A CONSISTENT NONPARAMETRIC TEST OF AFFILIATION IN AUCTION MODELS

SUNG JAE JUN, JORIS PINKSE AND YUANYUAN WAN

THE CENTER FOR AUCTIONS, PROCUREMENTS AND COMPETITION POLICY

DEPARTMENT OF ECONOMICS

THE PENNSYLVANIA STATE UNIVERSITY

ABSTRACT. We propose a new nonparametric test of affiliation, a strong form of positive dependence with independence as a special, knife–edge, case. The test is consistent against all departures from the null of affiliation, and its null distribution is standard normal. Like most nonparametric tests, a sample–size dependent input parameter is needed. We provide an informal procedure for choosing the input parameter and evaluate the test's performance using a simulation study. Our test can be used to test the fundamental assumptions of the auctions literature. We implement our test empirically using the Outer Continental Shelf (OCS) auction data.

Key Words: Auction Models; Affiliation; Empirical Distributions; Nonparametric Tests.

JEL Classification Codes: C12; C14.

Date: April 2010.

Jun (sjun@psu.edu) is the corresponding author. This paper is in part based on research supported by NSF grant SES-0922127. We thank the Human Capital Foundation for their support of CAPCP. We thank Ken Hendricks, Luciano de Castro, Vijay Krishna, Yosef Rinott, the associate editor, and two anonymous referees for helpful comments.

1. Introduction

In this paper we develop a test for the *affiliation* (Milgrom and Weber, 1982, MW) of the elements of observable vector–valued independent and identically distributed (i.i.d.) draws $\xi_1, \xi_2, \ldots \in \Re^d$ from a joint distribution F. Affiliation is a strong form of positive dependence with independence as a knife–edge case. We delay a formal definition of the notion of affiliation until the next section and for now focus on the properties of the test and its implications for the auctions literature.

Our test is consistent in the sense that it rejects any departure from the null hypothesis of affiliation with probability approaching one as the sample size n tends to infinity. It is nonparametric because it does not require parametric assumptions on the form of F. Indeed, it can be used for virtually any distribution F. In particular, the elements of ξ_1 can be any combination of continuous, discrete, or mixed variables. So provided that n is sufficiently large, the test proposed in this paper rejects at no more than (approximately) the specified significance level if the elements of ξ_1 are affiliated and with near certainty if they are not affiliated.

Our test is a general test of affiliation, which does not require any assumptions on the model structure. However, it should be emphasized that the relevance of the test must be understood within the context of a particular model, i.e. conditional on a set of maintained assumptions inherent in the model. This issue will be discussed in more detail later.

The proposed test is nonparametric and is based on the integration of (weighted) differences of products of indicator functions. The test requires the choice of a sample–size–dependent input parameter β_n. Although it is possible to construct a test without using an input parameter, we introduce β_n to achieve an asymptotically pivotal (standard normal) null distribution. Having a pivotal null distribution is helpful because it makes inference easier and it opens up the possibility of *asymptotic refinements* via the *bootstrap*.[1] We provide an informal argument on how to choose β_n. Simulation results show that the size of the test is relatively flat over a large range of β_n, but that the power drops off significantly if β_n is chosen too large. Other simulation results are encouraging for the performance of our test.

There are other ways of testing for affiliation. However, ours is the only nonparametric test on observables which is consistent when at least one of the variables is continuous. If all variables to be tested are discrete, then a method like Roosen and Hennessy (2004) or de Castro and Paarsch (2008)

[1] With asymptotic refinements the actual probability of rejection under the null hypothesis converges faster to the significance level than without.

can be used. Li and Zhang's (2008) test of affiliation is parametric and framed in the context of a specific entry model and is hence in a class of its own.[2]

We now discuss the relevance of our test for empirical studies of auctions of single indivisible units. The primitives of the model consist of L, the number of potential bidders, F, the joint distribution of the bidders' private (one–dimensional) signals, and (if present) a payoff–relevant variable that is common to every bidder. Bidders behave non–cooperatively and are assumed to be risk–neutral. The structural (econometric) exercise consists of estimating F on the basis of the bid data. Identification and estimation of F depends critically on the existence of a monotone pure strategy equilibrium. MW have shown that such an equilibrium exists in a symmetric environment if the private signals (and common value) are affiliated. Recently, Reny and Zamir (2004) extend their result to asymmetric environments, albeit only for first price auctions.

If bids are monotonic functions of signals then affiliation of signals implies affiliation of bids. A rejection of the test of affiliation of bids constitutes a rejection of the model. Signals may not be affiliated, in which case equilibrium bid functions may not be monotonic, or bidders may not be bidding competitively. In either case, estimating F from data on bids under the assumption that bidders are playing best replies against the empirical distribution of rival bids is called into question. What to do next is not clear. There is some work by de Castro (2007) and Monteiro and Moreira (2006) that investigates the existence of monotone bid functions in specific auction environments where signals are not affiliated.

In practice one observes data from a sequence of possibly heterogeneous auctions. Even if signals (and hence bids) are affiliated *conditional on* such heterogeneity, they may not be affiliated unconditionally. In other words, it is possible that signals are affiliated for each auction, but that such affiliation disappears upon aggregation across auctions.[3] It is likewise not necessary for signals to be affiliated conditional on heterogeneity in order for affiliation to hold unconditionally. In applications, it is therefore advisable to correct for any such heterogeneity before applying our test, e.g. by conditioning on covariates.

There are instances in which not all bids are observed for each auction. To test the affiliation of bids one clearly needs to observe at least two bids per auction. In principle these can be the winning bid and next highest bid, but their affiliation does not imply affiliation of the entire vector

[2]Of these, only Roosen and Hennessy (2004) predates our work.

[3]For instance, let s_1, s_2 are binary 0/1 variables with joint (conditional) probability function p and heterogeneity is represented by a binary 0/1 variable z taking either value with equal probability. Then, if $p(0,0|0) = p(0,0|1) = p(1,1|0) = p(1,1|1) = 0.1$, $p(1,0|0) = p(0,1|1) = 0.8$, $p(1,0|1) = p(0,1|0) = 0$, affiliation holds conditionally since $0.1 \times 0.1 - 0.8 \times 0 > 0$, but not unconditionally since $p(0,0) = p(1,1) = 0.1$, $p(1,0) = p(0,1) = 0.4$, and $0.1 \times 0.1 - 0.4 \times 0.4 < 0$.

of bids, or indeed signals. Likewise, two randomly selected bids per auction can be affiliated absent affiliation of the entire vector of bids. Under symmetry, since marginal operations preserve affiliation (Karlin and Rinott, 1980), the affiliation of the entire bid vector implies both the affiliation of two randomly selected bids and the affiliation of the two highest bids. If all bids are independent then two randomly selected bids are also independent whereas the two highest bids are affiliated, but not independent. Because independence is the knife–edge case, we expect a test of affiliation of two randomly selected bids to have more power than a test using the two highest bids.

Thus far the focus has been on testing the affiliation of bids. It is possible to test affiliation of other variables, e.g. of bids and the potential number of bidders L. For instance, Hendricks and Porter (1988) and Hendricks, Pinkse, and Porter (2003) provide evidence that the auctions in the Outer Continental Shelf (OCS) data set fall within the (pure) common values (CV) paradigm in which bidders' values are equal to the common value; their signal is informative about the common value through their affiliation with the common value. With common value auctions, bids and L are not affiliated. So in the OCS environment, given the external evidence in support of the common value paradigm, a failure to reject affiliation of bids and L suggests a problem with one of the other MW assumptions. If L is not observed, one can instead test the affiliation of the actual number of bidders L_a and bids; in the MW environment affiliation of bids and L implies affiliation of bids and L_a but not vice versa.

There are methods of testing the assumptions of an auction model other than testing for affiliation. Indeed, with some further assumptions (e.g. regarding the paradigm) one can test for the monotonicity of bid functions; see e.g. Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2005). Within the pure common value paradigm the equilibrium first order conditions provide a further way of testing the theory (see e.g. Hendricks, Pinkse, and Porter, 2003). After having tested (and having failed to reject) affiliation, such further tests may well be helpful.

The remainder of the paper is organized as follows. We present our test in section 2. An assumption needed for our test is explained in section 3. Finally, in sections 6 and 7 we study its performance in a modest simulation study and demonstrate its use in a small application, respectively.

2. A Consistent Nonparametric Test

Consider an i.i.d. sequence of random vectors $\xi_1, \ldots, \xi_n \in \mathbb{R}^d$. Our goal is to test for the affiliation of any of this sequence, say ξ.

The definition of affiliation we use is equivalent to that of Milgrom and Weber (1982), lemma 1. Let $a \vee b$ denote the element–wise maximum of a, b and $a \wedge b$ the element–wise minimum, and let $\mathscr{B}(a, \delta)$ be a cube with volume δ and centroid a.

Definition 1. *The elements of a random vector $\xi \in \mathbb{R}^d$ with joint distribution function F are affiliated if for any two vectors $a, b \in \mathbb{R}^d$ and any $\delta > 0$, $Q(a, b; \delta) = p_\delta(a) p_\delta(b) - p_\delta(a \vee b) p_\delta(a \wedge b) \leq 0$, where $p_\delta(a) = P[\xi \in \mathscr{B}(a, \delta)]$.*

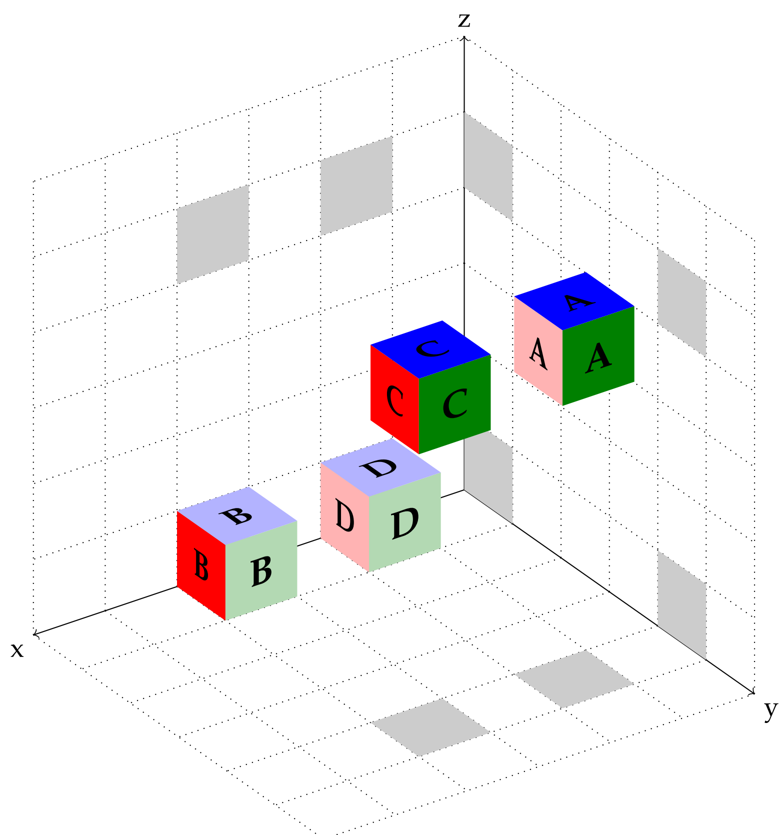

FIGURE 1. Affiliation

The definition is illustrated in figure 1 for $d = 3$. Boxes A,B,C,D correspond to $\mathscr{B}(a, \delta)$, $\mathscr{B}(b, \delta)$, $\mathscr{B}(a \vee b, \delta)$, and $\mathscr{B}(a \wedge b, \delta)$, respectively. Boxes A and B are arbitrary cubes of the same dimensions. Boxes C and D have the same dimensions as A and B. The centroid of box C (D) corresponds to the element–wise maximum (minimum) of the centroids of A and B. Q is the product of the probabilities that ξ takes a value in A, B, respectively, minus the corresponding product of C, D.

Under affiliation Q is nonpositive and under independence, $Q = 0$ for all a, b, δ. An equivalent definition for continuous F is that $f(a \wedge b)f(a \vee b) \geq f(a)f(b)$ for all a, b, where f is the density function corresponding to F.

Our test statistic is based on the quantity

$$T(Q) = \int m\big\{Q(a, b, \delta)\big\}w(a, b, \delta)dadbd\delta = \int mdW, \tag{1}$$

where $m(Q) = \max(Q, 0)$, w is some continuous nonnegative function such that for any a, b in the support of F, $\lim_{\delta \downarrow 0} w(a, b, \delta) = w^*(a, b) > 0$. By definition $T(Q)$ is positive whenever Q is positive on a set of nonzero measure.[4]

We achieve a pivotal limiting distribution under the null hypothesis by adjusting the 'kink' in m by the introduction of a sample–size–dependent function T_n. Dropping the (a, b, δ)–argument, let

$$T_n(Q) = \int_{Q > -\beta_n} QdW, \tag{2}$$

with $\beta_n \prec 1$ an input parameter, where '$\prec$' indicates that the quantity on the left converges faster than the quantity on the right.

We can then estimate the function Q by its sample analog, $\hat{Q}$, given by

$$\hat{Q}(a, b, \delta) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} h^*_{ij},$$

with $h^*_{ij} = I\big(\xi_i \in \mathscr{B}(a, \delta)\big)I\big(\xi_j \in \mathscr{B}(b, \delta)\big) - I\big(\xi_i \in \mathscr{B}(a \vee b, \delta)\big)I\big(\xi_j \in \mathscr{B}(a \wedge b, \delta)\big)$, where $I(A)$ is the standard indicator function that takes the value one when A is true and zero otherwise. Our test statistic then has the form

$$\hat{\tau}_n = \sqrt{n}\frac{T_n(\hat{Q})}{2\hat{v}}, \tag{3}$$

with $\hat{v} = \sqrt{\hat{v}^2}$, where $\hat{v}^2$ estimates some positive v^2. The idea is that as the sample size increases, $\hat{Q}$ converges to Q and T_n converges to T, so $T_n(\hat{Q})$ converges to $T(Q)$, which is zero under affiliation and positive absent affiliation. Since our (yet to be defined) $\hat{v}$ is bounded, the $\sqrt{n}$–norming ensures the consistency of $\hat{\tau}_n$.

So for consistency all that is required is that $\hat{v}$ is positive and bounded, but its choice is motivated by the test statistic properties under the null. Let $h_{ij} = \int_{Q=0}(h^*_{ij} + h^*_{ji})dW/2$ and $v^2 = V\left[E[h_{12}|\xi_1]\right] =$

[4]It would be possible to develop a Kolmogorov–Smirnov–style test instead of using our Cramér–von Mises approach. We have not done so because of the greater computational difficulty with Kolmogorov–Smirnov type tests.

$E[h_{12}h_{13}] - (Eh_{12})^2$. Further, let $h_{nij} = \int_{\hat{Q}>-\beta_n}(h^*_{ij} + h^*_{ji})dW/2$, and

$$\hat{\upsilon}^2 = \big(n(n-1)(n-2)\big)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{t\neq i,j} h_{nij}h_{nit} - \Big(\big(n(n-1)\big)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h_{nij}\Big)^2. \tag{4}$$

In large samples it is theoretically better to replace the condition $\hat{Q} > -\beta_n$ in the definition of $\hat{\upsilon}^2$ with $|\hat{Q}| < \beta_n$, because such a change does not affect the limit of $\hat{\upsilon}^2$ under the null, but converges to a smaller value under the alternative, thereby increasing power. However, (unreported) simulation results suggest that this approach requires a substantially larger choice of β_n to ensure an acceptable size in finite samples, which often offsets the potential power gain from a smaller denominator.

We are now in a position to state our consistency theorem.

Theorem I (Consistency). *For any $C < \infty$, under the alternative hypothesis ($T(Q) > 0$),* $\lim_{n\to\infty} P[\hat{\tau}_n > C] = 1$.

The consistency theorem is simple and essentially requires no assumptions other than i.i.d. data and the restriction that β_n be decreasing to zero. Asymptotic validity is trickier. The main problem is that if the elements of ξ are *strictly affiliated* (i.e. $Q < 0$ on a set of positive measure), then any test will necessarily be conservative in that the asymptotic rejection probabilities are less than the intended significance levels. A consequence of this is that if on large parts of the support ξ is *strictly affiliated* and affiliation is violated only in a small area, our test is likely to have little power in samples of moderate size. We make the following assumption.

Assumption A. *For $\xi \sim F$, either (i) the elements of ξ are independent (set $\gamma = 0$), or (ii) for some $0 < \gamma < \infty$ and some $0 < c_\gamma < 1$, the function $\psi(t) = \int_{-t<Q<-c_\gamma t} dW$ satisfies $\limsup_{t\downarrow 0}\big(t^\gamma/\psi(t)\big) < \infty$.*

Part (ii) of assumption A is nonprimitive and conditions under which it is satisfied are discussed in section 3. As we will argue in section 3, condition (ii) is satisfied when ξ contains at least one continuously distributed element. Typically γ can be taken to equal $1/2$.

The conditions on the choice of the β_n–sequence that we make in theorem II below are more restrictive for larger values of γ. This may be counter–intuitive since independence is the knife–edge case and requires the weakest condition on the choice of β_n. However, since one can create examples of scenarios in which $\hat{\upsilon}$ is poorly behaved under strict affiliation that are difficult to exclude with regularity conditions, our proofs guarantee that $\hat{\tau}_n$ is negative with probability approaching one under strict affiliation.

Let C_α be the $1 - \alpha$ quantile of the standard normal distribution.

Theorem II (Asymptotic Validity). *Under the null hypothesis ($T(Q) = 0$), if β_n is chosen such that $n^{-1/(2+2\gamma)} \prec \beta_n \prec 1$ and assumption A holds, then $\limsup_{n\to\infty} P\left[\hat{\tau}_n > C_\alpha\right] \leq \alpha$. The weak inequality becomes an equality in the case of independence.*

Note that in many instances our test is conservative in the sense that it rejects less often than the significance level would suggest, even asymptotically. This is natural if one considers the case of a t–test for the mean of an i.i.d. sequence, where the null hypothesis is that the population mean is either zero or negative. If the population mean in fact is negative, then such a t–test will reject less than 5% of the time at a 5% level of significance. Indeed, in the limit the probability of rejection is then zero. Our situation is similar, albeit that now independence is the knife–edge case.

3. Assumption A

In this section we investigate conditions under which part (ii) of assumption A is satisfied. The discussion below presumes throughout that the null of affiliation is satisfied, but that the vector ξ does not entirely consist of mutually independent variates. We first state and then explain a sufficient condition for assumption A(ii).

Assumption B. *There exists a set $S \in \mathbb{R}^d \times \mathbb{R}^d$ of positive measure and numbers $0 < \rho, \bar{\delta}, c_\rho, C_\rho < \infty$ such that*

$$\forall (a,b) \in S,\, 0 < \delta < \bar{\delta} : c_\rho \delta^\rho \leq -Q(a,b,\delta) \leq C_\rho \delta^\rho.$$

Theorem III. *If assumption B is satisfied then so is part (ii) of assumption A for $\gamma = 1/\rho$.*

To illustrate assumption B, first suppose that ξ has a continuous density f at $a^*, b^*, a^* \vee b^*, a^* \wedge b^*$ and $\lambda(a^*,b^*) = f(a^* \vee b^*)f(a^* \wedge b^*) - f(a^*)f(b^*) > 0$. Then for some $0 < \delta^*, \underline{\lambda}, \bar{\lambda} < \infty$, $\mathscr{B}(a^*, 2^d\delta^*), \ldots, \mathscr{B}(a^* \wedge b^*, 2^d\delta^*)$ do not intersect and for all $a \in \mathscr{B}(a^*,\delta^*), b \in \mathscr{B}(b^*,\delta^*)$ we have $\underline{\lambda} \leq \lambda(a,b) \leq \bar{\lambda}$. Then by construction,

$$\forall \delta \in (0,\delta^*), (a,b) \in S : \underline{\lambda}\delta^2 \leq -Q(a,b,\delta) \leq \bar{\lambda}\delta^2, \tag{5}$$

and assumption B is satisfied with $\rho = 2$.

So assumption B is necessarily satisfied if the elements of ξ_i are continuous. In fact, as we discuss in the next section, by extension assumption B is satisfied when at least one of the elements of ξ_i is continuous. Our test can be made to work when all elements of ξ_i are discrete, but then alternatives like de Castro and Paarsch (2008) specifically designed for discrete data are preferable. With at least one continuous variable, the relevance of assumption B is in its effect of its restrictions on ρ on the allowable choices of β_n.

4. Discrete Variables

The results derived in this paper do not make assumptions about the types of distributions the elements of ξ_i possess. However, if the vector ξ_i contains discrete as well as continuous variables, variable size boxes (δ) are not needed in the discrete dimensions. In this case one can put marginal point masses at each of the support points of the discrete variables and take δ to be the volume of a d_c–dimensional box, where d_c is the number of continuous variables. For instance, if one wishes to test the affiliation of a continuous and a discrete variable, one can use boxes of the form $\mathscr{B}\big((a_1,a_2),\delta\big) = \{(\tilde{a}_1,\tilde{a}_2) : |\tilde{a}_1 - a_1| \le \delta,\ \tilde{a}_2 = a_2\}$, where a_2 is a probability mass point of the discrete variable. A result similar to (5) then obtains for the redefined δ.

5. Choice of β_n

Above we have derived conditions on the convergence rate of β_n under which asymptotically correct size is guaranteed. These rate conditions do not, however, provide guidance on how to choose β_n in practice. This is a common, though not particularly desirable, characteristic of most nonparametric tests.

Since the unrestricted integral[5] of $\hat{Q}$ under independence is a standard U statistic to which a central limit theorem can be applied, $\beta_n = \infty$ provides optimal size. Since maximizing the power requires that large negative values of $\hat{Q}$ be truncated, $\beta_n = 0$ provides optimal power but the test statistic will then no longer have a limiting normal distribution. This trade–off makes the choice of β_n complicated. The purpose of this section is to provide a heuristic method for choosing the smallest value of β_n for which the size distortion is acceptable.

Because independence is the knife–edge case, our argument is motivated using the case of independence.[6] In lemma A4 we show that $\sqrt{n}(\hat{Q} - Q)$ converges weakly to a Gaussian process $\mathbb{G}$ with covariance kernel $\mathscr{K}(a,b,\delta,a^*,b^*,\delta^*) = 4\,\mathrm{Cov}[h_{12}(a,b,\delta), h_{13}(a^*,b^*,\delta^*)]$. A measure of what fraction of the distribution for given value of β_n is truncated is

$$\mathbb{C} = \frac{\int |\mathbb{G}| I(\mathbb{G} < -\sqrt{n}\beta_n) dW}{\int |\mathbb{G}| dW}. \tag{6}$$

[5] I.e. without requiring that $\hat{Q} > -\beta_n$.

[6] Although the conditions on β_n in the case of strict affiliation are stronger than those in the case of independence, in the bulk of cases of empirical interest rejection frequencies will be lower in the case of strict affiliation than in the case of independence for any value of β_n. Since our procedure is heuristic in any event, we thus focus our attention on the case of independence.

The motivation for (6) is that under independence, $\hat{Q} = \hat{Q} - Q$, and in large samples $\sqrt{n}(\hat{Q} - Q)$ behaves like $\mathbb{G}$. Hence $\hat{Q} < -\beta_n$ corresponds to approximately $\mathbb{G} < -\sqrt{n}\beta_n$; this is conservative under strict affiliation since then $Q < 0$ in some areas.

The larger is $\mathbb{C}$, the higher is the fraction of $\mathbb{G}$ that is truncated, which leads to greater power for a given value of β_n. Our idea is to choose β_n such that $\mathbb{C}$ takes the largest tolerable value in terms of its effect on the size. There are two issues: $\mathbb{C}$ is random, and the impact of $\mathbb{C}$ on the actual size is unclear. However, assuming that our test statistic is indeed standard normal but that it is contaminated by a factor of $1 - \mathbb{C}$ provides heuristic solutions to these issues. Specifically, treating $1 - \mathbb{C}$ as a draw of the distribution of factors by which the value of a standard normal variable is affected, a simple linearization argument (appendix D) suggests that the true rejection probability after the contamination is approximately equal to

$$\Phi(-C_\alpha) - \phi'(C_\alpha)E\mathbb{C}. \tag{7}$$

For $\alpha = 0.05$, (7) is equivalent to about $0.05 + 0.1E\mathbb{C}$. So if $E\mathbb{C} = 0.1$ and our test statistic is indeed standard normal, then the rejection frequency would be about 0.06 instead of 0.05. Our suggestion, then, is to choose β_n such that $E\mathbb{C}$ is in the order of 0.1.

Now, since $E\mathbb{C}$ is unobserved, it must be estimated. A simple procedure for doing so is in appendix E.

6. Experiments

To investigate the performance of our test statistic we have conducted a number of simulation experiments. In our experiments we generate artificial data $\{\xi_i\}_{i=1}^n$ using several models described further below. In all experiments we used 1,000 replications, the same W–function and the same procedure for choosing β_n.

In the simulations we integrate over $\bar{\delta} = \delta^{1/d}$ instead of δ. This does not affect the limit results but does change the constraints on β_n. Indeed, a change of variable type argument shows that the constraint when integrating over $\bar{\delta}$ instead of δ becomes $n^{-1/(2+1/d)} \prec \beta_n \prec 1$. For $d = 2$ the restriction becomes $n^{-2/5} \prec \beta_n \prec 1$.

We chose W as the distribution that arises when one draws random numbers a_1, a_2, b_1, b_2 from a $U(0, 1]$ distribution and $\bar{\delta}$ from a $U(0, \min\{|a_1 - b_1|, |a_2 - b_2|\})$ distribution. The choice of $\bar{\delta}$ is inspired by a desire to increase power by having nonoverlapping boxes (see figure 1).

Although the choice of W can affect the small sample performance of the statistic, it is of second order importance relative to the choice of β_n, much like the choice of a kernel versus that of a bandwidth in nonparametric kernel regression estimation. We hence focus on the choice of β_n here. For our experiments we used $\beta_n = cn^{-1/3}$ with various values of c.

Our first set of experiments is designed to investigate the behavior of the statistic under the null hypothesis. We consider the following model for such size experiments.

Model 1 Let u_{i1}, u_{i2}, u_{i3} be independent draws from $U(0,1)$. For $\lambda \in [0,1]$,

$$\begin{cases} \xi_{i1} = \lambda u_{i1} + (1-\lambda)u_{i3}, \\ \xi_{i2} = \lambda u_{i2} + (1-\lambda)u_{i3}. \end{cases}$$

Model 1 with $\lambda = 1$ represents the knife–edge case of the least favorable null, i.e. independence. When $\lambda < 1$, ξ_{i1} and ξ_{i2} are strictly affiliated. Since the rejection rates are expected to be close to zero if λ is small, we only studied $\lambda \in \{0.8, 0.9, 1.0\}$ for our experiments. The results are summarized in table 1.

	$\lambda = 1$			$\lambda = 0.9$			$\lambda = 0.8$		
$n =$	300	500	1,000	300	500	1,000	300	500	1,000
$c = 0.01$	0.075	0.075	0.064	0.070	0.062	0.051	0.024	0.014	0.012
$c = 0.02$	0.055	0.058	0.045	0.047	0.047	0.046	0.013	0.012	0.008
$c = 0.03$	0.050	0.054	0.054	0.041	0.045	0.044	0.012	0.012	0.005
$c = 0.05$	0.048	0.052	0.043	0.039	0.040	0.043	0.011	0.012	0.005
$c = 0.10$	0.048	0.052	0.042	0.037	0.040	0.043	0.011	0.012	0.005
$c = 0.20$	0.048	0.052	0.042	0.037	0.040	0.043	0.011	0.012	0.005
$c = 0.50$	0.048	0.052	0.042	0.037	0.040	0.043	0.011	0.012	0.005

TABLE 1. Rejection rates under the null of affiliation using $\beta_n = cn^{-1/3}$

$n = 1,000$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.8$
$c = 0.01$	0.316	0.323	0.326
$c = 0.02$	0.208	0.215	0.217
$c = 0.03$	0.140	0.142	0.148
$c = 0.05$	0.065	0.062	0.069
$c = 0.10$	0.009	0.009	0.010
$c = 0.20$	0.000	0.000	0.000
$c = 0.50$	0.000	0.000	0.000

TABLE 2. The values of $E[\mathbb{C}]$

To illustrate the heuristic method for choosing β_n based on $\mathbb{C}$ we proposed in section 5, we estimated $E\mathbb{C}$ for $n = 1,000$; the results are in table 2. Recall that we want $E\mathbb{C}$ to be around 0.1.

The rejection rates under independence are all close to the nominal 5% level; in fact, they are all within the range that the heuristic method using $E\mathbb{C}$ predicts. Under strict affiliation our test is conservative as it is a one–sided test of independence; indeed, the test rejects in considerably less than 5% of replications in our experiments. Note also that the rejection rates are larger for smaller choices of β_n, which is expected in view of the size–power trade–off discussion in section 5.

To investigate the power of the test we consider the following models.

Model 2 Let u_{i1} and u_{i2} be independent draws from $U(0,1)$. Then,

$$\begin{cases} \xi_{i1} = (u_{i1} + u_{i2})/2, \\ \xi_{i2} = (u_{i1} - u_{i2} + 1)/2. \end{cases}$$

Model 3 Let u_{i1}, u_{i2} be mean zero unit variance joint normal random variates with correlation ρ. Then ξ_{i1}, ξ_{i2} are given by

$$\begin{cases} \Phi(u_{i1})/2,\ \Phi(u_{i2})/2, & \text{w.p. } 0.09, \\ 3\Phi(-u_{i1})/10,\ 7\Phi(u_{i2})/10 + 3/10, & \text{w.p. } 0.21, \\ 7\Phi(-u_{i1})/10 + 3/7,\ 3\Phi(u_{i2})/10 & \text{w.p. } 0.21, \\ 7\Phi(-u_{i1})/10 + 3/7,\ 7\Phi(u_{i2})/10 + 3/7 & \text{w.p. } 0.49, \end{cases}$$

where Φ is the standard normal distribution function.

Model 2 is a simple example where ξ_{i1} and ξ_{i2} are uncorrelated, but not affiliated. Model 3 is designed in such a way that ξ_{i1} and ξ_{i2} are positively correlated, but not affiliated. Specifically, ρ equal to $-0.6, -0.8$ corresponds to $\text{Corr}(\xi_{i1}, \xi_{i2})$ equal to 0.109 and 0.213, respectively.[7] Tables 3–4 summarize our results.

	$n = 300$	$n = 500$	$n = 1,000$	$E\mathbb{C}$ ($n = 1,000$)
$c = 0.01$	1.000	1.000	1.000	0.300
$c = 0.02$	0.922	0.997	1.000	0.185
$c = 0.03$	0.554	0.917	1.000	0.118
$c = 0.05$	0.237	0.424	0.912	0.046
$c = 0.10$	0.124	0.184	0.370	0.005
$c = 0.20$	0.106	0.130	0.186	0.000
$c = 0.50$	0.093	0.120	0.162	0.000

TABLE 3. Rejection rates under Model 2 using $\beta_n = cn^{-1/3}$

[7]These are the sample correlation coefficients based on 10,000 random draws.

	$\rho = -0.6$ Corr$(\xi_{i1}, \xi_{i2}) = 0.109$				$\rho = -0.8$ Corr$(\xi_{i1}, \xi_{i2}) = 0.213$			
	$n = 300$	$n = 500$	$n = 1,000$	$E\mathbb{C}$	$n = 300$	$n = 500$	$n = 1,000$	$E\mathbb{C}$
$c = 0.01$	0.911	1.000	1.000	0.323	0.998	1.000	1.000	0.335
$c = 0.02$	0.494	0.859	1.000	0.212	0.869	0.995	1.000	0.223
$c = 0.03$	0.172	0.453	0.953	0.139	0.447	0.888	1.000	0.151
$c = 0.05$	0.007	0.043	0.311	0.063	0.019	0.155	0.808	0.071
$c = 0.10$	0.001	0.000	0.001	0.011	0.000	0.000	0.001	0.014
$c = 0.20$	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.002
$c = 0.50$	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Note: Corr(ξ_{i1}, ξ_{i2}) is computed by using 10,000 random draws.								

TABLE 4. Rejection rates under Model 3 using $\beta_n = cn^{-1/3}$

Table 3 illustrates that the power of our test may be hurt substantially when the chosen β_n is too large. We note here that the test shows reasonable power when the rule–of–thumb $E\mathbb{C} \approx 0.1$ is used for choosing β_n.

Recall that strict affiliation implies positive correlation but that the converse is not true. Table 4 shows that our test indeed has nontrivial power even in the presence of positive correlation. One would normally expect that a violation of affiliation becomes more difficult to detect as the correlation increases. This intuition is however not always correct as table 4 shows. In fact, model 3 is designed in such a way that finding a, b and δ for which $Q(a, b, \delta) > 0$ becomes easier as ρ becomes more negative (and hence the correlation between ξ_{i1} and ξ_{i2} becomes larger positive).

Although the experiments reported in this section are not exhaustive, they are supportive of our theoretical results. We also conducted experiments with $d = 3$, but we did not tabulate the results here since they carry largely the same message.

7. APPLICATION

We implemented our test in one data set, namely the OCS data set. Jun, Pinkse, and Wan (2008) contains two further applications to the California Department of Transport and Russian Federal Subsoil Resources Management Agency data sets. In the OCS data set, the object auctioned is the right to drill for oil (and gas). The data are available from the Center for Auctions, Procurements and Competition Policy website.

The OCS data have been used in a number of other papers, including Hendricks and Porter (1988) and Hendricks, Pinkse, and Porter (2003). The tracts were auctioned off using a first price sealed bid mechanism. Data are available on all bids, the identity of the bidders and any ex post revenue the tracts generated for the winner plus some tract–specific data such as its size and location. We

	$\beta_n \times 10^3$	$E\mathbb{C}$	τ_n	p–values
Whole Sample ($n = 841$)	0.424	0.052	−3.286	0.995
	0.318	0.102	−2.731	0.997
	0.212	0.200	−1.571	0.942
Higher Revenue ($n = 421$)	0.667	0.050	−4.298	1.000
	0.534	0.083	−3.774	1.000
	0.267	0.182	−2.009	0.978
Lower Revenue ($n = 420$)	0.668	0.051	−1.889	0.971
	0.534	0.086	−1.440	0.926
	0.267	0.186	−0.240	0.595
Big 12 Only ($n = 298$)	1.048	0.044	−1.912	0.972
	0.749	0.105	−1.263	0.897
	0.449	0.214	−0.172	0.568
Big 21 Only ($n = 551$)	0.610	0.050	−2.154	0.984
	0.488	0.102	−1.734	0.959
	0.244	0.210	−0.605	0.727

Note 1. The number of tracks with two or more actual bids is 841.
Note 2. Higher Revenue is the upper 50% revenue group.

TABLE 5. Testing pairwise affiliation.

used only data for wildcat tracts in the 1954–1970 period. The number of tracts used is 1,168. Twelve large firms, who collectively won the lion's share of auctions, and a large number of fringe firms participated in these auctions; joint bids were allowed. More details can be found in Hendricks, Pinkse, and Porter (2003).

We first used our test to test for affiliation between two randomly chosen (log) bids per auction, both on the raw data and conditional on ex post revenue. We then tested for affiliation between a randomly chosen bid and the number of bidders, again with and without conditioning on other variables. We conducted all our experiments both including and excluding the fringe bidders. The test statistics are summarized in tables 5 and 6. The reported statistics are based on choosing β_n as per our heuristic method. We also tried $\beta_n = 0.02n^{-1/3}$, but doing so did not change most of the results qualitatively.

For the test of affiliation of bids (table 5), conditioning affects the values of the test statistics, but the conclusions are identical across the board. In fact, the results are unambiguous; none of the test statistic values are even positive. For the test of affiliation of bids and potential number of bidders, affiliation can only be rejected when conditioning on zero revenue (table 6).

	$\beta_n \times 10^3$	EC	τ_n	p–values
Whole Sample ($n = 1168$)	0.161	0.051	−2.512	0.994
	0.114	0.108	−1.454	0.927
	0.076	0.192	−0.089	0.536
Group 1 ($n = 171$)	0.721	0.033	−1.566	0.941
	0.540	0.072	−1.351	0.912
	0.360	0.210	0.703	0.241
Group 2 ($n = 171$)	0.540	0.058	−1.126	0.870
	0.360	0.130	−0.164	0.565
	0.270	0.195	0.927	0.177
Group 3 ($n = 171$)	0.721	0.036	−2.496	0.994
	0.540	0.108	−1.384	0.917
	0.360	0.189	−0.480	0.684
Group 4 ($n = 171$)	0.721	0.054	0.084	0.467
	0.540	0.100	0.365	0.358
	0.360	0.181	1.537	0.062
Group 5 ($n = 172$)	0.899	0.058	−0.434	0.668
	0.629	0.112	−0.153	0.561
	0.360	0.223	0.943	0.173
Zero Revenue ($n = 312$)	0.590	0.047	1.481	0.069
	0.369	0.098	2.208	0.014
	0.221	0.199	3.012	0.001

Note 1. Group j is the subgroup whose revenue belongs to $(100 - 20 \cdot j,\ 100 - 20 \cdot (j-1)]$% of those with non–zero revenue.

TABLE 6. Testing affiliation of bids and the number of potential bidders

The fact that bids are affiliated provides support of the MW framework. But the fact that bids are found to be affiliated with the number of bidders, both actual and potential,[8] is troubling since Hendricks and Porter (1988) and Hendricks, Pinkse, and Porter (2003) have found empirical evidence which is consistent with the common value paradigm.[9] Taking the evidence in Hendricks and Porter (1988) and Hendricks, Pinkse, and Porter (2003) as given, assuming that the measure L constructed in Hendricks, Pinkse, and Porter (2003) is an accurate measure of the number of potential bidders, and discounting the possibility that the proposed test has no power in view of the results of section 6, then the failure to reject must be due to the presence of heterogeneity that is unaccounted for or a failure of one of the MW assumptions.

[8]See Hendricks, Pinkse, and Porter (2003) for how the number of potential bidders is generated.

[9]Li (2005) nevertheless use an *affiliated private values* (APV) model for the OCS data, but even with APV bids and number of bidders should not be affiliated; see Pinkse and Tan (2005).

References Cited

ATHEY, S., AND P. HAILE (2005): "Nonparametric approaches to auctions," in *Handbook of Econometrics*, vol. 6.

DE CASTRO, L. (2007): "Affiliation and dependence in auctions," Discussion paper, University of Illinois.

DE CASTRO, L., AND H. PAARSCH (2008): "Using grid distributions to test for affiliation in models of first–price auctions with private values," Discussion paper, University of Melbourne.

DE LA PEÑA, V., AND E. GINÉ (1999): *Decoupling, from dependence to independence*. Springer.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first–price auctions," *Econometrica*, 68, 525–574.

HENDRICKS, K., J. PINKSE, AND R. PORTER (2003): "Empirical implications of equilibrium bidding in first–price, symmetric, common value auctions," *Review of Economic Studies*, 70, 115–145.

HENDRICKS, K., AND R. PORTER (1988): "An empirical study of an auction with asymmetric information," *American Economic Review*, 78, 865–883.

JUN, S. J., J. PINKSE, AND Y. WAN (2008): "A consistent nonparametric test of affiliation in auction models," Discussion paper, The Pennsylvania State University.

——— (2009): "Cube–root–n and faster convergence, Laplace estimators and uniform inference," Discussion paper, The Pennsylvania State University.

KARLIN, S., AND Y. RINOTT (1980): "Classes of orderings of measures and related correlation inequalities. I Multivariate totally positive distributions," *Journal of Multivariate Analysis*, pp. 467–498.

LI, T. (2005): "Econometrics of first–price auctions with entry and binding reservation prices," *Journal of Econometrics*, 126, 173–200.

LI, T., AND B. ZHANG (2008): "Testing for affiliation in first–price auctions using entry behaviors," Discussion paper, Vanderbilt University.

MILGROM, P., AND R. WEBER (1982): "A theory of auctions and competitive bidding," *Econometrica*, 50(5), 1089–1122.

MONTEIRO, P., AND H. MOREIRA (2006): "First price auctions without affiliation," *Economics Letters*, 91, 1–7.

NOLAN, D., AND D. POLLARD (1987): "U–processes: rates of convergences," *Annals of Statistics*, 15(2), 780–799.

——— (1988): "Functional limit theorems for U–processes," *Annals of Probability*, 16(3), 1291–1298.

PINKSE, J., AND G. TAN (2005): "The affiliation effect in first–price auctions," *Econometrica*, pp. 263–277.

RENY, P. J., AND S. ZAMIR (2004): "On the existence of pure strategy monotone equilibria in asymmetric first–price auctions," *Econometrica*, 72, 1105–1126.

ROOSEN, J., AND D. HENNESSY (2004): "Testing for the monotone likelihood ratio assumption," *Journal of Business and Economic Statistics*, 22, 358–366.

SERFLING, R. (1980): *Approximation theorems of mathematical statistics*. Wiley.

APPENDIX A. CONSISTENCY

We first introduce some notation, which will be used in the proofs of the following lemmas. The following definition can also be found in Nolan and Pollard (1987, 1988).

Definition 2 (Nolan and Pollard (1987, 1988)). *(i) Let $N_p(\epsilon, R, \mathscr{F}, \mathscr{E})$ be the $(L_p(R))$ ϵ–covering number of a class of functions $\mathscr{F}$ with an envelope function $\mathscr{E}$; i.e. $N_p(\epsilon, R, \mathscr{F}, \mathscr{E})$ is the smallest cardinality for a subclass $\mathscr{F}^*$ of $\mathscr{F}$ such that $\min_{\mathscr{F}^*} E_R[|\phi - \phi^*|^p] \leq \epsilon^p E_R[\mathscr{E}^p]$ for each $\phi \in \mathscr{F}$, where $E_R[\cdot]$ denotes the integral with respect to the probability measure R.*
(ii) Let $J(\delta, R, \mathscr{F}, \mathscr{E})$ be the $L_2(R)$–covering integral $\int_0^\delta \log N_2(\epsilon, R, \mathscr{F}, \mathscr{E}) d\epsilon$.
(iii) The class $\mathscr{F}$ of functions is called Euclidean for the envelope $\mathscr{E}$ if there exist constants A and $\mathscr{V}$ such that $N_1(\epsilon, R, \mathscr{F}, \mathscr{E}) \leq A\epsilon^{-\mathscr{V}}$ for $0 < \epsilon < 1$ whenever $E_R[\mathscr{E}] < \infty$. The constants A and $\mathscr{V}$ are called the Euclidean constants for $\mathscr{E}$.

Remark 1. If $\mathscr{F}$ is Euclidean, then for each $p > 1$ $N_p(\epsilon, R, \mathscr{F}, \mathscr{E}) \leq A 2^{2p\mathscr{V}} \epsilon^{-p\mathscr{V}}$ for $0 < \epsilon < 1$ whenever $E_R[\mathscr{E}^p] < \infty$. See page 789 of Nolan and Pollard (1987).
Remark 2. The Euclidean constants A and $\mathscr{V}$ depend on R and $\mathscr{E}$ but only through the integral $E_R[\mathscr{E}]$. Therefore, if the Euclidean class $\mathscr{F}$ has a constant envelope function, then the upper bound of the $(L_p(R))$ ϵ–covering number is in fact uniform in the probability measure R.
Remark 3. A Vapnik–Çervonenkis (VÇ) subgraph class of functions is an example of a Euclidean class; see e.g. lemma 19 of Nolan and Pollard (1987). A collection of functions $\{I(C) : C \in \mathscr{C}\}$ is a VÇ subgraph class if and only if $\mathscr{C}$ is a VÇ class of sets. Also, if $\mathscr{C}_1$ and $\mathscr{C}_2$ are VÇ classes of sets, then so is the collection of all sets $C_1 \cap C_2$, $C_1 \cup C_2$, and C_i^c with $C_i \in \mathscr{C}_i$ for $i = 1, 2$.

Lemma A1. *$\mathscr{H}^* = \{h^*(\xi, \tilde{\xi}) = I\big(\xi \in \mathscr{B}(a, \delta)\big) I\big(\tilde{\xi} \in \mathscr{B}(b, \delta)\big) - I\big(\xi \in \mathscr{B}(a \vee b, \delta)\big) I\big(\tilde{\xi} \in \mathscr{B}(a \wedge b, \delta)\big) : (a, b, \delta) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^+\}$ forms a Euclidean class of functions with the envelope function $H(\xi, \tilde{\xi}) = 2$.*

Proof. Note first that

$$\mathscr{F} = \{\phi(\xi,\tilde{\xi}) = I\big(\xi \in \mathscr{B}(a,\delta)\big) I\big(\tilde{\xi} \in \mathscr{B}(b,\delta)\big) : (a,b,\delta) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^+\}$$

forms a Euclidean class of functions with an envelope function $\mathscr{E}(\xi,\tilde{\xi}) = 1$, because $\{\mathscr{B}(a,\delta) \times \mathscr{B}(b,\delta)\}$ is a collections of cells in $\mathbb{R}^{2d}$, which is a VÇ class of sets. To see this, recall that a collection of half–spaces forms a VÇ class of sets and note that a collection of cells can be formed by taking a finite number of intersections of half–spaces. Similarly,

$$\mathscr{F}^* = \{\phi^*(\xi,\tilde{\xi}) = -I\big(\xi \in \mathscr{B}(a \vee b,\delta)\big) I\big(\tilde{\xi} \in \mathscr{B}(a \wedge b,\delta)\big) : (a,b,\delta) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^+\}$$

forms a Euclidean class of functions with an envelope function $\mathscr{E}^*(\xi,\tilde{\xi}) = 1$. Since $\mathscr{H}^* \subset \mathscr{F} + \mathscr{F}^*$, the lemma follows from corollary 17 of Nolan and Pollard (1987). □

Lemma A2. $\mathscr{H} = \{h(\xi,\tilde{\xi}) = \big(h^*(\xi,\tilde{\xi}) + h^*(\tilde{\xi},\xi)\big)/2 : h^*(\cdot,\cdot) \in \mathscr{H}^*\}$ *forms a Euclidean class of symmetric functions with the envelope function* $H(\xi,\tilde{\xi}) = 2$.

Proof. It follows from lemma A1 and corollary 17 of Nolan and Pollard (1987). □

Lemma A3. *For each* ξ, *let* $P\mathscr{H}$ *be the class of functions* $Ph(\xi,\cdot)$ *with* $h(\cdot,\cdot) \in \mathscr{H}$. *Then,* $P\mathscr{H}$ *forms a Euclidean class of functions with the envelope function* $PH = 2$.

Proof. It follows from corollary 21 of Nolan and Pollard (1987), because $\mathscr{H}$ is a uniformly bounded Euclidean class of functions. □

The following lemma uses the fact that $\sqrt{n}(\hat{Q} - Q)$ is a U–process with respect to a function class $\mathscr{H}$. We will follow Nolan and Pollard (1987, 1988) (see also de la Peña and Giné (1999)).

Lemma A4. *(i)* $\sup |\hat{Q} - Q| \prec 1$ *and (ii)* $\sqrt{n}(\hat{Q} - Q) \xrightarrow{w} \mathbb{G}$ *for some Gaussian process* $\mathbb{G}$ *whose covariance kernel is* $\mathscr{K}(a,b,\delta,a^*,b^*,\delta^*) = 4\,\mathrm{Cov}[h_{12}(a,b,\delta), h_{13}(a^*,b^*,\delta^*)]$.

Proof. Let us consider (i) first. Since $\mathscr{H}$ is a Euclidean class with an envelope $H = 2$, there exist some constants K_1 and V_1 such that $N_1\big(\epsilon, R, \mathscr{H}, H\big) \le K_1 \epsilon^{-V_1}$ for $0 < \epsilon \le 1$. Since $N_1\big(\epsilon, R, \mathscr{H}, H\big)$ is non–increasing in ϵ, we in fact have

$$N_1\big(\epsilon, R, \mathscr{H}, H\big) \le K_1\big(\epsilon^{-V_1} + 1\big), \quad \text{for any } \epsilon > 0, \tag{8}$$

for any probability measure R. Since $E_R[H] = 2$ for any probability measure R, inequality (8) is in fact uniform in R:

$$\sup_R N_1\big(\epsilon, R, \mathscr{H}, H\big) \leq K_1\big(\epsilon^{-V_1} + 1\big), \quad \text{for any } \epsilon > 0. \tag{9}$$

Therefore, the three conditions of theorem 7 of Nolan and Pollard (1987) are all satisfied, and the result (i) follows from there.

For part (ii), we use theorem 5 of Nolan and Pollard (1988). Note first that there exist some constants K_2, K_2^*, V_2, and V_2^* such that when $0 < \epsilon \leq 1$,

$$\sup_R N_2\big(\epsilon, R, \mathscr{H}, H\big) \leq K_2\epsilon^{-2V_2} \quad \text{and} \quad \sup_R N_2\big(\epsilon, R, P\mathscr{H}, PH\big) \leq K_2^*\epsilon^{-2V_2^*}$$

from the fact that both $\mathscr{H}$ and $P\mathscr{H}$ form Euclidean classes that have constant envelope functions. It then follows that

$$\sup_R J(1, R, \mathscr{H}, H) \leq \int_0^1 \log K_2\epsilon^{-2V_2} d\epsilon < \infty \tag{10}$$

$$\sup_R J(1, R, \mathscr{H}, H)^2 \leq \int_0^1 \Big(\log K_2\epsilon^{-2V_2}\Big)^2 d\epsilon < \infty \tag{11}$$

$$\sup_R J(1, R, P\mathscr{H}, PH)^2 \leq \int_0^1 \Big(\log K_2^*\epsilon^{-2V_2^*}\Big)^2 d\epsilon < \infty. \tag{12}$$

Therefore, the first condition of theorem 5 of Nolan and Pollard (1988) is satisfied. Lastly, note that as $\delta \downarrow 0$,

$$\sup_R J(\delta, R, P\mathscr{H}, PH) \leq \int_0^\delta \log K_2^*\epsilon^{-2V_2^*} d\epsilon = \delta \log K_2^* - 2V_2^*\delta\log\delta \to 0, \tag{13}$$

which implies that the second condition of theorem 5 of Nolan and Pollard (1988) is also satisfied. Therefore, the result (ii) of the lemma follows, where the covariance kernel is

$$4\,\mathrm{Cov}\big[E[h_{12}(a,b,\delta)|\xi_1], E[h_{12}(a^*,b^*,\delta^*)|\xi_1]\big] = 4\,\mathrm{Cov}\big[h_{12}(a,b,\delta), h_{13}(a^*,b^*,\delta^*)\big].$$

□

Lemma A5. $T_n(\hat{Q}) - T(Q) \prec 1$.

Proof.

$$T_n(\hat{Q}) - T(Q) = \int_{\hat{Q}>-\beta_n} \hat{Q}dW - \int_{Q>0} QdW = \int_{\hat{Q}>-\beta_n} (\hat{Q}-Q)dW + \int_{\hat{Q}>-\beta_n, Q\leq -2\beta_n} QdW$$
$$+ \int_{\hat{Q}>-\beta_n, 0>Q>-2\beta_n} QdW - \int_{\hat{Q}<-\beta_n, Q\geq 0} QdW. \tag{14}$$

RHS1, RHS2 and RHS4 in (14) vanish by lemma A4. Finally, RHS3 is bounded above by zero and below by $-2\beta_n \prec 1$. □

Proof of Theorem I. Since h_{ij} is uniformly bounded, so is $\hat{v}$. Let C_v be an upper bound to $\hat{v}$, let $\mathscr{E}$ be the event that $|T_n(\hat{Q}) - T(Q)| \leq T(Q)/2$, $\mathscr{E}^c$ its complement, and write

$$P\big[\hat{\tau}_n \leq C\big] \leq P\big[\hat{\tau}_n \leq C,\, \mathscr{E}\big] + P\big[\mathscr{E}^c\big]. \tag{15}$$

Since $T(Q) > 0$, RHS2 in (15) vanishes by lemma A5. RHS1 in (15) is

$$P\left[\sqrt{n}\frac{T_n(\hat{Q})}{\hat{v}} \leq C,\, \mathscr{E}\right] = P\left[\sqrt{n}\frac{T_n(\hat{Q}) - T(Q)}{\hat{v}} + \sqrt{n}\frac{T(Q)}{\hat{v}} \leq C,\, \mathscr{E}\right] \leq P\left[\sqrt{n}\frac{T(Q)}{2C_v} \leq C\right] \prec 1.$$

□

APPENDIX B. VALIDITY

Let $R_n = \{(a,b,\delta) : 0 > Q(a,b,\delta) > -\beta_n/2\}$. Let $Z_n = \int_{Q=0}(\hat{Q} - Q)dW$, $N_n = \int_{R_n}(\hat{Q} - Q)dW$, $K_n = \int_{R_n} QdW$ and $I_n = \int_{Q\leq -\beta_n/2, -\beta_n < \hat{Q} \leq 0} \hat{Q}dW$.

Lemma B1. $\limsup_{n\to\infty} P[T_n(\hat{Q}) \neq Z_n + N_n + K_n + I_n] = 0$.

Proof. Note that

$$\begin{aligned} T_n(\hat{Q}) &= \int_{Q>-\beta_n/2, \hat{Q}>-\beta_n} \hat{Q}dW + \int_{Q\leq-\beta_n/2, \hat{Q}>-\beta_n} \hat{Q}dW \\ &= \int_{Q>-\beta_n/2} \hat{Q}dW - \int_{Q>-\beta_n/2, \hat{Q}\leq-\beta_n} \hat{Q}dW + I_n + \int_{Q\leq-\beta_n/2, \hat{Q}>0} \hat{Q}dW \\ &= Z_n + N_n + K_n + I_n - \int_{Q>-\beta_n/2, \hat{Q}\leq-\beta_n} \hat{Q}dW + \int_{Q\leq-\beta_n/2, \hat{Q}>0} \hat{Q}dW. \end{aligned} \tag{16}$$

The last two terms are zero with probability approaching one by lemma A4. □

B.1. Under condition (i) of assumption A.

Lemma B2. $\sqrt{n}T_n(\hat{Q}) \xrightarrow{d} N(0, 4v^2)$.

Proof. Since $N_n = K_n = I_n = 0$, it suffices to show that $\sqrt{n}Z_n \xrightarrow{d} N(0, 4v^2)$. Now, $\sqrt{n}Z_n = n^{-3/2}\sum_{i=1}^n \sum_{j\neq i} h_{ij}$, which is a standard nondegenerate U–statistic. Apply standard U–statistic theory (e.g. theorem A in section 5.5 of Serfling (1980)). □

Lemma B3. $\limsup_{n\to\infty} P\big[\max_{i\neq j} |h_{nij} - h_{ij}| > 0\big] = 0$.

Proof. Noting that $h_{nij} - h_{ij} = \int_{\hat{Q} \le -\beta_n, Q=0} (h^*_{ij} + h^*_{ji}) dW/2$, it follows that

$$P\big[\max_{i \neq j} |h_{nij} - h_{ij}| > 0\big] \le n^2 P\big[|h_{n12} - h_{12}| > 0\big] \le n^2 P\big[\sup |\hat{Q} - Q| > \beta_n\big] \prec 1,$$

by lemma A4. □

Lemma B4. $\hat{v}^2 - v^2 \prec 1$.

Proof. Let $\tilde{v}^2 = \big(n(n-1)(n-2)\big)^{-1} \sum_{i=1}^n \sum_{j \neq i} \sum_{t \neq i,j} h_{ij} h_{it}$. Then $\hat{v}^2 - \tilde{v}^2 \prec 1$ by lemma B3. Since $\tilde{v}^2$ is an asymmetric U–statistic, one can apply theorem A in section 5.5 of Serfling (1980). □

Lemma B5. $\hat{\tau}_n \xrightarrow{d} N(0,1)$.

Proof. Combine lemmas B2 and B4 and apply Cramér's theorem (**?**, p.355) to obtain the stated result. □

B.2. **Under condition (ii) of assumption A.**

Lemma B6. $|K_n| \succeq \beta_n^{1+\gamma}$.

Proof. By assumption A(ii),

$$|K_n| = -\int_{0 > Q > -\beta_n/2} Q dW \ge -\int_{-\beta_n/2 < Q < -c_\gamma \beta_n/2} Q dW \ge c_\gamma \beta_n \psi(\beta_n/2)/2 \succeq \beta_n^{1+\gamma}.$$ □

Lemma B7. $\limsup_{n \to \infty} P[T_n(\hat{Q}) > 0] = 0$.

Proof. By lemma A4, $Z_n + N_n \preceq n^{-1/2}$ and by lemma B6, $K_n \succeq \beta_n^{1+\gamma} \succ n^{-1/2}$, where K_n is negative. Since I_n is also negative, the stated result follows from lemma B1. □

Proof of Theorem II. In case of condition (i) of assumption A, apply lemma B5, otherwise apply lemma B7. □

Appendix C. Assumption A

Proof of Theorem III. Note that for $c_\gamma = c_\rho/(2^\rho C_\rho)$ and any $(a,b) \in S$, if $t_\rho = (t/C_\rho)^{1/\rho}$, then for sufficiently small $t > 0$,

$$\frac{t_\rho}{2} < \delta < t_\rho \Rightarrow \frac{t}{2^\rho C_\rho} < \delta^\rho < \frac{t}{C_\rho} \Rightarrow \frac{c_\gamma t}{c_\rho} < \delta^\rho < \frac{t}{C_\rho} \Rightarrow C_\rho \delta^\rho < t,\ c_\rho \delta^\rho > c_\gamma t \Rightarrow c_\gamma t < -Q < t.$$

So $\psi(t) = \int_{c_\gamma t < -Q < t} dW \ge \int_{(a,b) \in S;\, t_\rho/2 < \delta < t_\rho} dW \succeq t^{1/\rho} = t^\gamma$.

Appendix D. Linearization

Let $\mathscr{T} \sim N(0,1)$ and $\mathscr{Y}$ be arbitrarily distributed on $[0,1]$ and be independent of $\mathscr{T}$. Then

$$P[\mathscr{T} > C_\alpha \mathscr{Y}] = \int P[\mathscr{T} > C_\alpha y] dF_{\mathscr{Y}}(y) = \int \Phi(-C_\alpha y) dF_{\mathscr{Y}}(y)$$
$$\approx \Phi(-C_\alpha) - C_\alpha \phi(-C\alpha) \int (y-1) dF_{\mathscr{Y}}(y) = \Phi(-C_\alpha) + \phi'(C_\alpha)\big(E\mathscr{Y} - 1\big).$$

Take $\mathbb{C} = 1 - \mathscr{Y}$ to obtain (7).

Appendix E. Computation of $\hat{E}\mathbb{C}$

From Jun, Pinkse, and Wan (2009) it follows that for any process $\hat{\mathbb{G}}$ with covariance kernel $\hat{\mathscr{K}}$ converging to $\mathscr{K}$, then $\hat{\mathbb{G}}$ converges to $\mathbb{G}$. Use

$$\hat{\mathscr{K}}(a,b,\delta,a^*,b^*,\delta^*) = \frac{4}{n(n-1)^2}\Big(\sum_i \Sigma^*_{ni}(a,b,\delta)\Sigma^*_{ni}(a^*,b^*,\delta^*) - \frac{1}{n}\sum_i \Sigma^*_{ni}(a,b,\delta)\sum_i \Sigma^*_{ni}(a^*,b^*,\delta^*)\Big),$$

where $\Sigma^*_{ni}(a,b,\delta) = \sum_{j\neq i} h_{ij}(a,b,\delta)$.

Now,

(1) Draw random vectors $t_1, \dots, t_{\mathscr{S}}$ from W.
(2) Compute the matrix $\hat{\mathscr{M}} \in \mathbb{R}^{\mathscr{S}\times\mathscr{S}}$ with s_1, s_2 element $\hat{\mathscr{K}}(t_{s_1}, t_{s_2})$.
(3) Let $\hat{\mathscr{D}}$ be such that $\hat{\mathscr{M}} = \hat{\mathscr{D}}\hat{\mathscr{D}}'$ (Cholesky decomposition).
(4) For each replication $r = 1, \dots, \mathscr{R}$, take steps 5–7 below.
(5) Draw an i.i.d. standard normal vector $\zeta_r \in \mathbb{R}^{\mathscr{S}}$.
(6) Compute $\hat{\zeta}^*_r = \hat{\mathscr{D}}\zeta_r$.
(7) Compute $\hat{\mathbb{C}}_r = \sum_{s=1}^{\mathscr{S}} |\zeta_{rs}| I(\zeta_{rs} < -\sqrt{n}\beta_n) / \sum_{s=1}^{\mathscr{S}} |\zeta_{rs}|$.
(8) Then $\hat{E}\mathbb{C} = \mathscr{R}^{-1}\sum_{r=1}^{\mathscr{R}} \hat{\mathbb{C}}_r$.