ gives the following expression:

$$\int_{v_2=\underline{v}}^{\bar{v}}\{ v_1[G^{N-1}(b) - G^{N-1}(b')] - bG^{N-1}(b) + b'G^{N-1}(b') + \int_{t=b'}^{b}\int_{u=\underline{v}}^{s(v_1,v_2)}(s(v_1,v_2)-u)dH_1(u|t)dG^{N-1}(t) - \int_{t=b'}^{b}\int_{u=\underline{v}}^{v_2}(v_2-u)dH_2(u|t)dG^{N-1}(t)\}dF_2(v_2|v_1).$$

Then, after some algebra and integration by parts, $\kappa(v_1) - \kappa(v_1')$ is

$$(v_1 - v_1')[G^{N-1}(b) - G^{N-1}(b')] + \int_{t=b'}^{b}[\lambda(v_1,t) - \lambda(v_1',t)]\,dG^{N-1}(t) - \int_{t=b'}^{b}[\mu(v_1,t) - \mu(v_1',t)]\,dG^{N-1}(t),$$

where $\lambda(v_1,t) \equiv \int\int_{u=\underline{v}}^{s(v_1,v_2)} H_1(u|t)du\,dF_2(v_2|v_1)$ and $\mu(v_1,t) \equiv \int\int_{u=\underline{v}}^{v_2} H_2(u|t)du\,dF_2(v_2|v_1)$. First, since $v_1' > v_1$, $b' < b$, and $G^{N-1}(\cdot)$ is a cdf, the first part of the expression above is negative. Second, since H_1 and H_2 are non-negative, $F_2(v_2|v_1)$ is stochastically ordered in v_1, and $s(v_1,v_2)$ is weakly increasing in it arguments, both $\lambda(v_1,t)$ and $\mu(v_1,t)$ are weakly increasing in v_1. Hence the second part of the expression is negative and the third part is positive. Now I focus on this positive third part. $H_2(u|t) < 1$ for all $u < \bar{v}$ because it is a cdf. As a result, $-[\mu(v_1,t) - \mu(v_1',t)]$ is strictly bounded above by $\int\int_{u=\underline{v}}^{v_2} 1du\,dF_2(v_2|v_1') -$

$\int \int_{u=\underline{v}}^{v_2} 1du\, dF_2(v_2|v_1) = \int v_2 dF_2(v_2|v_1') - \int v_2 dF_2(v_2|v_1) = E[v_2|v_1'] - E[v_2|v_1]$. Hence,

$$\begin{aligned}\kappa(v_1) - \kappa(v_1') &< (v_1 - v_1')[G^{N-1}(b) - G^{N-1}(b')] + \int_{t=b'}^{b} \{E[v_2|v_1'] - E[v_2|v_1]\} dG^{N-1}(t) \\ &= [G^{N-1}(b) - G^{N-1}(b')][(v_1 - v_1') + E[v_2|v_1'] - E[v_2|v_1]] \\ &\leq 0.\end{aligned}$$

The last line comes from $E[v_2|v_1'] - E[v_2|v_1] \leq v_1' - v_1$, which is given by AS4 and AS5. Then according to the inequality above, $\kappa(v_1) - \kappa(v_1') < 0$. However, this contradicts $\kappa(v_1) \geq \kappa(v_1')$, which must be satisfied by definition of best response. Hence by contradiction, it must be that $b' \geq b$. Note that AS5 is stronger than necessary to arrive at this result; there is slack in the inequality above. In terms of the existing literature, the condition I have shown above implies that IRT-SCC as defined in Reny and Zamir (2004) is satisfied. Then, by their Theorem 2.1, the first auction possesses a monotone pure-strategy equilibrium. I continue below to show that all symmetric equilibria for this auction must be monotone.

Next, I show that two different values cannot share the same best response b. Consider $\pi_b(v_1, b)$, the derivative of the expected profit function with respect to b. For any $v_1' > v_1$, $\pi_b(v_1', b) - \pi_b(v_1, b)$ is

$$\begin{aligned}&\pi_b(v_1', b) - \pi_b(v_1, b) \\ &\quad = \{v_1' - v_1 + \lambda(v_1', b) - \lambda(v_1, b) - [\mu(v_1', b) - \mu(v_1, b)]\}(N-1)G^{N-2}(b)g(b) \\ &\quad > \{v_1' - v_1 + \lambda(v_1', b) - \lambda(v_1, b) - [E[v_2|v_1'] - E[v_2|v_1]]\}(N-1)G^{N-2}(b)g(b) \\ &\quad \geq 0.\end{aligned}$$

Again, the second line comes from the fact that $\mu(v_1', b) - \mu(v_1, b)$ is strictly bounded above by $E[v_2|v_1'] - E[v_2|v_1]$, and the third line comes from $E[v_2|v_1'] - E[v_2|v_1] \leq v_1' - v_1$, $\lambda(v_1', b) - \lambda(v_1, b) \geq 0$, and $(N-1)G^{N-2}(b)g(b) > 0$. So $\pi_b(v_1, b)$ is strictly increasing in v_1. Hence, for any given bid b and bid distribution $G(\cdot)$, there can only be one v_1 that satisfies equation (3); two different values cannot share the same best response b. This rules out $b' = b$. We already established that $b' \geq b$, so it must be that $b' > b$.

Finally, I show that for each v_1, there cannot be more than one $b \in BR(v_1)$. Suppose not; suppose $b'' \in BR(v_1)$ as well, and without loss of generality, $b'' > b$. Given what we already established, the probability of winning does not change whether the bidder bids b'' or b; bidders with values lower than v_1 will bid lower than all elements of $BR(v_1)$ and bidders with higher values will bid higher than all elements of $BR(v_1)$. On the other hand, bidding more still decreases the bidder's payoff upon winning. As a result, $\pi(v_1, b'') < \pi(v_1, b)$. However, this contradicts the premise that $b'' \in BR(v_1)$. Therefore, there cannot be more than one $b \in BR(v_1)$. Hence we can define a bid function $b(v_1)$, which is strictly increasing.

Checking the second-order condition In the proof of Proposition 1, I established that the derivative $\pi_b(v_1, b)$ is strictly increasing in v_1. Let $\xi(\cdot)$ be the inverse bid function. By monotonic bidding, if $x < b(v_1)$, then $\xi(x) < v_1$. Then, since $\pi_b(v_1, b)$ is strictly increasing in v_1, $\pi_b(v_1, x) > \pi_b(\xi(x), x) = 0$. Likewise, if $x > b(v_1)$, then $\xi(x) > v_1$ by monotonic bidding, so $\pi_b(v_1, x) < \pi_b(\xi(x), x) = 0$. In summary, $\pi_b(v_1, x) > 0$ if $x < b(v_1)$ and $\pi_b(v_1, x) < 0$ if $x > b(v_1)$, so $b(v_1)$ does achieve the global maximum of $\pi(v_1, \cdot)$.

Proof of Proposition 2 First I derive (5). Replacing $H_1(u|b)$, $H_2(u|b)$, $G(b)$, and $g(b)$ with $\tilde{H}_1(u|v_1)$, $\tilde{H}_2(u|v_1)$, $F_1(v_1)$, and $f_1(v_1)/b'(v_1)$, respectively in (10) gives $0 = -F_1(v_1)^{N-1} + (N-1)F_1(v_1)^{N-2}\frac{f_1(v_1)}{b'(v_1)}(T(v_1) - b)$. Rearranging terms leads to $b'(v_1)F_1(v_1)^{N-1} + b(N-1)F_1(v_1)^{N-2}f_1(v_1) = T(v_1)(N-1)F_1(v_1)^{N-2}f_1(v_1)$, which is exactly (5). With the boundary condition $P(\underline{v}) \equiv b(\underline{v})F_1(\underline{v})^{N-1} = 0$, the solution to that differential equation is $P(v_1) = \int_{\underline{v}}^{v_1} T(x)dF_1(x)^{N-1}$. Now, since $P(v_1) \equiv b(v_1)F_1(v_1)^{N-1}$, this means $b(v_1)F_1(v_1)^{N-1} = \int_{\underline{v}}^{v_1} T(x)dF_1(x)^{N-1}$. Hence $b(v_1) = \int_{\underline{v}}^{v_1} T(x)dF_1(x)^{N-1}/F_1(v_1)^{N-1}$. This is a bid function that must be satisfied in any symmetric equilibrium; therefore the equilibrium given by this bid function is the only symmetric Bayes-Nash equilibrium.

Proof of Proposition 3 Step (i): For a fixed set of first auction bids $\{b_i\}$, values in the second auction are drawn from $\tilde{D}(\cdot|b_{w1})$ for the A1-winner $w1$, and from $\tilde{F}_2(\cdot|b_i)$ each loser $i \neq w1$. These draws are independent across bidders. Furthermore, by assumption AS3, all value distributions involved are continuous and have the same support. Hence, we can apply Theorem 2 of Athey and Haile (2002), which establishes identification of asymmetric value distributions from transaction prices and bidder identities. Theorem 3 of Athey and Haile (2002) extends this to auctions with auction-specific covariates.

Step (ii): By assumption AS6, $\tilde{s}(b, v_2)$ is weakly increasing in v_2. So if we define $v_2(\alpha|b) \equiv \tilde{F}_2^{-1}(\alpha|b)$, i.e. the α-quantile of v_2 conditional on b, then $\tilde{s}(b, v_2(\alpha|b))$ must be the α-quantile of s conditional on b, $\tilde{D}^{-1}(\alpha|b)$. That is, for any quantile α, $\tilde{s}(b, \tilde{F}^{2,-1}(\alpha|b)) = \tilde{D}^{-1}(\alpha|b)$. Since b is observed and $\tilde{F}_2(\cdot|b)$ and $\tilde{D}(\cdot|b)$ are identified from step (i), we know the function $\tilde{s}(\cdot, \cdot)$.

Step (iii): Consider (4), the inverse bid function. From steps (i) and (ii), every component of the right-hand side is either observed or identified from data, so $\xi(b)$ can be computed. Since bids are monotonic in v_1, the α-quantile of v_1, $v_1(\alpha)$, corresponds to $\xi(b(\alpha))$. Now, since the distribution of b is observed and $\xi(b)$ can be computed for any b, we can compute $v_1(\alpha)$ for any quantile α. Hence, the distribution of v_1 is identified nonparametrically.

Step (iv) is explained fully in the text.

Role of AS3 in identification Revisiting step (i), one assumption that goes into Athey and Haile (2002)'s result is that all value distributions have the same support. I list this assumption in AS3 to provide for full identification. What are the consequences if this assumption does not hold? Suppose $D(\cdot|\cdot)$ has a larger support than $F_2(\cdot|\cdot)$ with a greater supremum $\bar{v}_D > \bar{v}_{F_2}$. Since English auction prices only reveal the second highest value, we never observe prices above $\bar{v}_{F_2}$, meaning we gather no information on the shape of $D(\cdot|\cdot)$ in the interval $(\bar{v}_{F_2}, \bar{v}_D)$. So $D(\cdot|\cdot)$ will be identified on $[\underline{v}, \bar{v}_{F_2}]$, where the two supports overlap, but not on $(\bar{v}_{F_2}, \bar{v}_D)$. Meilijson (1981), on which Theorem 2(a) of Athey and Haile (2002) is based, discusses identification on non-identical supports in section 6.

Proof of Remark 1 Consider the $N = 2$ case as an example. The distribution of the second highest value out of $\{s(v_1, v_2),\ v_2\}$, $J(\cdot|z'\beta)$, can be rewritten $\{F_2^{-1}(\alpha_s|z'\beta), F_2^{-1}(\alpha_2|z'\beta)\}$, where the -1 superscript indicates the inverse function. Now for any v, define $\alpha \equiv F_2(v|z'\beta)$ and $\tilde{\alpha} \equiv J(v|z'\beta)$. Then $\tilde{\alpha} \equiv J(v|z'\beta) = J(F_2^{-1}(\alpha|z'\beta)|z'\beta)=$ $\Pr(\{F_2^{-1}(\alpha_s|z'\beta), F_2^{-1}(\alpha_2|z'\beta)\}_{(2)} \leq F_2^{-1}(\alpha|z'\beta)|z'\beta) = \Pr(\{\alpha_s, \alpha_2\}_{(2)} \leq \alpha|z'\beta)$.[28] From AS8, the distributions of α_s, α_2 are invariant to $z'\beta$, so we can simplify $\tilde{\alpha} = \Pr(\{\alpha_s, \alpha_2\}_{(2)} \leq \alpha|z'\beta)$ to $\Pr(\{\alpha_s, \alpha_2\}_{(2)} \leq \alpha)$. Hence $\tilde{\alpha}$ is a function only of α, invariant to z. Furthermore, since $C(\alpha_1, \alpha_2)$ is invariant to z according to AS8, and $\tilde{\alpha}$ is a function only of α, $C(\alpha_1, \tilde{\alpha}_2)$ is also invariant to z. The same applies for $C(\alpha_1, \tilde{\alpha}_s)$.

Bid homogenization Haile et al. (2003)'s method can be used for bid homogenization if their assumptions of additive separability and independence are extended to a joint distribution of values and a synergy function. A proof follows.

The assumptions needed are (1) Additively separable structure on values: $v_1 = \Gamma(z)+\epsilon_1$; $v_2 = \Gamma(z) + \epsilon_2$; $s(v_1, v_2) = s(\Gamma(z) + \epsilon_1, \Gamma(z) + \epsilon_2) = \Gamma(z) + \check{s}(\epsilon_1, \epsilon_2)$; and (2) Independence of joint distribution of ϵ from z: $\check{F}(\epsilon_1, \epsilon_2; z) = \check{F}(\epsilon_1, \epsilon_2)$.

Note that as a direct consequence of these assumptions, $\check{F}_1(\epsilon_1)$, $\check{F}_2(\epsilon_2|\epsilon_1)$, and $\check{D}(s(\epsilon_1, \epsilon_2)|\epsilon_1)$ are all independent of z. For simplicity assume there exists a z_0 such that $\Gamma(z_0) = 0$. Given these assumptions, I show that the expected benefit in A2 of winning A1 - the bracketed part of $T(v_1)$ - is a function of ϵ_1. From this, the additive separability of the bid function follows easily. First, I perform some necessary change-of-variables algebra. Notation-wise, I use η_1 to stand in for a generic draw from $\check{F}_1(\cdot)$.

$$
\begin{aligned}
\tilde{H}_1(u|v_1; z) &= \tilde{H}_1(\Gamma(z) + \epsilon_u|\Gamma(z) + \epsilon_1; z) \\
&\equiv F_2(\Gamma(z) + \epsilon_u|\eta_1 \leq \epsilon_1; z)^{N-2} F_2(\Gamma(z) + \epsilon_u|\eta_1 = \epsilon_1; z) \\
&= \check{F}_2(\epsilon_u|\eta_1 \leq \epsilon_1)^{N-2} \check{F}_2(\epsilon_u|\eta_1 = \epsilon_1) \\
&\equiv \check{H}_1(\epsilon_u|\epsilon_1).
\end{aligned}
$$

[28] The $\{\}_{(2)}$ subscript indicates the second order statistic out of the values in $\{\}$.

$$\begin{aligned}\tilde{H}_2(u|v_1;z) &= \tilde{H}_2(\Gamma(z)+\epsilon_u|\Gamma(z)+\epsilon_1;z)\\ &\equiv F_2(\Gamma(z)+\epsilon_u|\eta_1 \le \epsilon_1;z)^{N-2}D(\Gamma(z)+\epsilon_u|\eta_1=\epsilon_1;z)\\ &= \check{F}_2(\epsilon_u|\eta_1 \le \epsilon_1)^{N-2}\check{D}(\epsilon_u|\eta_1=\epsilon_1)\\ &\equiv \check{H}_2(\epsilon_u|\epsilon_1).\end{aligned}$$

For both $\tilde{H}_1(u|v_1;z)$ and $\tilde{H}_2(u|v_1;z)$ above, the first line comes from applying additive separability, the second line comes from the definition of $\tilde{H}(\cdot|\cdot)$, and the third line comes from applying independence from z. Then as a result we have $\int_{\underline{v}}^{s(v_1,v_2)} \tilde{H}_1(u|v_1;z)du = \int_{\Gamma(z)+\underline{\epsilon_2}}^{\Gamma(z)+\check{s}(\epsilon_1,\epsilon_2)} \tilde{H}_1(u|v_1;z)du = \int_{\underline{\epsilon_2}}^{\check{s}(\epsilon_1,\epsilon_2)} \check{H}_1(\epsilon_u|\epsilon_1)d\epsilon_u$. Similarly, $\int_{\underline{v}}^{v_2} \tilde{H}_2(u|v_1;z)du = \int_{\Gamma(z)+\underline{\epsilon_2}}^{\Gamma(z)+\epsilon_2} \tilde{H}_2(u|v_1;z)du = \int_{\underline{\epsilon_2}}^{\epsilon_2} \check{H}_2(\epsilon_u|\epsilon_1)d\epsilon_u$.

Then,

$$\begin{aligned}&\int_{v_2=\underline{v}}^{\bar{v}} \left\{ \int_{u=\underline{v}}^{s(v_1,v_2)} \tilde{H}_1(u|v_1;z)du - \int_{u=\underline{v}}^{v_2} \tilde{H}_2(u|v_1;z)du \right\} dF_2(v_2|v_1;z)\\ &= \int_{\epsilon_2=\underline{\epsilon_2}}^{\bar{\epsilon}_2} \left\{ \int_{\underline{\epsilon_2}}^{\check{s}(\epsilon_1,\epsilon_2)} \check{H}_1(\epsilon_u|\epsilon_1)d\epsilon_u - \int_{\underline{\epsilon_2}}^{\epsilon_2} \check{H}_2(\epsilon_u|\epsilon_1)d\epsilon_u \right\} d\check{F}_2(\epsilon_2|\epsilon_1).\\ &\equiv EB(\epsilon_1).\end{aligned}$$

The second line above is a function of ϵ_1, just as the first line is a function of v_1. Let's call this function $EB(\cdot)$.

Now, $T(v_1)$ as defined in section 3.3 can be rewritten $T(v_1) = v_1 + EB(\epsilon_1) = \Gamma(z) + \epsilon_1 + EB(\epsilon_1)$. Define $\epsilon_1(x;z) \equiv x - \Gamma(z)$. Then, by Proposition 2, the bid $b(v_1;z) = \int_{\underline{v}}^{v_1} T(x)dF_1(x)^{N-1}/F_1(v_1)^{N-1} = \int_{\underline{v}}^{v_1} \{\Gamma(z)+\epsilon_1(x;z)+EB(\epsilon_1(x;z))\}dF_1(x)^{N-1}/F_1(v_1)^{N-1} = \Gamma(z)\int_{\underline{v}}^{v_1} dF_1(x)^{N-1}/F_1(v_1)^{N-1} + \int_{\underline{v}}^{v_1} \{\epsilon_1(x;z) + EB(\epsilon_1(x;z))\}dF_1(x;z)^{N-1}/F_1(v_1;z)^{N-1} = \Gamma(z) + \int_{y=\underline{\epsilon_1}}^{\epsilon_1} \{y + EB(y)\}d\check{F}_1(y)^{N-1}/\check{F}_1(\epsilon_1)^{N-1} = \Gamma(z) + b(\epsilon_1;z_0)$.

So the bid function has the additively separable form $b(v_1;z) = \Gamma(z) + b(\epsilon_1;z_0)$ as in Haile et al. (2003).

A.2 Model extensions

A.2.1 Asymmetric bidders

I extend the model of sequential auctions to bidders having asymmetric value distributions and synergy functions. For simplicity, I illustrate the case of two asymmetric subgroups with at least two bidders per group, but this is not essential. The asymmetric model requires additional notation. First, a superscript or subscript m will indicate the subgroup to which value distributions and synergy functions belong, so $v_1 \sim F_1^m(\cdot)$, $v_2 \sim F_2^m(\cdot|v_1)$, $s_m(v_1,v_2)$, and $D_m(x|v_1) \equiv prob(s_m(v_1,v_2) \le x|v_1)$. I assume the support of these distributions is the same across groups. Then the distribution of the highest competing bid in A2 conditional on the highest competing bid in A1 being t can be expressed as one of the following, depending on whether the bidder wins A1 and depending on the subgroup of his highest competitor. Namely,

$$
\begin{aligned}
H_1^{m,m}(u|t) &= F_2^m(u|b_m \le t)^{N_m-2} F_2^{-m}(u|b_{-m} \le t)^{N-m} F_2^m(u|b_m = t) \\
H_1^{m,-m}(u|t) &= F_2^m(u|b_m \le t)^{N_m-1} F_2^{-m}(u|b_{-m} \le t)^{N-m-1} F_2^{-m}(u|b_{-m} = t) \\
H_2^{m,m}(u|t) &= F_2^m(u|b_m \le t)^{N_m-2} F_2^{-m}(u|b_{-m} \le t)^{N-m} D^m(u|b_m = t) \\
H_2^{m,-m}(u|t) &= F_2^m(u|b_m \le t)^{N_m-1} F_2^{-m}(u|b_{-m} \le t)^{N-m-1} D^{-m}(u|b_{-m} = t).
\end{aligned}
$$

Subscript 1 on $H(\cdot|\cdot)$ applies if the bidder wins A1, and subscript 2 applies if the bidder loses A1. The first superscript on $H(\cdot|\cdot)$ indicates the subgroup of the bidder being analyzed, and the second superscript indicates the subgroup of his highest competing bidder in A1.

Now, for a bidder from subgroup m, the probability that the highest competing bid in A1 is less than or equal to t is $G_m(t)^{N_m-1}G_{-m}(t)^{N-m}$, where $G_m(\cdot)$ is the distribution of A1 bids from subgroup m. Then, for a bidder from subgroup m, the probability (density) that the highest competing bid in A1 is equal to t is $\partial[G_m(t)^{N_m-1}G_{-m}(t)^{N-m}]/\partial t$ and can be expressed as $j_m(t)+k_m(t)$, where $j_m(t) \equiv (N_m-1)G_m(t)^{N_m-2}g_m(t)G_{-m}(t)^{N-m}$ is the probability that the highest competing bid in A1 is equal to t and from subgroup m, and $k_m(t) \equiv N_{-m}G_{-m}(t)^{N-m-1}g_{-m}(t)G_m(t)^{N_m-1}$ is the probability that the highest competing bid in A1 is equal to t and from subgroup $-m$. Using this notation, the expected profit at the time of A1 for a bidder from subgroup m is

$$
\pi_m(v_1,b) = \int_{v_2=\underline{v}}^{\bar{v}} X_m(v_1,v_2,b)\,dF_2^m(v_2|v_1),
$$

where

$$
\begin{aligned}
X_m(v_1,v_2,b) \equiv\ & \textstyle\int_{t=\underline{b}}^{b}[v_1 - b + \int_{u=\underline{v}}^{s_m(v_1,v_2)}(s_m(v_1,v_2)-u)\,dH_1^{m,m}(u|t)]j_m(t)dt \\
&+ \textstyle\int_{t=\underline{b}}^{b}[v_1 - b + \int_{u=\underline{v}}^{s_m(v_1,v_2)}(s_m(v_1,v_2)-u)\,dH_1^{m,-m}(u|t)]k_m(t)dt \\
&+ \textstyle\int_{t=b}^{\bar{b}}\int_{u=\underline{v}}^{v_2}(v_2-u)\,dH_2^{m,m}(u|t)j_m(t)dt \\
&+ \textstyle\int_{t=b}^{\bar{b}}\int_{u=\underline{v}}^{v_2}(v_2-u)\,dH_2^{m,-m}(u|t)k_m(t)dt.
\end{aligned}
$$

In the equation defining $X_m(\cdot,\cdot,\cdot)$, the first two parts account for the probability that the bidder wins A1 and the last two parts account for the probability that he loses A1. There are two parts to each case because with asymmetry, the identity (subgroup) of the highest competing bidder in A1 matters for the bidder's expected profit in A2.

Taking a derivative of the expected profit function $\pi_m(v_1,b)$ with respect to b yields the first-order condition for bidding for subgroup m. After simplifying and rearranging, the first-order condition for subgroup m can be rewritten as

$$
\begin{aligned}
&G_m(b)^{N_m-1}G_{-m}(b)^{N_{-m}} \\
&= (v_1 - b)(j_m(b) + k_m(b)) \\
&+ \textstyle\int_{v_2=\underline{v}}^{\bar{v}} \{ j_m(b)[\int_{u=\underline{v}}^{s_m(v_1,v_2)} H_1^{m,m}(u|b)du - \int_{u=\underline{v}}^{v_2} H_2^{m,m}(u|b)du] \\
&+ \qquad k_m(b)[\textstyle\int_{u=\underline{v}}^{s_m(v_1,v_2)} H_1^{m,-m}(u|b)du - \int_{u=\underline{v}}^{v_2} H_2^{m,-m}(u|b)du]\} dF_2^m(v_2|v_1).
\end{aligned}
\tag{11}
$$

This is structurally similar to the FOC for symmetric bidders in (3), but breaks down terms to account for differences between subgroups. The logic of Proposition 1 still applies in the asymmetric case, so bidding in A1 is monotonic in v_1 within each subgroup and the single crossing condition[29] is satisfied. By Reny and Zamir (2004), there exists a monotone pure-strategy equilibrium.

Provided that a single equilibrium is being played, the primitives of the asymmetric model, $F_1^m(\cdot)$, $F_2^m(\cdot|\cdot)$, and $s_m(\cdot,\cdot)$, are identified from the observables, which are all the bids in A1 and the transaction price in A2, along with bidder identities.

To see this, split the data conceptually into two subsamples, one where the first-auction winner is from subgroup m, and the other where the first-auction winner is from subgroup $-m$. Take the first subsample. In the first subsample, bidders in A2 are either the A1-winner from subgroup m, an A1-loser from subgroup m, or an A1- loser from subgroup $-m$. Following Proposition 3(i), the value distributions from which each of these bidders draws their A2 values conditional on their observed A1 bids - $\tilde{D}_m(\cdot|b)$ and $\tilde{F}_2^{-m}(\cdot|b)$ at least[30] - are identified. Similarly, $\tilde{D}_{-m}(\cdot|b)$ and $\tilde{F}_2^m(\cdot|b)$ are additionally identified from the second subsample. Then, following Proposition 3(ii), the synergy function $\tilde{s}_m(b,\cdot)$ is identified from the difference between $\tilde{D}_m(\cdot|b)$ and $\tilde{F}_2^m(\cdot|b)$, and $\tilde{s}_{-m}(b,\cdot)$ is identified from $\tilde{D}_{-m}(\cdot|b)$ and $\tilde{F}_2^{-m}(\cdot|b)$. This compares an A1-winner and A1-loser conditional on the same subgroup and same first-auction bid. Finally, $F_1^m(v_1)$ and $F_1^{-m}(v_1)$ are identified using each subgroup's first-order condition (11) for bidding in A1; that is, if we replace $s_m(v_1,v_2)$ with $\tilde{s}_m(b,v_2)$ and $F_2^m(v_2|v_1)$ with $\tilde{F}_2^m(v_2|b)$, every component of (11) other than v_1 is either observed or identified. Therefore, we can recover any quantile of v_1 by computing this equation at the same quantile of b_m. Once $F_1^m(v_1)$ and $F_1^{-m}(v_1)$ are identified, we can convert $\tilde{s}_m(b,v_2)$ and $\tilde{F}_2^m(v_2|b)$ back to $s_m(v_1,v_2)$ and $F_2^m(v_2|v_1)$ by replacing the α-quantile of b_m with the α-quantile of v_1. This completes identification with asymmetric bidders.

[29] Specifically, the IRT-SCC as defined by Reny and Zamir (2004).

[30] I say "at least" because this is what is identified in the minimal case of one bidder per subgroup. In that scenario, the first subsample contains exactly one subgroup m winner and one subgroup $-m$ loser, and vice versa for the second subsample.

A.2.2 Sequence of two second-price auctions

Consider a sequence of two second-price auctions. Bidding in the second auction does not change from section 3.2. The first auction does change. Retaining the same notation as before, the expected profit from the two auctions at the time of the first auction, where the bidder bids b is

$$\begin{aligned}\pi(v_1, b) &= \int_{v_2=\underline{v}}^{\bar{v}} \left\{ \int_{t=\underline{b}}^{b} \left(v_1 - t + \int_{u=\underline{v}}^{s(v_1,v_2)} (s(v_1, v_2) - u) dH_1(u|t) \right) dG^{N-1}(t) \right. \\ &\left. + \int_{t=b}^{\bar{b}} \int_{u=\underline{v}}^{v_2} (v_2 - u) dH_2(u|t) dG^{N-1}(t) \right\} dF_2(v_2|v_1).\end{aligned}$$

Setting $\partial\pi(v_1, b)/\partial b = 0$ yields the following first-order condition for bidding in the first auction. It says bidders in the first auction bid v_1 plus the expected benefit in the second auction of winning the first auction (contrast to (3)). Namely,

$$b = v_1 + \underbrace{\int_{v_2=\underline{v}}^{\bar{v}} \left\{ \int_{u=\underline{v}}^{s(v_1,v_2)} H_1(u|b) du - \int_{u=\underline{v}}^{v_2} H_2(u|b) du \right\} dF_2(v_2|v_1)}_{\text{expected benefit in A2 of winning A1}}.$$

The inverse bid function (contrast to (4)) is

$$\xi(b) \equiv b - \int_{v_2=\underline{v}}^{\bar{v}} \left\{ \int_{u=\underline{v}}^{\tilde{s}(b,v_2)} H_1(u|b) du - \int_{u=\underline{v}}^{v_2} H_2(u|b) du \right\} d\tilde{F}_2(v_2|b) = v_1.$$

For identification, the key data requirement is that all bids in the first auction be observed. Turning to the identification strategy in section 4.1, steps (i), (ii), and (iv) carry through. Only step (iii) changes; one would use the simpler inverse bid function above rather than (4) to identify the quantiles of v_1. Then all the primitives of the model are identified as before.

A.2.3 Sequence of two first-price auctions

Auction theory shows that when the second auction is a first-price auction, characterizing equilibria is even more challenging than for other sequences due to the "ratchet effect": information about private values revealed by first-auction bids affects rivals' strategies in the second auction, so bidders have an incentive to deceive their rivals, leading to pooling equilibria. Due to the difficulties, the theory literature often focuses on stylized models to garner insights, e.g. models with binary valuations. As strict monotonicity of bids is key to point identification in all structural auction analysis, it is natural that this paper's methods cannot be applied to pooling equilibria.

On the other hand, if information disclosure between auctions is limited (see Bergemann and Horner (2014) for a list of empirical examples), equilibrium can still consist of pure and monotone strategies in both auctions; Février (2003) demonstrates this for an auction in which the seller reveals the name of the first-auction winner but does not reveal first-auction bids. If bidding strategies are strictly monotonic, this paper's method applies: synergy is identified nonparametrically and distinguished from affiliation across auctions.

To illustrate, consider a sequence of first-price auctions with $N = 2$ ex-ante symmetric bidders, where only the name of the A1-winner is revealed to bidders between auctions, and all bid functions are strictly monotonic in values. The econometrician observes all bids in both auctions. Let b_{1W} and b_{1L} indicate the first-auction bids of the first-auction winner and loser, respectively. The A1-winner's synergy-inclusive value for the second item has distribution $D(\cdot|b_{1W})$, and his corresponding monotonic second-auction bids have distribution $G_{2W}(\cdot|b_{1W})$. Similarly, the A1-loser has distribution of values for second item $F_2(\cdot|b_{1L})$ and corresponding bid distribution $G_{2L}(\cdot|b_{1L})$.

Now consider the A1-winner's bidding decision in A2. He does not know the losing first-auction bid - I emphasize this with upper case B_{1L} - knowing only that it was less than his own bid. His expected profit from bidding b_2 is $\pi_W(s(v_1, v_2), b_2|b_{1W}) \equiv (s(v_1, v_2) - b_2)G_{2L}(b_2|B_{1L} < b_{1W})$. This leads to the winner's first-order condition that $s(v_1, v_2) = b_2 + G_{2L}(b_2|B_{1L} < b_{1W})/g_{2L}(b_2|B_{1L} < b_{1W})$. Meanwhile, the A1-loser does not know the winning first-auction bid, knowing only that it was greater than his own bid. His expected profit from bidding b_2 is $\pi_L(v_2, b_2|b_{1L}) \equiv (v_2 - b_2)G_{2W}(b_2|B_{1W} > b_{1L})$. This leads to the loser's first-order condition that $v_2 = b_2 + G_{2W}(b_2|B_{1W} > b_{1L})/g_{2W}(b_2|B_{1W} > b_{1L})$.

As the econometrician observes all bids, the distributions $G_{2W}(\cdot|b_{1W})$, $G_{2L}(\cdot|b_{1L})$, $G_{2W}(\cdot|B_{1W} > b_{1L})$, and $G_{2L}(\cdot|B_{1L} < b_{1W})$ are estimated directly from the data. Then, following the logic of Guerre et al. (2000), $D(\cdot|b_{1W})$ and $F_2(\cdot|b_{1L})$ are identified. Once this is done, the synergy function $s(\cdot, \cdot)$ is identified nonparametrically by comparing $D(\cdot|b_{1W})$ and $F_2(\cdot|b_{1L})$ conditional on $b_{1W} = b_{1L}$ as explained in section 4.1.

A.3 Proof of consistency of estimators for $F_2(\cdot|\cdot)$, $D(\cdot|\cdot)$, $s(\cdot, \cdot)$, and $F_1(\cdot)$

Proposition 4. *The estimators for $F_2(\cdot|\cdot)$, $D(\cdot|\cdot)$, $s(\cdot, \cdot)$, and $F_1(\cdot)$ are consistent given the following conditions.*

(C1) AS1-AS3, AS5, AS6 as presented in section 3.1.
(C2) AS4′: Either $F_2(x|v_1) = F_2(x)\ \forall x, v_1$ (independence) or $F_2(x|v_1') < F_2(x|v_1)$ if $v_1' > v_1$ $\forall x$ (strict stochastic ordering).
(C3) The estimated distribution of observed A1 bids, $\hat{G}(\cdot)$, is consistent, continuous, and has strictly positive density on the whole support.
(C4) Observations are i.i.d., where one observation refers to one pair (or sequence) of auctions A1-A2.

Recall that for the sieve maximum likelihood estimation step, bid data are normalized to have support [0,1], and $F_2(\cdot|\cdot)$, $D(\cdot|\cdot)$, and $F_1(\cdot)$ have range [0,1] as cdfs. I use the following notation in the proof. The parameter θ is an infinite-dimensional parameter representing nonparametric $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$. As the identity function is a continuous function, $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ are continuous in θ. The function $d(\cdot,\cdot)$ is a metric on parameter space Θ. Specifically, I use the uniform norm here. The function $Q(\theta) \equiv \mathbb{E}[\log(L(\theta))]$, where $L(\cdot)$, the likelihood of an A2 price and winner given θ, is defined in section 5.1. The function $\hat{Q}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \log(L_i(\theta))$ where n is the number of observations.

Proof of consistency of the sieve maximum likelihood estimators for $\tilde{F}_2(\cdot|\cdot)$ and $\tilde{D}(\cdot|\cdot)$ I establish this by checking the conditions of Corollary 2.6 of White and Wooldridge (1991) as presented by Chen (2007).
CONDITION 3.1′ *(Identification).*
(i) $Q(\theta)$ is continuous at θ_0 in Θ, $Q(\theta_0) > -\infty$;
(ii) for all $\epsilon > 0$, $Q(\theta_0) > \sup_{\{\theta\in\Theta: d(\theta,\theta_0)\geq\epsilon\}} Q(\theta)$.
CONDITION 3.2′ *(Sieve spaces).*
(i) $\Theta_k \subseteq \Theta_{k+1} \subseteq \Theta$ for all $k \geq 1$;
(ii) for any $\theta \in \Theta$ there exists $\pi_k\theta \in \Theta_k$ such that $d(\theta, \pi_k\theta) \to 0$ as $k \to \infty$.
CONDITION 3.3′ *(Continuity).* For each $k \geq 1$,
(i) $\hat{Q}_n(\theta)$ is a measurable function of the data $\{Z_t\}_{t=1}^n$ for all $\theta \in \Theta_k$;
(ii) for any data $\{Z_t\}_{t=1}^n$, $\hat{Q}_n(\theta)$ is upper semicontinuous on Θ_k under the metric $d(\cdot,\cdot)$.
CONDITION 3.4 *(Compact sieve space).* The sieve spaces, Θ_k, are compact under $d(\cdot,\cdot)$.
CONDITION 3.5(i) *(Uniform convergence over sieves).* For all $k \geq 1$, $\text{plim}_{n\to\infty} \sup_{\theta\in\Theta_k} |\hat{Q}_n(\theta) - Q(\theta)| = 0$.

By (C1), continuity of $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ in θ, and the definition of likelihood $L(\cdot)$, the following are true: (a) $\log(L(\theta))$ is continuous in θ and (b) $\mathbb{E}[\sup_\theta |\log(L(\theta))|] < \infty$. Let Θ' be a compact subset of Θ containing θ_0. Then Condition 3.1′(i) is satisfied, since $Q(\theta)$ is continuous over Θ' by Lemma 2.4 of Newey and McFadden (1994). Condition 3.1′(ii) is established by the nonparametric identification proof of Proposition 3 and the fact that for maximum likelihood, identification is a sufficient condition for a unique maximum; see Lemma 2.2 of Newey and McFadden (1994). Condition 3.2′(i) is satisfied by the definition of Bernstein polynomials shown in (8). Condition 3.2′(ii) is established by (C1) (specifically continuity in AS3) and the Stone–Weierstrass theorem (in fact, Bernstein polynomials are the polynomials constructed to prove the Weierstrass Approximation Theorem). Condition 3.3′ is satisfied immediately by (C1), continuity of $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ in θ, and the definition of likelihood $L(\cdot)$. Condition 3.4 is satisfied as follows. By properties of Bernstein polynomials (see Lorentz (1986)), the parameters $\vec{\gamma} \equiv [\gamma_0, \gamma_1, ..., \gamma_k]$ of the Bernstein polynomial of order k that approximates function $F(\cdot)$ satisfy $\gamma_i = F(\frac{i}{k})$. Since $F(\cdot)$ here is a cdf, $F(\frac{i}{k}) \in [0,1]$ and hence $\vec{\gamma} \in [0,1]^{k+1}$. Let Y be the space of Bernstein polynomials of order k, and define a mapping $m : \mathbb{R}^{k+1} \to Y$ that maps a vector of parameters $\vec{\gamma}$ to the polynomial bearing those parameters. Then map m is a continuous map between topological spaces, which has the property of preserving compactness. Now, since $[0,1]^{k+1}$, being closed and bounded, is compact in $\mathbb{R}^{k+1}$, $m([0,1]^{k+1})$ is compact in Y. Finally, defining $\Theta_k \equiv m([0,1]^{k+1})$, the sieve space Θ_k is compact in Y. Condition 3.5(i) is satisfied for any given k according to the uniform law of large numbers in Lemma 2.4 of Newey and McFadden (1994), since (a) $\log(L(\theta))$ is continuous in θ, (b) $\mathbb{E}[\sup_\theta |\log(L(\theta))|] < \infty$, and (c) Θ_k is compact. Since conditions 3.1′-3.5(i) are satisfied, $d(\hat{\theta}_n, \theta_0) = o_P(1)$ by Corollary 2.6 of White and Wooldridge (1991). Finally, by continuity of $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ in θ and the continous mapping theorem, the estimators for $\tilde{F}_2(\cdot|\cdot)$ and $\tilde{D}(\cdot|\cdot)$ are consistent.

During the remainder of the proof of consistency, I will make use of the following remark.

Remark 2. (Consistency of inverse). If $\hat{F}_n(\cdot) \xrightarrow{p} F(\cdot)$ and $F(\cdot)$, $\{\hat{F}_n(\cdot)\}_{n=1,...,\infty}$, $F^{-1}(\cdot)$ and $\{\hat{F}_n^{-1}(\cdot)\}_{n=1,...,\infty}$ exist, are well defined, are continuous in their arguments, and have compact support, then $\hat{F}_n^{-1}(\cdot) \xrightarrow{p} F^{-1}(\cdot)$.

Proof. Defining $\eta_n(v) \equiv \hat{F}_n(v) - F(v)$, $\hat{F}_n(\cdot) \xrightarrow{p} F(\cdot)$ means $\lim_{n\to\infty} \Pr(\| \eta_n(\cdot) \|_\infty > \epsilon) = 0$ for any arbitrary ϵ. Meanwhile, since $\hat{F}_n^{-1}(\cdot)$ is continuous, there exists an $\epsilon(v) > 0$ such that if $\eta_n(v) < \epsilon(v)$, $|\hat{F}_n^{-1}(F(v)) - \hat{F}_n^{-1}(F(v) + \eta_n(v))| < \delta$ for arbitrary $\delta > 0$. By definition of $\eta_n(v)$, $\hat{F}_n^{-1}(F(v) + \eta_n(v)) = \hat{F}_n^{-1}(\hat{F}_n(v)) = v = F^{-1}(F(v))$. So there exists an $\epsilon(v) > 0$ such that if $\eta_n(v) < \epsilon(v)$, $|\hat{F}_n^{-1}(F(v)) - F^{-1}(F(v))| < \delta$ for arbitrary $\delta > 0$. By the extreme value theorem, $\inf_v \epsilon(v) > 0$. Define $\epsilon \equiv \inf_v \epsilon(v)$. Now, if $\| \eta_n(\cdot) \|_\infty < \epsilon$, $\| \hat{F}_n^{-1}(\cdot) - F^{-1}(\cdot) \|_\infty < \delta$. Given that $\lim_{n\to\infty} \Pr(\| \eta_n(\cdot) \|_\infty > \epsilon) = 0$ for any arbitrary ϵ, we have that $\lim_{n\to\infty} \Pr(\| \hat{F}_n^{-1}(\cdot) - F^{-1}(\cdot) \|_\infty > \delta) = 0$ for any arbitrary δ. □

Proof of consistency of the estimator for $\tilde{s}(\cdot,\cdot)$ Recall $\tilde{s}(b, v_2) = \tilde{D}^{-1}(\tilde{F}_2(v_2|b)|b)$ and $\hat{\tilde{s}}_n(b, v_2) = \hat{\tilde{D}}_n^{-1}(\hat{\tilde{F}}_{2n}(v_2|b)|b)$. By AS3, the inverse $\tilde{D}^{-1}(\cdot|\cdot)$ exists, is well defined, and is continuous in its arguments. Moreover, it has compact domain and range. Therefore, by

Remark 2 and consistency of $\hat{\tilde{D}}_n(\cdot|\cdot)$, $\hat{\tilde{D}}_n^{-1}(\cdot|\cdot) \xrightarrow{p} \tilde{D}^{-1}(\cdot|\cdot)$. Consequently, $\hat{\tilde{D}}_n^{-1}(\hat{\tilde{F}}_{2n}(\cdot|\cdot)|\cdot) \xrightarrow{p} \tilde{D}^{-1}(\hat{\tilde{F}}_2(\cdot|\cdot)|\cdot)$. Meanwhile, $\tilde{D}^{-1}(\hat{\tilde{F}}_{2n}(\cdot|\cdot)|\cdot) \xrightarrow{p} \tilde{D}^{-1}(\tilde{F}_2(\cdot|\cdot)|\cdot)$ by the continuous mapping theorem,. Therefore, $\hat{\tilde{s}}_n(b, v_2) \xrightarrow{p} \tilde{s}(b, v_2)$.

Proof of consistency of the estimators for $H_1(\cdot|\cdot)$ and $H_2(\cdot|\cdot)$ Since $H_1(u|t)$ and $H_2(u|t)$ are products of $\tilde{F}_2(u|t)$, $\tilde{D}(u|t)$, and $\tilde{F}_2(u|b \leq t)$, by the continuous mapping theorem they are consistently estimated if each component is consistently estimated. Having already established the consistency of estimators for $\tilde{F}_2(\cdot|\cdot)$ and $\tilde{D}(\cdot|\cdot)$, I show below that the estimator for $\tilde{F}_2(u|b \leq t)$ is consistent. Since $\tilde{F}_2(u|b \leq t) = \tilde{F}_2(u)$ if $v_1 \perp v_2$, I focus on the case in which v_2 is strictly stochastically ordered in v_1 (see (C2) above).

Proof of consistency of the estimator for $\tilde{F}_2(u|b \leq t)$ Recall $\tilde{F}_2(u|b \leq t) \equiv \int_{b=\underline{b}}^{t} \tilde{F}_2(u|b)dG(b)/G(t)$. I show that $\int_{b=\underline{b}}^{t} \hat{\tilde{F}}_{2n}(u|b)d\hat{G}(b)/\hat{G}(t) \xrightarrow{p} \int_{b=\underline{b}}^{t} \tilde{F}_2(u|b)dG(b)/G(t)$. Note that since $\tilde{F}_2(u|b)$ is strictly decreasing in b by (C2), $\Pr[\tilde{F}_2(u|b) \leq x] = \Pr[b \leq \tilde{F}_2^{-1,b}(x|u)] = G(\tilde{F}_2^{-1,b}(x|u))$, where $\tilde{F}_2^{-1,b}(\cdot|u)$ is the inverse of $\tilde{F}_2(u|\cdot)$ and consistently estimated according to Remark 2. Then by the mathematical property that $\mathbb{E}[z] \equiv \int z dF_Z(z) = \int (1 - F_Z(z))dz$ for $z \geq 0$, $\int_{b=\underline{b}}^{t} \tilde{F}_2(u|b)dG(b) = \int_{\tilde{F}_2(u|\underline{b})}^{\tilde{F}_2(u|t)} (1 - G(\tilde{F}_2^{-1,b}(x|u)))dx$. If $\| \hat{G}(\cdot) - G(\cdot) \|_\infty < \epsilon$, $| \int_{\hat{\tilde{F}}_2(u|\underline{b})}^{\hat{\tilde{F}}_2(u|t)} (1 - \hat{G}(\hat{\tilde{F}}_{2n}^{-1,b}(x|u)))dx - \int_{\hat{\tilde{F}}_2(u|\underline{b})}^{\hat{\tilde{F}}_2(u|t)} (1 - G(\hat{\tilde{F}}_{2n}^{-1,b}(x|u)))dx | < \int_{\hat{\tilde{F}}_2(u|\underline{b})}^{\hat{\tilde{F}}_2(u|t)} \epsilon dx = \epsilon[\hat{\tilde{F}}_{2n}(u|t) - \hat{\tilde{F}}_{2n}(u|\underline{b})] \leq \epsilon$ for all u, t. This means $\int_{\hat{\tilde{F}}_2(u|\underline{b})}^{\hat{\tilde{F}}_2(u|t)} (1 - \hat{G}(\hat{\tilde{F}}_{2n}^{-1,b}(x|u)))dx$ is continuous in $\hat{G}(\cdot)$ under the uniform norm. It is also continuous in $\hat{\tilde{F}}_{2n}^{-1,b}(\cdot|\cdot)$ and $\hat{\tilde{F}}_{2n}(\cdot|\cdot)$. Then by the continuous mapping theorem, $\int_{\hat{\tilde{F}}_{2n}(u|\underline{b})}^{\hat{\tilde{F}}_{2n}(u|t)} (1 - \hat{G}(\hat{\tilde{F}}_{2n}^{-1,b}(x|u)))dx \xrightarrow{p} \int_{\tilde{F}_2(u|\underline{b})}^{\tilde{F}_2(u|t)} (1 - G(\tilde{F}_2^{-1,b}(x|u)))dx$. Therefore, $\int_{b=\underline{b}}^{t} \hat{\tilde{F}}_2(u|b)d\hat{G}(b) \xrightarrow{p} \int_{b=\underline{b}}^{t} \tilde{F}_2(u|b)dG(b)$. I conclude that $\hat{\tilde{F}}_{2n}(u|b \leq t) \xrightarrow{p} \tilde{F}_2(u|b \leq t)$.

Proof of consistency of the estimator for $\xi(b)$ Define $\psi_1(b, v_2) \equiv \int_{u=\underline{v}}^{\tilde{s}(b,v_2)} H_1(u|b)du$ and $\psi_2(b, v_2) \equiv \int_{u=\underline{v}}^{v_2} H_2(u|b)du$. $\psi_1(\cdot, \cdot)$ and $\psi_2(\cdot, \cdot)$ are continuous in $\tilde{s}(\cdot, \cdot)$ and $H_1(\cdot|\cdot)$ or $H_2(\cdot|\cdot)$ under the uniform norm and have compact domain and range. By the continuous mapping theorem, $\hat{\psi}_{1n}(\cdot|\cdot) \xrightarrow{p} \psi_1(\cdot|\cdot)$ and $\hat{\psi}_{2n}(\cdot|\cdot) \xrightarrow{p} \psi_2(\cdot|\cdot)$.

Now define $\psi_3(b) \equiv \int_{v_2=\underline{v}}^{\bar{v}} \psi_1(b, v_2)d\tilde{F}_2(v_2|b)$. I proceed to show that $\psi_3(b)$ is consistently estimated. If $\frac{\partial \tilde{s}(b,v_2)}{\partial v_2} = 0$, $\psi_3(b) = \psi_1(b)$, which has already been shown to be consistently estimated. So I focus on the case in which $\frac{\partial \tilde{s}(b,v_2)}{\partial v_2} > 0$, and show that $\hat{\psi}_{3n}(b) \equiv \int_{v_2=\underline{v}}^{\bar{v}} \hat{\psi}_{1n}(b, v_2)d\hat{\tilde{F}}_{2n}(v_2|b) \xrightarrow{p} \int_{v_2=\underline{v}}^{\bar{v}} \psi_1(b, v_2)d\tilde{F}_2(v_2|b)$. I make use of the mathematical property that $\mathbb{E}[z] \equiv \int z dF_Z(z) = \int (1 - F_Z(z))dz$ for $z \geq 0$. Given $\frac{\partial \tilde{s}(b,v_2)}{\partial v_2} > 0$, $\psi_1(b, v_2)$ is strictly increasing in v_2, so the inverse with respect to v_2 given b, $\psi_1^{-1,v_2}(\cdot|b)$, exists and is well defined. Moreover by the properties of $\psi_1(\cdot, \cdot)$, $\psi_1^{-1,v_2}(\cdot|\cdot)$ is continuous in its arguments, has compact support, and is therefore consistently estimated by

Remark 2. Now, $\Pr[\psi_1(b,v_2) \leq x] = \Pr[v_2 \leq \psi_1^{-1,v_2}(x|b)] = \tilde{F}_2(\psi_1^{-1,v_2}(x|b)|b)$. Then $\int_{v_2=\underline{v}}^{\bar{v}} \psi_1(b,v_2) d\tilde{F}_2(v_2|b) = \int_{x=\psi_1(b,\underline{v})}^{\psi_1(b,\bar{v})} [1 - \tilde{F}_2(\psi_1^{-1,v_2}(x|b)|b)]dx$. Since this last expression is continuous in $\psi_1^{-1,v_2}(\cdot|\cdot)$, $\psi_1(\cdot,\cdot)$, and $\tilde{F}_2(\cdot|\cdot)$ under the uniform norm, $\int_{x=\hat{\psi}_{1n}(b,\underline{v})}^{\hat{\psi}_{1n}(b,\bar{v})} [1 - \hat{\tilde{F}}_{2n}(\hat{\psi}_{1n}^{-1,v_2}(x|b)|b)]dx \xrightarrow{p} \int_{x=\psi_1(b,\underline{v})}^{\psi_1(b,\bar{v})} [1 - \tilde{F}_2(\psi_1^{-1,v_2}(x|b)|b)]dx$ by the continuous mapping theorem. This is equivalent to $\int_{v_2=\underline{v}}^{\bar{v}} \hat{\psi}_{1n}(b,v_2) d\hat{\tilde{F}}_{2n}(v_2|b) \xrightarrow{p} \int_{v_2=\underline{v}}^{\bar{v}} \psi_1(b,v_2) d\tilde{F}_2(v_2|b)$, i.e. $\hat{\psi}_{3n}(b) \xrightarrow{p} \psi_3(b)$.

Following analogous logic, define $\psi_4(b) \equiv \int_{v_2=\underline{v}}^{\bar{v}} \psi_2(b,v_2) d\tilde{F}_2(v_2|b)$. I proceed to show that $\psi_4(b)$ is consistently estimated. $\psi_2(b,v_2)$ is strictly increasing in v_2, so the inverse with respect to v_2 given b, $\psi_2^{-1,v_2}(\cdot|b)$, exists and is well defined. Moreover by the properties of $\psi_2(\cdot,\cdot)$, $\psi_2^{-1,v_2}(\cdot|\cdot)$ is continuous in its arguments, has compact support, and is therefore consistently estimated by Remark 2. Now, $\Pr[\psi_2(b,v_2) \leq x] = \Pr[v_2 \leq \psi_2^{-1,v_2}(x|b)] = \tilde{F}_2(\psi_2^{-1,v_2}(x|b)|b)$. Then $\int_{v_2=\underline{v}}^{\bar{v}} \psi_2(b,v_2) d\tilde{F}_2(v_2|b) = \int_{x=\psi_2(b,\underline{v})}^{\psi_2(b,\bar{v})} [1 - \tilde{F}_2(\psi_2^{-1,v_2}(x|b)|b)]dx$. Since this last expression is continuous in $\psi_2^{-1,v_2}(\cdot|\cdot)$, $\psi_2(\cdot,\cdot)$, and $\tilde{F}_2(\cdot|\cdot)$ under the uniform norm, $\int_{x=\hat{\psi}_{2n}(b,\underline{v})}^{\hat{\psi}_{2n}(b,\bar{v})} [1 - \hat{\tilde{F}}_{2n}(\hat{\psi}_{2n}^{-1,v_2}(x|b)|b)]dx \xrightarrow{p} \int_{x=\psi_2(b,\underline{v})}^{\psi_2(b,\bar{v})} [1 - \tilde{F}_2(\psi_2^{-1,v_2}(x|b)|b)]dx$ by the continuous mapping theorem. This is equivalent to $\int_{v_2=\underline{v}}^{\bar{v}} \hat{\psi}_{2n}(b,v_2) d\hat{\tilde{F}}_{2n}(v_2|b) \xrightarrow{p} \int_{v_2=\underline{v}}^{\bar{v}} \psi_2(b,v_2) d\tilde{F}_2(v_2|b)$, i.e. $\hat{\psi}_{4n}(b) \xrightarrow{p} \psi_4(b)$.

Finally, $\hat{\xi}(b) \equiv b + \frac{\hat{G}(b)}{(N-1)\hat{g}(b)} - \hat{\psi}_{3n}(b) + \hat{\psi}_{4n}(b)$. By the continous mapping theorem, $\hat{\xi}(\cdot) \xrightarrow{p} \xi(\cdot)$.

This inverse bid function $\xi(\cdot)$ is strictly increasing and continuous in its argument with compact domain and range. Hence the bid function $\xi^{-1}(\cdot)$ exists, is continuous in its argument and has compact domain and range. By Remark 2, the estimator for $\xi^{-1}(\cdot)$ is also consistent.

Proof of consistency of the estimators for $F_1(\cdot)$ and $F_2(\cdot|\cdot)$, $D(\cdot|\cdot)$, $s(\cdot,\cdot)$
Recall $F_1(v_1) = G(\xi^{-1}(v_1))$. As both $G(\cdot)$ and $\xi^{-1}(\cdot)$ are consistently estimated, $\hat{F}_1(\cdot) = \hat{G}(\hat{\xi}_n^{-1}(\cdot)) \xrightarrow{p} G(\hat{\xi}_n^{-1}(\cdot))$ and $G(\hat{\xi}_n^{-1}(\cdot)) \xrightarrow{p} G(\xi^{-1}(\cdot))$ by the continuous mapping theorem. Hence $\hat{F}_1(\cdot) \xrightarrow{p} F_1(\cdot)$.

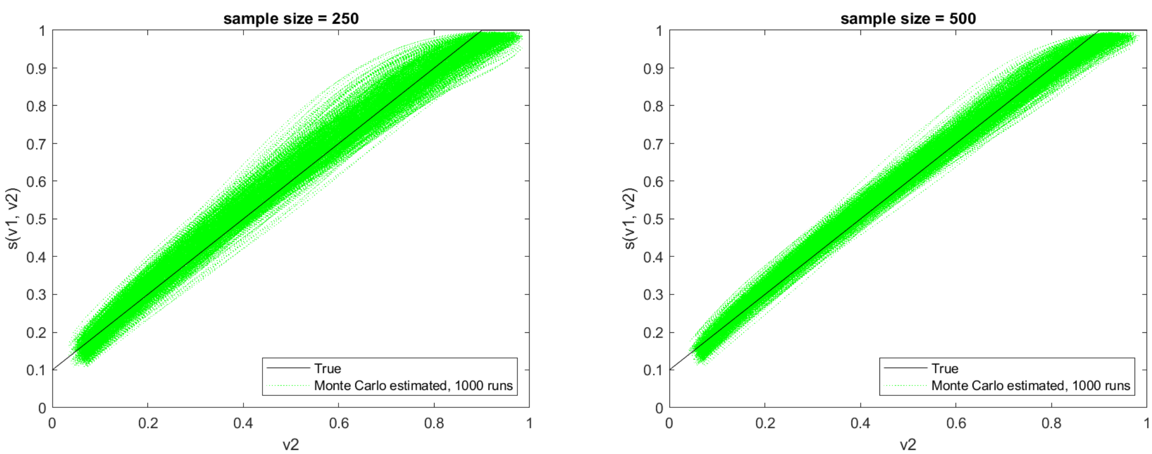
Recall $F_2(v_2|v_1) = \tilde{F}_2(v_2|\xi^{-1}(v_1))$. As both $\tilde{F}_2(\cdot|\cdot)$ and $\xi^{-1}(\cdot)$ are consistently estimated, $\hat{F}_2(\cdot|\cdot) = \hat{\tilde{F}}_{2n}(\cdot|\hat{\xi}_n^{-1}(\cdot)) \xrightarrow{p} \tilde{F}_2(\cdot|\hat{\xi}_n^{-1}(\cdot))$ and $\tilde{F}_2(\cdot|\hat{\xi}_n^{-1}(\cdot)) \xrightarrow{p} \tilde{F}_2(\cdot|\xi^{-1}(\cdot))$ by the continuous mapping theorem. Hence $\hat{F}_2(\cdot|\cdot) \xrightarrow{p} F_2(\cdot|\cdot)$. The argument for consistency of $D(\cdot|\cdot)$ is exactly analogous.

Recall $s(v_1,v_2) = \tilde{s}(\xi^{-1}(v_1),v_2)$. As both $\tilde{s}(\cdot,\cdot)$ and $\xi^{-1}(\cdot)$ are consistently estimated, $\hat{s}(\cdot,\cdot) = \hat{\tilde{s}}_n(\hat{\xi}_n^{-1}(\cdot),\cdot) \xrightarrow{p} \tilde{s}(\hat{\xi}_n^{-1}(\cdot),\cdot)$ and $\tilde{s}(\hat{\xi}_n^{-1}(\cdot),\cdot) \xrightarrow{p} \tilde{s}(\xi^{-1}(\cdot),\cdot)$ by the continuous mapping theorem. Hence $\hat{s}(\cdot,\cdot) \xrightarrow{p} s(\cdot,\cdot)$.

A.4 Monte Carlo studies

I conduct Monte Carlo studies to evaluate the ability of the estimation procedure to recover the synergy function. The model underlying the simulation is specified as follows: $F(v_1, v_2)$ is represented by Epanechnikov marginal distributions on support [0,1] and a Gaussian copula with Kendall's tau of 0.3, constant synergy subject to support [0,1] i.e. $s(v_1, v_2) = \min(v_2 + 0.1, 1)$, and $N = 2$. Figure 5 displays the synergy functions estimated from 1000 Monte Carlo runs against the true synergy function used to generate the data. The sieve orders were selected by minimizing the criterion $0.5k - 2\ln(L)$, where k is the number of parameters and $ln(L)$ is the log likelihood of the data. This is the same criterion I use in my application. I compare this criterion against other criteria below.

Figure 5: Monte Carlo estimates of synergy function



Selection criteria for sieve orders I compare the following selection criteria: the AIC (Akaike information criterion), BIC (Bayesian information criterion), and two other criteria that have a weaker penalty on the number of parameters k, namely "IC3" = $k - 2ln(L)$ and "IC4" = $0.5k - 2ln(L)$. To perform this comparison, I first simulate the (integrated) MSE of estimators for $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ over 1000 Monte Carlo runs, varying sample size for each model selection criterion. Specifically, the MSE is computed as $\frac{1}{m}\sum_{i=1}^{m} \iint \{\hat{F}_{2i}(v_2|v_1) - F_2(v_2|v_1)\}^2 d^2F(v_1, v_2)$ for $F_2(\cdot|\cdot)$ and analogously for $D(\cdot|\cdot)$, where m is the number of Monte Carlo runs. For the largest sample size tried of 10,000, I use 750 rather than 1000 Monte Carlo runs due to longer computational time. Next, I hold the sample size constant at 500 and repeat the MSE computation using a different $F(v_1, v_2)$ distribution: normal marginal distributions distributed $\sim N(0.5, 0.15)$ related by a Clayton copula with Kendall's tau of 0.3. Figures 6 and 7 display the results. The criterion $0.5k - 2\ln(L)$ generates the smallest mean squared error across these specifications.

Figure 6: Selection criterion comparison by mean squared error, varying sample size

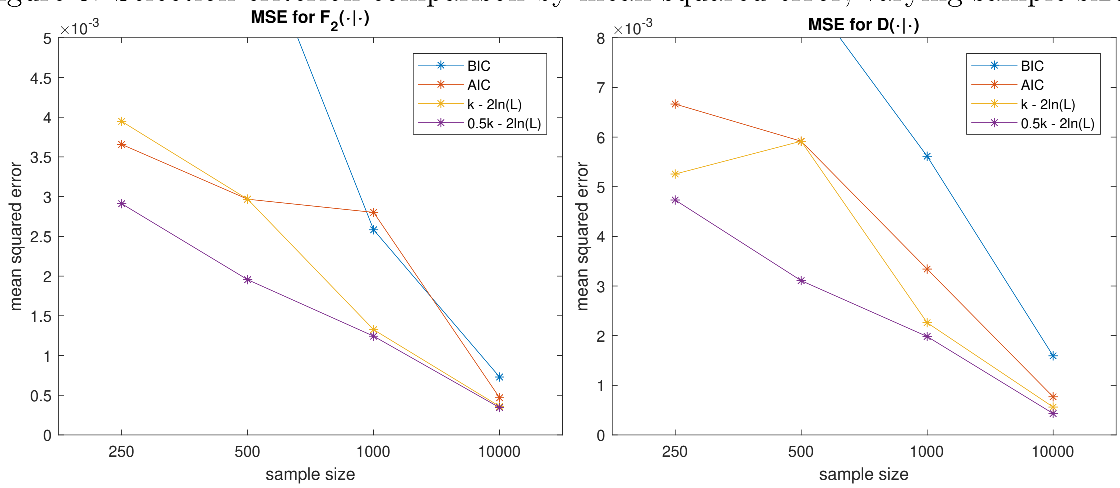
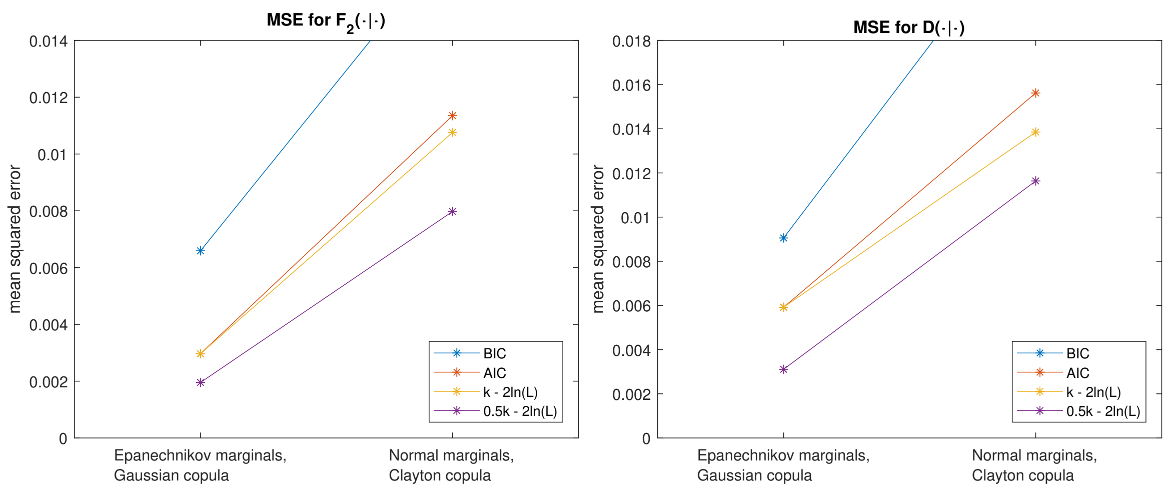
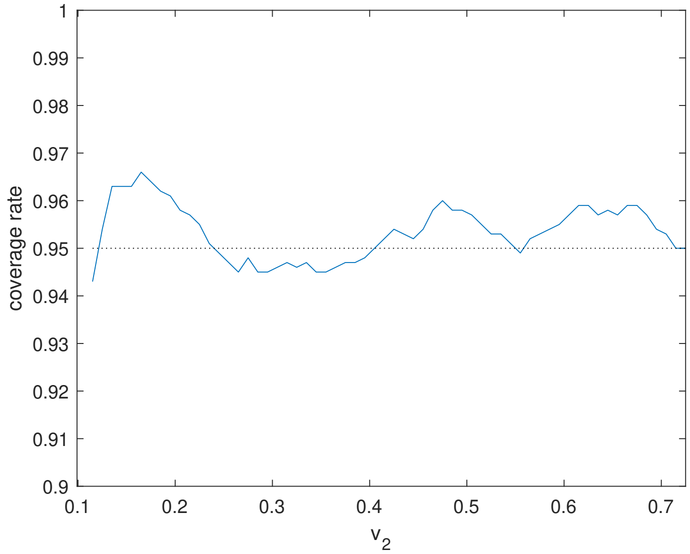


Figure 7: Selection criterion comparison by mean squared error, varying distribution



Bootstrap percentile interval coverage rate I also use the model specified above to perform a Monte Carlo simulation of the coverage rate of a 95% bootstrap percentile interval for $s(v_1, v_2)$. The simulated sample size is 500. I conduct 1000 Monte Carlo runs of estimating $s(v_1, v_2)$, in which I compute a bootstrap percentile interval based on 1000 bootstrap samples of each Monte Carlo run, i.e. I estimate $s(v_1, v_2)$ a total of 1 million times, each time using a different bootstrapped Monte Carlo data sample. Then, for a grid of v_2 values, I compute the probability across the 1000 Monte Carlo runs that the 95% bootstrap percentile interval generated from that run contains the true $s(\text{median } v_1, v_2)$ used to generate the data. Note that for each simulated dataset, functions of v_2 are identified only for v_2 in the A2 price support of that dataset. Therefore, the coverage rate is computed for values of v_2 in the intersection of the 1 million different A2 price supports generated by the Monte Carlo-of-bootstrap procedure. Figure 8 displays the coverage rates computed. The average coverage rate across a uniformly spaced grid of v_2 is 0.9534. It appears that the bootstrap interval behaves well.

Figure 8: Coverage rate of 95% bootstrap percentile interval for $s(v_1, v_2)$



A.5 Additional empirical details

Table 8: Probit regression results for probability of winning second auction

	(1) $N \geq 2$	(2) $N \geq 2$	(3) $N \geq 2$	(4) $N = 2$	(5) $N = 3$
Won first auction	1.561	2.045	2.041	1.723	1.769
	(0.093)	(0.197)	(0.201)	(0.169)	(0.194)
Number of bidders fixed effects	Y	Y	Y	-	-
Bidder fixed effects	Y	N	N	Y	Y
Bidder-date fixed effects	N	Y	Y	N	N
Lease descriptive covariates†	N	N	Y	N	N
Observations	1557	612††	612	381	405

Standard errors in parentheses

†Lease descriptive covariates are described in section 6.2.

††Adding bidder-date fixed effects decreases the number of usable observations.

Comparing post-auction production data for A1 and A2 leases

I look to post-auction production data for extra-auction evidence that $F_1(\cdot)$ and $F_2(\cdot)$ should be similar. In Table 9, an OLS regression of log production (in barrel of oil equivalents from the auction date through 2014) on a dummy variable for A1 (first auction) indicates no statistically significant difference between tracts auctioned in A1 versus in A2. In the same table, a probit regression of whether production exists similarly indicates no statistically significant difference. Finally, in Figure 9 the histograms compare the distribution of log production, conditional on production being positive, for A1 tracts and A2 tracts, respectively. The production distributions look similar visually. All of this suggests that there is no difference in quality between A1 and A2 leases.

Table 9: Regression of log production on A1 dummy

	OLS ln(prod)	probit prod>0
A1 dummy	-0.156	-0.066
	(0.220)	(0.077)
Constant	1.788	-1.112
	(0.155)	(0.054)
Observations	1744	1744

Standard errors in parentheses.

Figure 9: Histograms of log production, A1 and A2

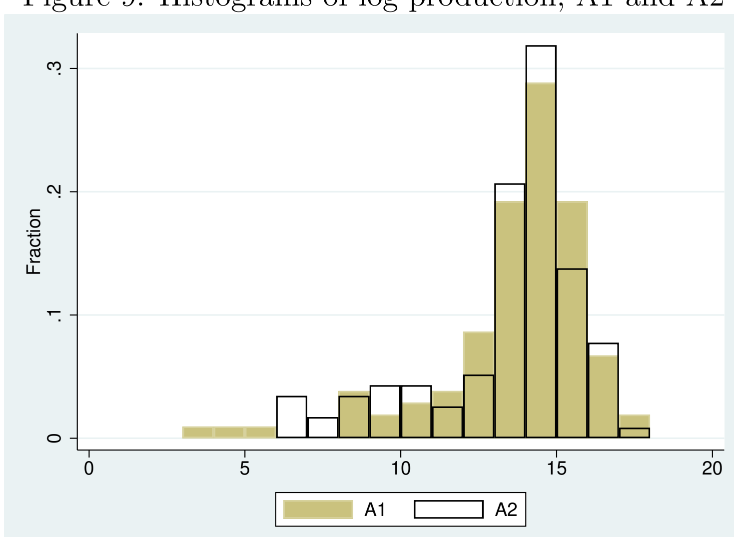
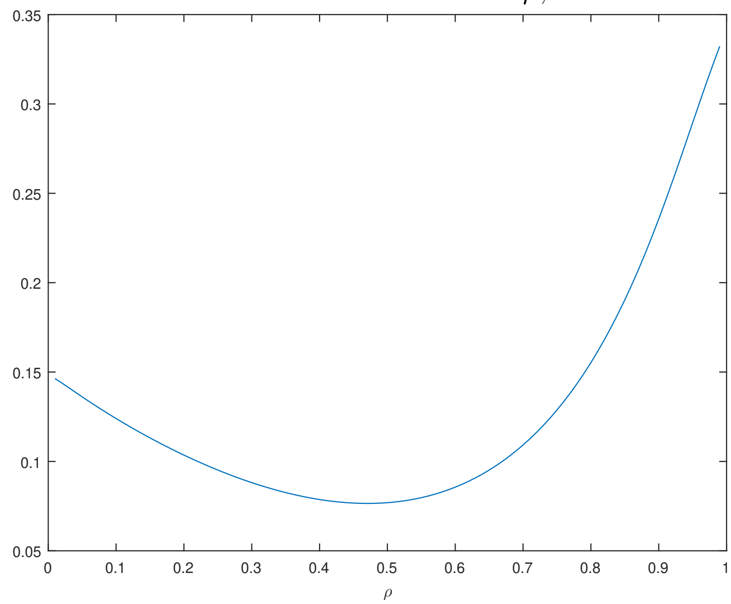


Table 10: Regression of ln(sealed bid) on observable characteristics

	lnbid
heatmap index	0.710
	(0.057)
ln(oil prod) 1970-auction date	0.022
	(0.014)
ln(oil prod) auction date-2014	0.012
	(0.007)
lease prefix V0 pre-2005	0.091
	(0.065)
lease prefix VB pre-2005	0.102
	(0.280)
lease prefix VB post-2005	0.330
	(0.076)
ln(nat gas 1 mo futures)	0.317
	(0.132)
ln(WTI oil price)	0.134
	(0.221)
ln(prior month price/acre)	0.123
	(0.061)
ln(BLM price/acre)	0.267
	(0.055)
Constant	0.157
	(0.966)
Year fixed effects	Y
Month fixed effects (seasonality)	Y
Observations	2095
R^2	0.261
Adjusted R^2	0.248

Heteroskedasticity robust standard errors in parentheses

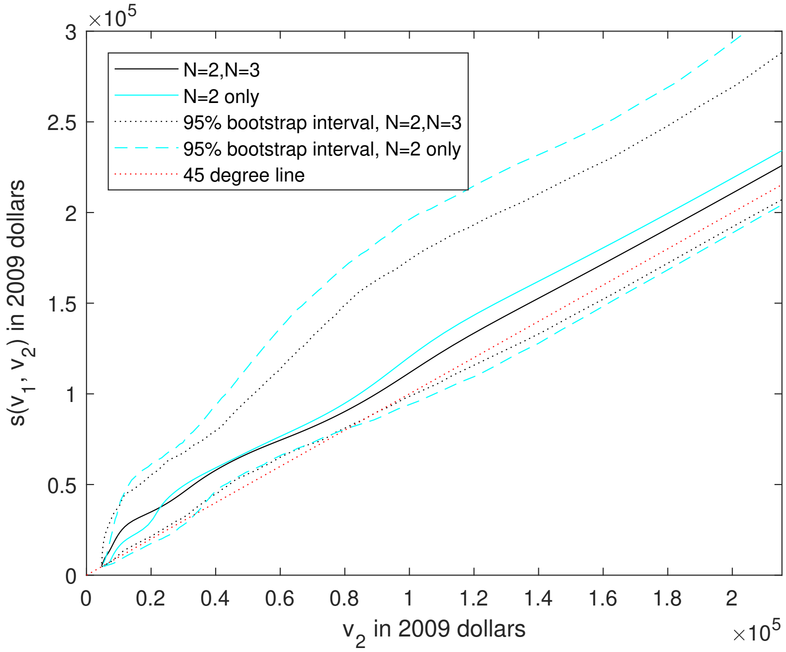
Figure 10: Minimand of the estimator for ρ, as a function of ρ



A.6 Alternative specifications

I estimate the synergy function under three alternative specifications to see how much the estimates differ from that of Figure 3. First, I restrict the estimation sample to auctions with two bidders ($N = 2$) only. This is a subsample for which the evidence in Table 2 and A2 prices above the minimum acceptable bid provide sufficient confidence that A1 and A2 typically share the same bidders. The $N = 3$ sample does not share this property because there is no way to disprove attrition, i.e. from three bidders in A1 to two bidders in A2. The caveat is that this reduces the estimation sample to 250 pairs of auctions, leading to less precision (wider bootstrap intervals). Figure 11 displays the resulting synergy plot. Recalling that the 90th percentile of v_2 is roughly \$85,000, the $N = 2$ synergy estimate diverges from the pooled estimate at large values of v_2 where data is sparse and confidence intervals are wider. On the other hand, it is close to the pooled estimate below the 90th percentile where data is concentrated.

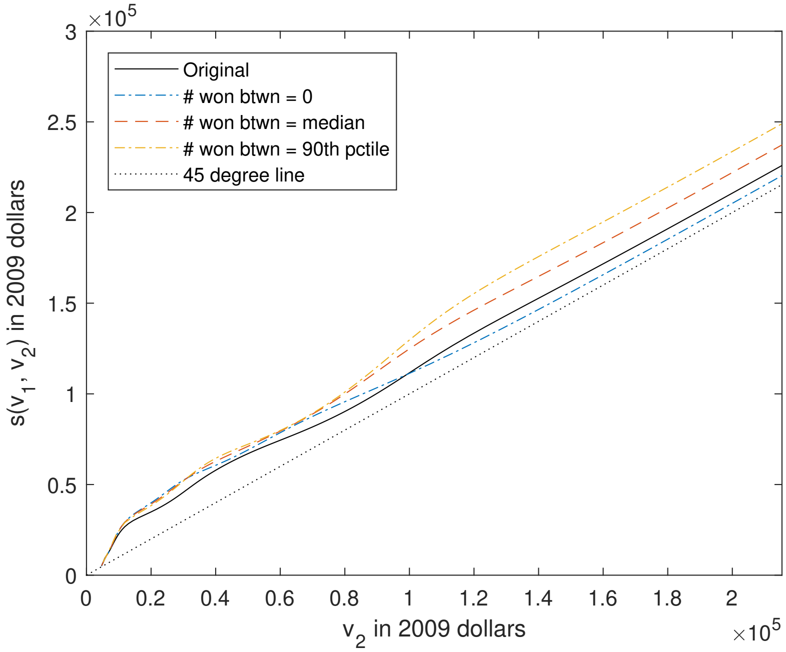
Figure 11: $\hat{s}(v_1, v_2)$ at median v_1 and $z'\beta$, $N = 2$ only



Second, I repeat the structural estimation procedure conditioning $F_2(\cdot|\cdot)$ and $D(\cdot|\cdot)$ on the number of auctions a bidder wins between A1 and A2, w; that is, I estimate $F_2(\cdot|v_1, w)$ and $D(\cdot|v_1, w)$ and the resulting synergy function. A caveat is that this leads to using trivariate Bernstein polynomials rather than bivariate in the sieve maximum likelihood estimation step, increasing the number of parameters to be estimated and reducing precision.[31] The median value of w is 3. Figure 12 plots at different values of w the synergy functions estimated conditional on w along with the original, unconditional estimate. Up to the 90th percentile of v_2, synergy estimates do not vary much with w. At very large values of v_2, synergy estimates increase with w. This suggests that winning other auctions increases the boost in value of the second item that comes from winning A1, perhaps due to combined synergies across multiple items. However, data is sparse for $v_2 > \$85,000$. For $v_2 < \$85,000$ where data is concentrated, the synergy plots for different values of w overlap.

[31] I choose a polynomial degree of 1 for w, because even this low degree doubles the number of polynomial parameters to be estimated.

Figure 12: $\hat{s}(v_1, v_2)$ at median v_1 and $z'\beta$, conditional on number of auctions won



Third, rather than excluding observations in which the A2 winner did not bid in A1 (refer to section 6.1), I include these observations and estimate a model in which there is always one additional bidder in A2 than there is in A1. The "extra" bidder draws his value from a distribution $F_{ex}(\cdot)$, which I estimate via a Bernstein polynomial of degree 10. Figure 13 displays the resulting synergy plot. Including the extra data makes little difference to the estimated synergy function.

Figure 13: $\hat{s}(v_1, v_2)$ at median v_1 and $z'\beta$, allowing for extra A2 bidder