DECIDING BETWEEN THE COMMON AND PRIVATE VALUES PARADIGM : AN APPLICATION TO EXPERIMENTAL DATA.

ARMANTIER Olivier

State University of New York at STONY BROOK
November 2000

Abstract

In the application of auction theory to the real world, a practitioner must choose ex ante which of the Common Value or the Private Values paradigm applies. Since intuition may fail, some authors have proposed to decide statistically between the two models. These studies however did not confirm whether the tests validated the true model since field data were used. I propose to use experimental data as well as Monte Carlo simulations to study different non structural rules to decide between the two paradigms. I find that regressions are inconclusive while a non-parametric procedure seems to be powerful and robust.

Keyword: Auction, Common Value, Private Values, non-parametric estimation.

JEL Classification Number: C51, C14.

I would like to thank Jean-Francois Richard, John Kagel, Soiliou Namoro, Christopher Swan, Jaideep Roy, Robert Slonim and two anonymous referees for helpful comments and discussions. All remaining errors are mine.

1 Introduction

The application of auction theory to the real world is an exercise to be undertaken with caution. To suit the situation under study one must make realistic assumptions, but also decide between the private values (PV) or common value (CV) paradigms. The choice between assumptions is a well-known problem common to most game theoretical applications. Typically, it is assumed in first price auctions that bidders are risk neutral, symmetric and act non-cooperatively. On the other hand, the choice between PV and CV is specific to a given market.[1] The object of the present paper is to develop a robust non-parametric test to decide statistically between the two paradigms.

Theoretical models, optimal behavior and predictions are fundamentally different under the PV and CV paradigms. The selection of the proper paradigm is therefore crucial to make sound theoretical predictions and to derive a correctly specified structural econometric model. One may sometimes find an intuitive argument to select one model over the other. In most cases however, both paradigms are plausible. For instance, consider a procurement where bidders compete to obtain a mineral right. This right may seem to have a single objective value determined by the actual amount of oil beneath the ground. The bidders do not know ex-ante this true value, but, having access to different information sources, they make different estimates. The CV setting seems to fit this model. However, one may assume that the companies participating in the auction have unknown heterogenous costs of extraction. The expected profit, and, consequently, the value of the mineral right is now different for each bidder. In that case, the PV paradigm applies. It can be argued that the CV and PV assumptions are overly restrictive and that a real auction includes both CV and PV elements. Theoretical models explicitly accounting for CV and PV components in first price auctions have been recently proposed (e.g. Wilson (1992), Goeree and Offerman (1999)). Such models however raise serious questions regarding the identification and estimation of the structural parameters. Typically, researchers consider that either the CV or PV component "dominates", and the empirical model is assumed to behave as a "pure" CV or PV model.

Since intuition may fail, some authors have proposed to decide statis-

[1] See Riley and Samuelson (1981), Milgrom and Weber (1982) or Klemperer (1999) for surveys on auction theory.

tically between the competing models in the context of first price auction. Two different approaches have been explored: the first one is based upon structural econometrics and model validation (Paarsch (1992)). The success of this method depends on the correct specification of the structural model. In other words, this approach is inefficient if the econometric (e.g. the distribution of the private signals) or the theoretical model (e.g. the risk neutrality or symmetry assumption) is not correctly specified. Moreover, specification tests are based upon asymptotic results and are known to be of low power here since auction samples are typically small and bids may no be independently distributed. Finally, Paarsch compares the CV model to the Independent Private Values (IPV) model. However, the natural alternative to a CV model is the Affiliated Private Values (APV) model, since they are informationally equivalent to the players. Indeed, the distribution of the private signals is perfectly known to the players in the IPV model. On the other hand, the distribution of types under the APV and CV models is typically defined up to an unobserved affiliation factor. The information available to individual bidders about their opponents is therefore the same under the CV and APV models. Additionally, the APV includes as a special case the IPV model.

The second approach (Giliberto and Varaiya (1989) and Brannman et al. (1987), GVB hereafter) consists of running regressions in order to determine and compare the empirical and theoretical relationship between bids (or winning bids) and the the number of bidders. It is unclear however that the correct relationship between bids and the number of bidders can be found by running regressions. Indeed, this relationship is determined through the equilibrium bid function which may be highly non-linear. GVB do not formally derive the bid functions nor do they observe private signals which is a critical explanatory variable. Their regression models are therefore misspecified and nothing ensures that the correct relationship between bids and the number of bidders will be found.

The method proposed in the present paper combines a non-parametric estimation of the bid function and a formal statistical test in order to identify whether the relationship between expected bids and the number of bidders is consistent with the CV or PV paradigm. The improvement upon the previous literature is threefold. Firstly, the approach is fully non-parametric. It does not require to specify the theoretical model or to estimate the private signal's distribution. Secondly, the method entails resampling independent data from a smooth non-parametric estimation of the bids distribution. This approach provides accurate estimates and the correct relationship between bids and the number of bidders is more likely to be identified. Finally, the robustness of the method is tested on Monte Carlo simulations and experimental data. The object is to analyze how decision rules perform when players systematically use the equilibrium strategy (Monte Carlo simulations) and when data are noisy (e.g. experimental subjects behave slightly out of equilibrium) which is common with field studies.

The paper is structured as follows. Section 2, discusses theoretical predictions. The experiment from Kagel and Levin (1986) is introduced in section 3. Section 4 analyzes the method of GVB. In section 5, a non-parametric estimation method and a statistical test are proposed to distinguish between the CV and PV models. Section 6 concludes.

2 Theoretic Predictions

The nature of the relationship between bids and the number of bidders in PV and CV first price auctions is widely accepted in the literature and it has been confirmed in many applications.[2] These general predictions however are not always unconditionally supported by auction theory, although the theoretical limitation appears to be in practice an exception more than a rule. Let us try here to clarify what theory predicts about the relationship between individual bids or expected winning bids and the the number of bidders. The theoretical models in the present paper are built upon the assumptions that players are risk neutral, symmetric and act non-cooperatively. Each auction is treated as a one shot game and repeated auctions are assumed to be independent and without synergy. The the number of bidders is assumed to be known ex-ante and exogenously determined. This framework allows for unobserved heterogeneity across auctions as in the CV or APV model as long as the affiliation factor is independent of the the number of bidders. The assumptions also rule out cases such as endogenous entry in auctions where more valuable objects attract more bidders.

[2] See Kagel (1995) for experimental evidence or laffont (1997) for empirical evidence.

2.1 Theoretic Prediction about the Individual Bid

In the IPV model, an increase in the the number of bidders (n), caeteris paribus, increases optimal individual bids, as it is more likely that another bidder's estimate is close to one's own. Matthews (1987) extended the result to APV models. This generalization has been validated in several applications.[3]

In a CV model there should be a negative relationship between the individual bids and the the number of bidders, since the highest signal is more likely to be on the upper tail of the distribution as n increases. To avoid the "Winner's Curse" bidders must therefore discount their bids further.[4] It has been shown, however, that the bid function in the CV model may first increase, reflecting more competition, but then decrease in n. This result appears to crucially depend upon the accuracy of the signals and the degree of uncertainty regarding the range of the true value. Giliberto and Varaiya's test is valid when the signals are drawn from a Weibull or a Log-normal distribution but may become invalid with a Normal distribution and a small number of players.[5] Note, that there is no general theoretical threshold value, above which any CV bid function decreases in n. However, the threshold value is generally believed to be reasonably small in practice. For instance, Wilson (1992) considers a standard Normal distribution and finds the bids to decrease when n is greater than 4.

2.2 Theoretic Prediction about the Expected Winning Bid

The expected winning bid in the first price IPV auction is equal to the expected second highest order statistic and, consequently, is an increasing function of n. When private signals are not independent, the theoretical relationship between the expected winning bid and the number of bidders is expected to increase although it has not been proven formally.

For the CV model, it is generally believed that the expected winning bid decreases with the the number of bidders as it converges almost surely to the true value at the rate of $1/n$ (e.g. Laffont (1997)). Wilson (1992) however

[3] See Wilson (1977) or Kagel et al. (1987).

[4] See Capen et al. (1971) as well as Kagel (1995) for empirical and experimental evidence of the "Winner's Curse".

[5] See Capen et al. (1971) or Wilson (1992).

notes that "there seems to be no general results on the rate of convergence" of the expected winning bid toward the true value as n increases. Therefore, the generality of this result is questionable although it appears to be mostly true in practice. Besides, Milgrom's (1979) theorem shows that for a sequence of first price symmetric auctions with an increasing the number of bidders and the same true value, the expected winning bid converges to the true value if and only if the signal sequence is an extremal-consistent estimator of the true value.[6] An extremal-consistent estimator exists for most of the familiar distributions, such as the Normal or Lognormal distributions. There are however important exceptions such as the Exponential distribution. In such cases, it may not be possible to predict the convergence and the rate of convergence of the expected winning bid towards the true value. As a consequence it seems difficult to draw a general theoretical prediction linking the expected winning bid and the the number of bidders in the CV model. More importantly from an empirical perspective, the theoretical prediction requires observable true values. This issue is of importance in practice since, as discussed in section 4, true values are unobserved and difficult to approximate from field data.

The tests in the present paper are based upon what appears to be the most reasonable prediction: the expected bid is an increasing (decreasing) function of the the number of bidders when the PV (CV) model applies and when the number of participants is "sufficiently" large. The relationship between bids and the number of bidders may be non-monotonic when players are clearly out of equilibrium, the auction is a mixture of PV and CV, or the assumptions of the model are not valid.

3 The Model and the Data

3.1 The Model

Kagel and Levin (1986) run an experiment consisting of a series of first price common value sealed bid auctions. At auction j ($j = 1, ..., m$), n_j ($n_j = 1, ..., N$) players are randomly matched and the true value of the item, x_{0_j}, is generated from a uniform distribution on the interval $[\underline{x}_j, \overline{x}_j]$. A private signal s_{ij} ($i = 1, ..., n_j$), is drawn independently from a uniform distribution

[6] An extremal-consistent estimator is a consistent estimator of the mean of the distribution (the true value) based upon the highest order statistic (the highest signal).

on $[x_{0j} - \varepsilon_j, x_{0j} + \varepsilon_j]$ and is sent to bidder i. The distribution of the private signals, the distribution of the true value, the value of ε_j, and the interval $[\underline{x}_j, \overline{x}_j]$ are common knowledge. The true value x_{0j} and the other bidders private signals are not known to individual bidders at the time they submit their bids. The high bidder earns a profit equal to the true value of the item less the amount he bids. The subjects receive a starting balance and are no longer permitted to bid if their balance becomes negative during the course of the experiment. Therefore, as often is the case with field data, only the "best" players survive.

This setting clearly corresponds to the CV auction model with strict positive affiliation described by Milgrom and Weber (1982). If it is assumed that bidders are symmetric, risk neutral and act non-cooperatively, then under the boundary condition $B^{cv}(\underline{x}_j - \varepsilon_j) = \underline{x}_j - \varepsilon_j$, the unique Bayesian Nash equilibrium is:[7]

$$B^{cv}(s_{ij}) = b_{ij}^{cv} = \begin{cases} \underline{x}_j + \frac{1}{n_j+1}\left(s_{ij} - \underline{x}_j + \varepsilon_j\right) & if\ \underline{x}_j - \varepsilon_j \leq s_{ij} \leq \underline{x}_j + \varepsilon_j\ (region\ 1) \\ s_{ij} - \varepsilon_j + Y & if\ \underline{x}_j + \varepsilon_j \leq s_{ij} \leq \overline{x}_j - \varepsilon_j\ (region\ 2) \\ s_{ij} + \varepsilon_j - G(s_{ij}) & if\ \overline{x}_j - \varepsilon_j \leq s_{ij} \leq \overline{x}_j + \varepsilon_j\ (region\ 3) \end{cases} \tag{1}$$

where

$$Y = \frac{2\varepsilon_j}{n_j + 1} \exp\left\{-\frac{n_j}{2\varepsilon_j}\left(s_{ij} - \underline{x}_j + \varepsilon_j\right)\right\} \quad , \tag{2}$$

and

$$\begin{array}{l} G(s_{ij}) = \left\{\left[\overline{x}_j - B^{cv}(\overline{x}_j - \varepsilon_j)\right] P_{n_j}(0) + 2n_j\varepsilon_j \int_0^{\gamma(s_{ij})} P_{n_j}(u)du\right\} P_{n_j}(\gamma(s_{ij})) \\ \gamma(s_{ij}) = \frac{s_{ij} - (\overline{x}_j - \varepsilon_j)}{2\varepsilon_j} \\ P_{n_j}(u) = \exp\left[\ln(1 - u^{n_j}) + n_j \int \frac{du}{1-u^{n_j}}\right] \quad . \end{array} \tag{3}$$

When bidder i receives a signal s_{ij} in region 2, he can infer that the true value x_{0_j} is in the interval $[s_{ij} - \varepsilon_j, s_{ij} + \varepsilon_j]$. When s_{ij} is in region 1 (region 3) bidder i knows that x_{0_j} belongs to $[\underline{x}_j, s_{ij} + \varepsilon_j]$ ($[s_{ij} - \varepsilon_j, \overline{x}_j]$). Heuristically, the bid function differs across regions because bidders receive different information regarding the possible range of the true value.

Now, suppose that the true underlying model is unknown. One could assume that bidders are engaged in a PV auction with positive affiliation.

[7] See Kagel and Richard (1998) or Armantier (1998) for the derivation of the bid function.

In this case, the experiment would be seen as follow: first, an affiliation factor x_{0j}, representing the unobserved heterogeneity across auctions, is drawn from a uniform distribution on the interval $[\underline{x}_j, \overline{x}_j]$; then, private values s_{ij} $(i = 1, ..., n_j)$ are drawn independently from a uniform distribution on $[x_{0j} - \varepsilon_j, x_{0j} + \varepsilon_j]$. The value of the item for bidder i is represented by his private signal s_{ij}, and no longer by x_{0j} as in the CV model; x_{0j} is now simply the mean of the private signals distribution at auction j. This APV model is a plausible alternative to the CV model described previously. Indeed, it corresponds to the extreme case of a CV model where players see their signals as an exact estimate of the true value, rather than an unbiased estimate of x_{0j}. The corresponding APV Nash equilibrium is:

$$B^{pv}(s_{ij}) = b_{ij}^{pv} = \begin{cases} s_{ij} - \frac{1}{n_j+1}\left(s_{ij} - \underline{x}_j + \varepsilon_j\right) & if \quad \underline{x}_j - \varepsilon_j \leq s_{ij} \leq \underline{x}_j + \varepsilon_j \\ s_{ij} - \frac{2\varepsilon_j}{n_j} + \frac{Y}{n_j} & if \quad \underline{x}_j + \varepsilon_j \leq s_{ij} \leq \overline{x}_j - \varepsilon_j \\ s_{ij} + \varepsilon_j - G(s_{ij}) & if \quad \overline{x}_j - \varepsilon_j \leq s_{ij} \leq \overline{x}_j + \varepsilon_j \\ \end{cases} \tag{4}$$

where

$$Y = \frac{2\varepsilon_j}{n_j + 1} \exp\left\{-\frac{n_j}{2\varepsilon_j}\left(s_{ij} - \underline{x}_j + \varepsilon_j\right)\right\} \tag{5}$$

and

$$\begin{cases} G(s_{ij}) = \left\{\left[\overline{x}_j - B^{pv}(\overline{x}_j - \varepsilon_j)\right] P_{n_j}(0) - n_j\varepsilon_j \int_0^{\gamma(s_{ij})} P_{n_j}(u)du\right\} P_{n_j}(\gamma(s_{ij})) \\ \gamma(s_{ij}) = \frac{s_{ij} - (\overline{x}_j - \varepsilon_j)}{2\varepsilon_j} \\ P_{n_j}(u) = \exp\left[\ln(1 - u^{n_j}) + n_j \int \frac{du}{1-u^{n_j}}\right] \end{cases} . \tag{6}$$

The bid functions in region 2 under the CV and PV models are linear in s_{ij} up to a factor Y that decreases rapidly towards 0 as s_{ij} increases. As illustrated in Kagel and Richard (1998), the CV (and, by analogy, the PV) bid functions in region 3 are essentially linear in s_{ij} despite their analytical complexity. Finally, the bid functions across regions and models are non-linear in n_j, as well as in the other parameters of the model ($\underline{x}_j$, $\overline{x}_j$ and ε_j).

The derivative of the CV (PV) bid functions with respect to n_j are uniformly negative (positive) across the three regions.[8] As suggested in Section

[8] See Armantier (1998) for detailed calculations of the bid functions derivatives.

2, the bids and the number of bidders are positively related for the PV model, and negatively related for the CV model. We can therefore attempt to distinguish the two paradigms in the experimental model based upon the theoretical relationship between bids and the number of bidders.

3.2 The Data

There are two experimental samples: the first one (1003 bids), involves inexperienced bidders who exhibit a strong “Winner’s Curse”, since, on average, they make negative profits. The second (2159 bids) involves experienced bidders who are largely free of the ”Winner’s Curse”. The parameters of the model vary in each treatment: for the inexperienced sample $n_j \in \{3, 4, 5, 6, 7\}$, $\varepsilon_j \in \{12, 18, 24, 30\}$, $\underline{x}_j \in \{15, 25, 30\}$ and $\overline{x}_j \in \{100, 225, 500\}$; for the experienced sample $n_j \in \{4, 6, 7\}$, $\varepsilon_j \in \{18, 30\}$, $\underline{x}_j \in \{25, 30, 50\}$ and $\overline{x}_j \in \{225, 230, 550\}$. For each bid b_{ij} submitted during the experiments we observe s_{ij}, $\underline{x}_j$, $\overline{x}_j$, ε_j, x_{0j} and n_j.

In order to compare actual behavior with that in equilibrium, I compute two benchmark samples b_{ij}^{cv} and b_{ij}^{pv} with the bid functions (1) and (4) and with the same s_{ij}, $\underline{x}_j$, $\overline{x}_j$, ε_j, and n_j used by the subject who submitted b_{ij}. Based on these benchmark samples it is possible to determine how close the bids observed during the experiments are to the CV and PV benchmarks. Table 1 summarizes the regressions results of the experimental individual bids on the CV and PV benchmarks and a constant C. The ”nesting” model is $b_{ij} = \alpha b_{ij}^{cv} + (1-\alpha) b_{ij}^{pv} + C + \omega_{ij}$ where ω_{ij} is a normally distributed error term. If the bids were generated from the CV (PV) model then $\{\alpha = 1, C = 0\}$ ($\{\alpha = 0, C = 0\}$). In each regression α exceeds 0.7 but it is significantly lower than 1. In addition, C is reasonably close to 0 and mostly insignificant. This means that both samples are close to the CV benchmark over the three regions. Moreover, there is almost a perfect fit between bids and the CV benchmark in region 2 and 3 for the experienced sample. Note also that experience drives behavior toward the CV paradigm since $\{\alpha, C\}$ get closer to $\{1, 0\}$ with the experience sample. This finding is explained by Kagel and Levin (1986) who note that the inexperienced subjects face a stronger ”Winner’s Curse” than the experienced subjects. It is typically recognize that players may strive to behave optimally but in practice, the implementation of the strategy selected entails small mistakes. The coefficient of correlation between the CV (PV) bidding error $e_{ij}^{cv} = \left(b_{ij}^{cv} - b_{ij}\right)$ ($e_{ij}^{pv} = \left(b_{ij}^{pv} - b_{ij}\right)$) and the the number of bidders is 0.085 (0.101) in the experienced sample and 0.134

Inexperienced Sample				
	Region 1	Region 2	Region 3	All Regions
α	0.702*	0.801*	0.776*	0.785*
Std	(0.112)	(0.049)	(0.039)	(0.051)
C	0.281*	0.186	0.123	0.175
Std	(0.115)	(0.095)	(0.087)	(0.101)
R^2	0.856	0.921	0.932	0.906
Experienced Sample				
	Region 1	Region 2	Region 3	All Regions
α	0.797*	0.912*	0.873*	0.883*
Std	(0.061)	(0.017)	(0.024)	(0.019)
C	0.115	0.032	0.086	0.061
Std	(0.078)	(0.021)	(0.053)	(0.047)
R^2	0.884	0.948	0.951	0.947

* indicates parameters significant at a 5% level.

Table 1: Regression of the actual individual bids on the CV and PV benchmarks

(0.128) in the inexperienced sample. This indicates that observed behavior is non-optimal but the bias is uncorrelated with n_j. It is therefore still possible to use the theoretical prediction to distinguish PV and CV models.

Typically an econometrician does not know the private signals and is unable to compute the benchmark samples and run this type of comparison. Nevertheless, these regressions indicate that the quality of the samples differs. There are indeed two samples of perfectly rational players (the benchmark samples), a slightly noisy sample (the experienced sample) and a sample affected by significant errors (the inexperienced sample). These noisy samples will test the robustness of the decision rule as long as the noise is uncorrelated with the number of players. In practice, such imperfect data commonly originate from players' mistakes, measurement errors or small unobserved heterogeneity across players and/or auctions.

4 Decision Rule Based on Regressions

GVB run a series of regressions to determine the relationship between n_j and the individual bids. They use different specifications (a logarithmic and a the number of bidders dummy variables specifications) to account for non-linearities. The baseline explanatory variables are the the number of bidders, a proxy for the "objective value" or "quality" of the item, and a proxy for the uncertainty about the true value of the item. They conclude in favor of the CV (PV) model if the coefficient of n_j is negative (positive). Note that the "objective value" and the uncertainty about the true value of the item are difficult to estimate with field data. Variables that account for the quality of the objects offered for sale may be used as a proxy for the "objective value". These variables however, would not account for unobserved characteristics among bidders which constitute the core of the PV paradigm.

Following GVB, let us regress the individual bids on the the number of bidders n_j, x_{0_j} (the objective true value), and ($\underline{x}_j$,$\overline{x}_j$,ε_j) (the uncertainty). The best case scenario is considered here since exact values instead of proxies are used in the regressions. In addition to a simple linear model, let us also consider a logarithmic specification, and a linear specification with a number-of-bidders dummy variables to account for other non-linearities.[9] The results of the regressions are presented in Tables 2 and 3. The coefficient of n_j is positive in every regressions. In addition, the dummy variables when significant tend to follow an increasing pattern (e.g. the experience sample and PV benchmark in Table 3). In other words, the regressions always favor the PV model even when bidders use the CV equilibrium bid function. This result is also inconsistent with the CV experimental samples. The GVB regression method is therefore unable to discriminate between the CV and PV paradigms.

I have run some regressions with different combinations of explanatory variables (e.g. including only the the number of bidders, leaving out the true value, with the private signals, with standardized data) and with different specifications (e.g. quadratic, reciprocal, including different sets of dummy variables). In addition, I have simulated larger samples and considered bid functions with simpler structural form and different distributions (e.g. Weibull, Lognormal, truncated Normal).[10] Even though some of these regressions fit the data well, most of the estimated coefficients of n_j are statistically insignificant and/or their signs are inconsistent with the theoretical

[9]The coefficients of the dummy variables allow to compare the effect of various levels of competition with the case $n = 3$ for the inexperienced sample and $n = 6$ for the experienced sample. In Tables 2 and 3, the variable D_k is defined as follow: $D_k = 1$ when $n = k$.

[10]The results of all these regressions are available upon request.

Data	Benchmark CV			Inexperienced Sample			Benchmark PV		
Specification	Lin	Log	Dummy	Lin	Log	Dummy	Lin	Log	Dummy
coefficient of n_j	0.156*	0.038*	.	0.220	0.044*	.	0.394	0.105*	.
Std error	(0.073)	(0.016)	.	(0.108)	(0.020)	.	(0.197)	(0.038)	.
coefficient of ε_j	-1.195*	-0.207*	-1.173*	-1.001*	-0.174*	-1.000*	-0.586*	-0.078*	-0.582*
Std error	(0.058)	(0.013)	(0.065)	(0.062)	(0.013)	(0.069)	(0.057)	(0.011)	(0.063)
coefficient of $\underline{x}_j$	0.558*	0.177*	0.334*	0.331	0.127	0.312*	0.027	0.039	0.013
Std error	(0.157)	(0.082)	(0.131)	(0.169)	(0.089)	(0.138)	(0.152)	(0.074)	(0.129)
coefficient of $\overline{x}_j$	0.022*	-0.030	0.029*	0.036*	-0.012	0.033*	0.035*	0.050	0.026*
Std error	(0.010)	(0.045)	(0.011)	(0.014)	(0.052)	(0.009)	(0.011)	(0.039)	(0.010)
coefficient of x_{0j}	0.960*	1.084*	0.963*	0.963*	1.060*	0.981*	0.967*	0.982*	0.968*
Std error	(0.012)	(0.017)	(0.012)	(0.012)	(0.019)	(0.013)	(0.014)	(0.015)	(0.011)
D4	.	.	2.012*	.	.	3.492*	.	.	2.635*
Std error	.	.	(0.806)	.	.	(0.946)	.	.	(0.583)
D5	.	.	-0.612	.	.	0.513	.	.	1.569*
Std error	.	.	(0.559)	.	.	(0.902)	.	.	(0.706)
D6	.	.	0.126	.	.	-0.117	.	.	-0.430*
Std error	.	.	(0.219)	.	.	(0.097)	.	.	(0.193)
D7	.	.	-1.264*	.	.	-2.087*	.	.	0.006
Std error	.	.	(0.443)	.	.	(0.651)	.	.	(0.154)
R^2	0.989	0.994	0.991	0.988	0.994	0.986	0.990	0.995	0.991

* indicates parameters significant at a 5% level.

Table 2: Regressions accounting for the "objective value" and the uncertainty about the true value. Inexperience and Benchmark Samples.

Data	Benchmark CV			Experienced Sample			Benchmark PV		
Specification	Lin	Log	Dummy	Lin	Log	Dummy	Lin	Log	Dummy
coefficient of n_j	0.119	0.091*	.	0.387*	0.117	.	0.196*	0.141*	.
Std error	(0.056)	(0.041)	.	(0.063)	(0.055)	.	(0.067)	(0.045)	.
coefficient of ε_j	-1.121*	-0.181*	-1.104*	-1.062*	0.175*	1.060*	-0.529*	-0.067*	-0.500*
Std error	(0.098)	(0.024)	(0.099)	(0.113)	(0.024)	(0.113)	(0.097)	(0.021)	(0.098)
coefficient of $\underline{x}_j$	0.264	0.080	0.453	0.239	-0.041	0.209	-0.228	0.040	0.181
Std error	(0.244)	(0.065)	(0.281)	(0.273)	(0.073)	(0.315)	(0.239)	(0.061)	(0.282)
coefficient of $\overline{x}_j$	-0.014	-0.085*	-0.028	-0.023	-0.021	-0.002	0.027	-0.031	0.002
Std error	(0.019)	(0.030)	(0.026)	(0.021)	(0.036)	(0.029)	(0.013)	(0.028)	(0.026)
coefficient of x_{0j}	0.985*	1.103*	0.991*	0.988*	1.099*	0.995*	0.984*	1.011*	0.978*
Std error	(0.006)	(0.009)	(0.005)	(0.008)	(0.011)	(0.007)	(0.006)	(0.009)	(0.006)
D4	.	.	-0.152	.	.	-0.624*	.	.	-0.656*
Std error	.	.	(0.097)	.	.	(0.247)	.	.	(0.173)
D7	.	.	1.721*	.	.	1.623*	.	.	2.559*
Std error	.	.	(0.643)	.	.	(0.687)	.	.	(0.541)
R^2	0.992	0.996	0.991	0.993	0.995	0.991	0.993	0.996	0.992

* indicates parameters significant at a 5% level.

Table 3: Regressions accounting for the "objective value" and the uncertainty about the true value. Experience and Benchmark Samples.

predictions. In other words, it was not possible to draw a definitive conclusion based upon a regression model.

These results may be explained by the following arguments:

(1) The regression model is misspecified. It fails to account for the exact non-linearity of the bid function, and it does not include relevant explanatory variables such as private signals. When a model is misspecified, the estimated signs and values may not be consistent with the theory.

(2) The coefficients of n_j are significant in most cases, but close to zero, and the corresponding standard deviations are rather large. This suggests that the bids may depend weakly on n_j. In this case, the experimental samples may be too small to identify this relationship. Note however that the samples considered here are unusually large compared to the auction literature. In particular, GVB used samples of size 322 and 572. Besides, regressions on large simulated samples did not provide better results.

5 Non-Parametric Approach

This section introduces a non-parametric estimation of the expected bids and a statistical test based on multivariate linear inequalities to verify whether the expected bid increases or decreases with the the number of bidders.

5.1 Non Parametric Estimation of the Expected Bids

In typical auction models, the equilibrium bid vector $b = (b_1, ..., b_n)$ can be written as

$$b = B(s, n, f(. \mid \theta)) \tag{7}$$

where $B(.)$ is the equilibrium bid function, s a vector of private signals, n the the number of bidders, $f(. \mid \theta)$ the joint probability distribution function (p.d.f) of the private signals and θ a parameter. $B(.)$ is assumed to be continuous and non decreasing in s. In the experimental models $B(.)$ is given by either equation (1) or (4), $\theta = \{\underline{x}, \overline{x}, \varepsilon\}$, and $f(. \mid \theta)$ is given by

$$f(s_1, ..., s_n \mid \theta) = \int_{\underline{x}}^{\overline{x}} f_{x_0}(u \mid \theta) \prod_{i=1}^{n} f_{s_i|x_0}(s_i \mid \theta) du \quad , \tag{8}$$

where $f_{x_0}(. \mid \theta)$ (the p.d.f of the true value) and $f_{s_i|x_0}(. \mid \theta)$ (the conditional p.d.f of each private signal) are uniformly distributed on $[\underline{x}, \overline{x}]$ and

$[x_0 - \varepsilon, x_0 + \varepsilon]$.

The bids are the transformation of a random variable s, by a monotonic function $B(.)$. Consequently b is also a random variable with joint p.d.f $g(. \mid n)$ with support $\left[b_n^{\min}, b_n^{\max}\right]$ defined as

$$g\left(B(s,n,f(. \mid \theta)) \mid n\right) = f(s).\left(\frac{\partial B(s,n,f(. \mid \theta))}{\partial s}\right)^{-1} \quad . \tag{9}$$

The expected bid for a given value of n can be expressed as

$$E\left[b \mid n\right] = \int_{b_n^{\min}}^{b_n^{\max}} ug\left(u \mid n\right) du \quad , \tag{10}$$

or equivalently, after substitution

$$E\left[b \mid n\right] = \int_{\underline{s}}^{\overline{s}} B(s,n,f(. \mid \theta)) f\left(s\right) ds \quad . \tag{11}$$

Under the usual regularity assumption on $B(.)$ (i.e. the derivative of $B(.)$ with respect to n is continuous and bounded by an integrable function) we can differentiate (11) under the integral sign. If $B(.)$ is monotonic in n, then so is the expectation of the bids, since

$$\frac{\partial E\left[b \mid n\right]}{\partial n} = \int_{\underline{s}}^{\overline{s}} \frac{\partial B(s,n,f(. \mid \theta))}{\partial n} f\left(s\right) ds \quad . \tag{12}$$

This result suggests the following method to differentiate the CV and PV models: One can estimate non-parametrically the sequence of expected bids $\widehat{E}\left[b \mid n\right]$, for $n = 2, ..., N$, and verify whether this sequence is increasing or decreasing. Working with non-parametric estimates of the expected bid, rather than the bid function, is a convenient way to aggregate observations without requiring any theoretical structure, additional information, or distributional assumption.

Note that any transformation of the bid function by a real increasing function would still be monotonic in n. In other words, such transformation would preserve the theoretical prediction between bids and the number of bidders. One could for instance concentrate on the median bid whose estimation is less sensitive to outliers. The highest order statistic might also be a good candidate when only winning bids are observed. The selection of the bids transformation should be based upon four criteria: convenience, the reliability of the theoretical prediction, the precision of the estimation, and the quality of the relationship with n (e.g. the bid function might be steeper for high private signals or for low bids close to the reservation price). The expected bid is perfectly suited for the present application but the non-parametric estimation method developed subsequently trivially generalizes to different transformations of the bids.

Let us first divide the original sample of size $T = \sum_{j=1}^{m} n_j$ into sub-samples of size $T_n = \sum_{j=1}^{m} n_j I_{(n_j=n)}$ (where $I_{(n_j=n)}$is the indicator function) regrouping the observations in auctions with exactly n bidders. The empirical mean, $\overline{b}_n = T_n^{-1} \left(\sum_{j=1}^{m} \sum_{i=1}^{n_j} b_{ij} I_{(n_j=n)} \right)$ is the obvious non-parametric estimate of the expected bid. This estimator is perfectly appropriate when T_n is large for all values of n. Typical auction samples however are small and data may not be independently distributed. These characteristics may significantly impair the accuracy of the empirical mean and, as we shall see, the sequence $\overline{b}_n$ is not monotonic in our application. In order to increase and homogenize accuracy across n, I propose to derive $\widehat{E}\left[b \mid n\right]$ from the following non-parametric estimation of the p.d.f $g\left(. \mid n\right)$:

$$\widehat{g}\left(b \mid n\right) = \frac{1}{\gamma} \sum_{j=1}^{m} \sum_{i=1}^{n_j} K_b \left(\frac{b_{ij} - b}{h_b} \right) K_n \left(\frac{n_j - n}{h_n} \right) \quad , \tag{13}$$

where

$$\gamma = \frac{1}{h_b} \sum_{j=1}^{m} \sum_{i=1}^{n_j} K_n \left(\frac{n_j - n}{h_n} \right) \quad , \tag{14}$$

$K_b\left(.\right)$ and $K_n\left(.\right)$ are arbitrary kernel density functions, h_b and h_n are bandwidths controlling the smoothness of the kernel estimates, and $\{b_{ij}, n_j\}$ are observed data points. Heuristically, to estimate accurately $g\left(. \mid .\right)$ at a given point (b, n) we define a two dimensional neighborhood $(b \pm \varepsilon_1, n \pm \varepsilon_2)$, where ε_1 and ε_2 are functions of h_b and h_n respectively; then some "mass" is added to the estimated p.d.f, proportionally to the number of data observed in this neighborhood. If there are numerous (few) observations in the defined neighborhood then the density $\widehat{g}\left(b \mid n\right)$ is high (low) since it appears more (less) likely to observe bids around (b, n). The contribution of data points away from (b,n) is down-weighted by the kernel functions.[11] The introduction of $K_n(.)$ differentiates the present estimation method from other non-parametric approaches (e.g. Vuong et al. 2000). The kernel $K_n(.)$ plays a crucial role in the subsequent application as it provides the smoothness and homogeneity necessary to identify the monotonic relationship between bids and the number of bidders.

Finally, the expected bid can be estimated by

$$\widehat{E}[b \mid n] = \int_{b_n^{\min}}^{b_n^{\max}} u\widehat{g}(u \mid n)\, du \qquad . \tag{15}$$

Equation (15) cannot be solved analytically in most cases. The integral is approximated by the Monte Carlo method where the simulated points are generated from $\widehat{g}(. \mid n)$ over the interval $\left[b_n^{\min}, b_n^{\max}\right]$. Note that $b_n^{\min}$ and $b_n^{\max}$ are typically unknown, but they can be estimated consistently by the two extrema among observed bids: $b_n^{\min} = \underset{(j=1,m;i=1,n_j)}{Max} b_{ij} I_{(n_j=n)}$, $b_n^{\max} = \underset{(j=1,m;i=1,n_j)}{Min} b_{ij} I_{(n_j=n)}$. The non-parametric estimate $\widehat{E}[b \mid n]$ possesses two key characteristics: it involves the entire sample and it relies upon a resampling technique that generates large simulated samples of independent data. The empirical mean $\overline{b}_n$ on the other hand relies in the application on T_n non independent observations.

The choice of kernel functions has been shown to be of limited importance to obtain an accurate estimate, compared to the choice of bandwidth. Therefore, Kernel functions are selected here based on pure convenience. The explanatory variables in non-parametric estimations are usually continuous. Here, n is an integer and one must choose a Kernel function $K_n(.)$ to control easily the neighborhood around n to be included in the smoothing process. The Epanechnikov Kernel is perfectly suited:

$$K_n(u) = \frac{3}{4}(1-u^2)I(\mid u \mid \leq 1) \quad . \tag{16}$$

As for $K_b(.)$, let us consider a Gaussian kernel of the form :

$$K_b(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) \quad . \tag{17}$$

[11] See Hardle (1995) for discussions of non parametric density estimation.

An important aspect of Kernel estimation is the selection of the bandwidths. A very large bandwidth results in an oversmoothed estimate which obscures the detailed structure of the data, while a bandwidth that is too small results in an undersmoothed estimate producing a wavy curve. An optimal bandwidth minimizes the mean square error of the estimate. This minimization requires knowledge of the true p.d.f which is typically unknown. In practice, the chosen bandwidth is only an approximation of the optimal bandwidth. Following Silverman (1986), let us consider the "normal reference bandwidth selector", which, in the case of a one dimensional Gaussian Kernel is given by

$$h_b = 1.06 S_T T^{-\frac{1}{5}} \tag{18}$$

where S_T is the sample standard deviation. This bandwidth is nearly optimal when the data are Gaussian distributed and is considered reasonable in many applications.

The bandwith h_n is fixed at 1.5 in the subsequent application in order to include only $(n+1)$ and $(n-1)$ in the neighborhood of n.[12] One may wonder whether the chosen h_n is too large compared to the optimal bandwidth. This risk is partially reduced by the fact that the optimal bandwidth h_n^* is expected to be large since the influence of n on the bid function is typically small. Note that when h_n becomes smaller than one the effect of $K_n(.)$ vanishes. To preserve the consistency of the estimate, h_n must converge toward 0 as T_n increases. In other words, when the number of observations is large, $\widehat{E}\left[b \mid n\right]$ relies only on the sub-sample of size T_n.

5.2 Statistical Test of the PV and CV Paradigms

To decide between the CV and PV model, one can simply verify whether the sequence of average bids $\widehat{E}\left[b \mid n\right]$ estimated in the previous section is increasing or decreasing in n. In this section I formalize this idea by proposing a test to decide statistically between the paradigms.

Consider the random vector $Y = (y_1, ..., y_n, ..., y_{N-1})$ where $y_n = \widehat{E}\left[b \mid n+1\right] - \widehat{E}\left[b \mid n\right]$. Assume for a moment that Y follows a multinormal distribution with mean μ ($\mu_n = E\left[b \mid n+1\right] - E\left[b \mid n\right]$ $\forall n = 1, ..., N-1$) and a known non-singular variance-covariance matrix Σ. As seen in section 2, μ should be

[12] I have also tried $h_n \geq 1.5$. The results are not significantly different.

positive when the PV model applies. Therefore, we want to test

$$\begin{array}{l} H_0 : \mu_n \leq 0 \text{ for some } n = 1, ..., N-1 \\ H_1 : \mu > 0 \end{array} \tag{19}$$

We conclude in favor of the PV model when the null hypothesis is rejected. When the the number of bidders is reasonably high, we can also test for the CV paradigm by simply changing the signs of the inequalities in (19). Note that the null hypotheses may be accepted in both tests. This would suggest that the bids are out of equilibrium, the auction is a combination of CV and PV and/or the assumptions of the model are not valid. Finally, the statistical test is built such that the null hypotheses under the PV and CV paradigms cannot be rejected simultaneously.

Berger (1989) developed a procedure to test linear inequalities such as (19). For a significance level $\alpha \in \{0.1, 0.05\}$ Berger defines $J = \frac{1}{2\alpha}$ polyhedron shaped rejection regions $R_j = \left\{ y \in \Re^{N-1} : c_j \leq Z_n \leq c_{j-1}, \text{ for all } n = 1, ..., N-1 \right\}$ where $Z_n = y_n / \sqrt{\Sigma_{n,n}}$ is the likelihood ratio test statistic, and c_j $(\forall j = 1, ..., J-1)$ is the percentile $z_{(j*\alpha)}$ of a standard normal distribution, while $c_0 = \infty$ and $c_J = 0$. The null hypothesis H_0 is then rejected if Y belongs to one of these rejection regions $(Y \in R_1 \cup ... \cup R_J)$.

The actual distribution of Y and the covariance matrix Σ are next to impossible to determine since they involve a combination of non-parametric estimators (the estimation of the bids distribution and the estimation of the bids support). The first two moments of Y, however, may be estimated by a bootstrap method with the non parametrically estimated distribution (for details on the bootstrap method see Shao and Tu (1995)). It is well known that the normal distribution maximizes the entropy when only the first two moments of a random variable are known (see Berger (1985)). We can therefore assume if we rely on pure informational criteria that Y follows a Normal distribution. This type of Gaussian approximation is common in the applied statistics literature when the distribution of a random variable cannot be calculated (e.g. Csorgo and Revesz 1981).

5.3 Application to the Experimental Data

A non-parametric method requires a high ratio of data over the number of explanatory variables in order to avoid the "curse of dimensionality". Moreover, a neighborhood has to be defined in each variable dimension. In other words,

n	# b_{ij} observed	h_b	$\overline{b}_n$	$\widehat{E}[b \mid n]$	y_n	Z_n
3	60	15.76	107.41 (23.45)	115.29 (12.62)	0.16 (10.31)	0.0155
4	176	19.63	108.07 (19.63)	115.45 (4.21)	-1.35 (6.83)	-0.1976
5	25	21.06	152.19 (33.02)	114.10 (6.95)	-0.5 (7.18)	-0.0696
6	210	22.34	120.34 (17.26)	113.60 (5.13)	-1.12 (8.65)	-0.1294
7	278	23.01	96.82 (16.74)	112.48 (8.61)	—	—

The numbers in parenthesis refer to the MC standard deviations

Table 4: Non Parametric Estimation of The Average Bid. $\{[\underline{x}, \overline{x}] = [100, 500], \varepsilon = 50, MC = 1000\}$

we must observe data that are reasonably close to each other for each explanatory variable. The only sub-sample one can extract from the experimental samples, satisfying these requirements is $\{[\underline{x}, \overline{x}] = [25, 225], \varepsilon = (12, 18, 24, 30)\}$ from the inexperienced sample. In considering this sub-sample, I implicitly assume that the bids computed for different values of ε's are not significantly different (on average).[13]

The following information is presented in Table 4: the number of data observed in the experimental sample for each n, the value of the bandwidth h_b, the empirical mean of observed bids $\overline{b}_n$, the average bid evaluated with the non-parametric procedure $\widehat{E}[b \mid n]$, the difference between two estimated conditional means $\left(y_n = \widehat{E}[b \mid n+1] - \widehat{E}[b \mid n]\right)$, and the likelihood ratio statistic $Z_n = y_n / \sqrt{\Sigma_{n,n}}$. The standard deviations are calculated by a bootstrap technique where the simulated samples are generated with $\widehat{g}(. \mid n)$.[14]

The average bid $\overline{b}_n$ computed with actual data is neither an increasing nor a decreasing function of n. On the other hand, except for $n = 3$, the estimated mean $\widehat{E}[b \mid n]$ decreases in n, which is consistent with the CV model. The empirical mean is an inaccurate estimator of the expectation because the samples for a given n are small and observed bids are not independent. Indeed, twenty five observations for $n = 5$ correspond to five auctions. These 25 bids are divided in five groups each roughly centered around its own true value x_{0j}. The comparison of the empirical mean is therefore hazardous when

[13]This assumption does not affect the results presented below.

[14]Note that the non parametric estimates of the expected bid are not independent across n in the bootstrap estimation and therefore $Var(y_n) < Var\left(\widehat{E}[b \mid n+1]\right) + Var\left(\widehat{E}[b \mid n]\right)$.

α	J	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
0.05	10	∞	1.64	1.28	1.03	0.84	0.67	0.52	0.38	0.25	0.12	0
0.1	5	∞	1.28	0.84	0.52	0.25	0	—	—	—	—	—

Table 5: Parameters of Berger (1989) Linear Inequalities Test.

the number of observations differs (e.g. $n = 5$ and $n = 6$). Non-parametric methods are known to have slower rates of convergence. The non-parametric estimation of the expected bids for a given n however is more homogenous since cells with few observations (e.g. $n = 5$) benefit from neighboring cells with more observations. This non-parametric method is also more accurate because it relies on large simulated samples of independent bids spread over their support. Note that the non-parametric estimation for $n = 3$ ($n = 7$) should be considered more cautiously since $n = 3$ ($n = 7$) is a bound and the estimation benefits only from $n = 4$ ($n = 6$) its unique neighbor. The estimation for $n = 3$ also relies on a smaller sample and it will not be included in the subsequent test.

The parameters from Berger's (1989) statistical test are presented in Table 5. The rejection region consists then of 10 (5) polyhedral sets defined by the bounds c_j when $\alpha = 0.05$ $(\alpha = 0.1)$. The likelihood ratio statistic Z_n in Table 4 is negative for $n \in \{4, 5, 6\}$. Therefore, the vector Y does not belong to the rejection region for any significance level and the null hypothesis in (19) is always accepted. In other words, the data strongly reject the PV model. To test the CV model let us consider the statistic $-Z$. The likelihood ratio statistic $-Z_n$ ($\forall n \in \{4, 5, 6\}$) belongs to the rejection region $R_5 = [0, 0.25]$ for $\alpha = 0.1$. In other words, we accept the CV paradigm with a 10% risk.

To summarize, from the sole observation of bids and the number of bidders, we recognize that the inexperienced bidders are involved in a CV auction even though they are slightly out of equilibrium since they suffer from the "winner's curse".[15]

[15] As a test, I have run the regressions described in section 4, on the same sub-sample used here. I was not able to conclude in favor of the PV or CV paradigm.

CV Auction Simulation					
n	# of b_{ij} observed	h_b	$\overline{b}_n$	$\widehat{E}[b \mid n]$	$E[b \mid n]$
3	150	33.98 (1.24)	242.52 (14.56)	287.12 (7.23)	286.77
4	152	28.53 (2.18)	269.50 (19.67)	286.05 (4.51)	285.66
5	150	20.92 (1.73)	237.99 (28.74)	284.92 (3.93)	284.71
6	150	16.50 (1.12)	262.36 (13.08)	243.89 (5.29)	284.02
7	154	14.32 (1.36)	277.59 (23.41)	281.28 (8.18)	283.39
PV Auction Simulation					
n	# of b_{ij} observed	h_b	$\overline{b}_n$	$\widehat{E}[b \mid n]$	$E[b \mid n]$
3	150	36.17 (1.94)	293.64 (16.25)	298.34 (8.26)	300.09
4	152	25.83 (2.63)	314.26 (10.65)	307.86 (3.45)	308.37
5	150	22.56 (1.43)	321.43 (19.27)	313.75 (5.21)	313.35
6	150	14.37 (1.02)	314.69 (8.41)	316.32 (4.36)	316.08
7	154	17.68 (1.39)	315.91 (13.56)	321.18 (6.97)	319.05

The numbers in parenthesis refer to the MC standard deviations

Table 6: Monte Carlo Simulations.$\{[\underline{x}, \overline{x}] = [100, 500], \varepsilon = 50, MC = 1000\}$.

5.4 Robustness of the Non-Parametric Method

The fact that the average bid does not change substantially in n, combined with the relatively important standard deviations of the estimates, lead to question the robustness of the non-parametric method. To test this hypothesis, let us conduct a simple Monte-Carlo simulation consisting of 1000 repetitions of the following experiment: for $[\underline{x}, \overline{x}] = [100, 500]$, and $\varepsilon = 50$, we simulate a total of 165 auctions in order to obtain approximately as many bids for $n = 3, 4, 5, 6, 7$; two samples are then simulated, one with the CV bid function (1) and the other with the PV bid function (4); finally, the non-parametric procedure and the statistical test presented in sections 5.1 and 5.2 are applied to these samples. The simulations' outputs are summarized in Table 6.[16]

The results have two key characteristics. Firstly, the empirical mean is much more volatile than the non-parametric estimate. The inaccuracy of the empirical mean is mostly due to the non-independent nature of bids. As noted previously, the non-parametric method provides slightly biased estimates for

[16] The benchmark mean $E[b \mid n]$ is set equal to the theoretical mean of the bids, conditional on having a signal in region 2.

$n = 3$ and $n = 7$. Secondly, the non-parametric method allows one to find a decreasing (increasing) relationship between the average bids and the the number of bidders when the CV (PV) model applies. When $n = 3$ and $n = 7$ are not taken into account, the mean decreases (increases) with n when the CV (PV) model applies in 91.8% (95.2%) of the Monte-Carlo simulations. On the other hand, the empirical mean decreases (increases) with n when the CV (PV) model applies in only 32.1% (44.7%) of the simulations. The statistical tests conclude in favor of the true model 97.1% (88.3%) of the time and reject the alternative model 99.7% (94.8%) of the time with a significance level $\alpha = 0.1$ $(\alpha = 0.05)$.

I have also compared the regression and the non-parametric approach in simulations of different first price auctions models. The non-parametric approach consistently outperformed the regression method which was mostly inconclusive.

6 Conclusion

The goal of this paper was to analyze whether the theoretical prediction that the expected bid increases (decreases in general) with the the number of bidders when the PV (CV) assumption applies can be used to identify that a set of experimental data has been generated from a CV model with unobserved heterogeneity across auctions.

I have shown that existing regression methods are mostly inconclusive and sometimes misleading. I have also shown that, under the sole observation of bids and the number of bidders, a simple method based upon the non-parametric estimation of the expected bid validates the CV model with the experimental data. Moreover, this method appears to be robust relative to three elements: highly non-linear bid functions such as the one derived from Kagel and Levin experimental model; bid functions weakly increasing or decreasing in n; and noisy data such as slightly out of equilibrium bids.

The paper addresses the problem of specification in the particular case of an auction model. The methodology developed here could however be extended to other situations. There are indeed many open problems in economics where the relationship between two variables has not been clearly identified empirically. For instance, regression based tests of the "Risk Model" in agency theory (risk sharing increases with risk) or the "Phillips curve" (unemployment decreases with inflation) have produced mixed results. These examples share two characteristics with the auction model considered here: the analytical structure of the relationship is unknown and some relevant explanatory variables are unobserved (e.g. the risk aversion in the risk model). In such cases, a non-parametric estimation may be more reliable than regressions to identify the nature of the relationship between variables.

Armantier Olivier, 1998, "Empirical Application of Game Theory : Three Essays", Ph.D.. Dissertation, University of Pittsburgh.

Berger R., 1989, "Uniformly More Powerful Tests for Hypotheses Concerning Linear Inequalities and Normal Means", Journal of the American Statistical Association, Vol 84.

Berger R. 1985, "Statistical Decision Theory and Bayesian Analysis", Springer Verlag.

Brannaman L., Douglas Klein J. and L. Weis, 1987, "The Price Effects of Increased Competition in Auction Markets", The Review of Economics and Statistics, 24-32.

Capen E., Clapp R. and W. Campbell, 1971, "Competitive Bidding in High Risk Situation", Journal of Petroleum Technology 23, 641-653.

Csorgo M. and Revesz P., 1981, "Strong Approximation in Probability and Statistic", Academic, New York.

Giliberto M. and N. Varaiya, 1989, "The Winner's Curse and Bidder Competition in Acquisitions: Evidence from Fail Bank Auctions", The Journal of Finance, Vol. 54, No. 1, 59-75.

Goeree, J. and T. Offerman, 1999, "Competitive bidding in Auctions with Private and Common Values", Mimeo, University of Amsterdam.

Hardle Wolfgang, 1995, "Applied Nonparametric Regression", Econometric Society Monographs.

Kagel J., 1995, "Auctions: A Survey of Experimental Research" in "handbook of Experimental Economy".

Kagel J., Harstad R. and D.Levin, 1987, "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: a Laboratory Study", Econometrica, Vol 55, No. 6, 1275-1304.

Kagel J. and D. Levin, 1986, "The Winner's Curse and Public Information in Common Value Auctions", The American Economic Review, Vol 76 No.5, 894-919.

Kagel J. and J-F Richard, 1998, "Super-Experience Bidders in First-Price Common Value Auctions: Rules-Of-Thumb, Nash Equilibrium Bidding and the Winner's Curse", Mimeo, University of Pittsburgh.

Klemperer, 1999, "Auction Theory: A Guide to the Literature", Journal of Economic Survey.

Laffont J-J., 1997, "Game Theory and Empirical Economics : the Case of Auction Data", European Economic Review, 41, 1-35..

Matthews, S.A., 1984, "Comparing Auctions for Risk Averse Buyers: a Buyers Point of View", Econometrica, Vol 55, 633-646.

Milgrom Paul, 1979, "A Convergence Theorem for Competitive Bidding with Differential Information", Econometrica, Vol47 No 3.

Milgrom P. and R. Weber, 1982, "A Theory of Auction and Competitive Bidding", Econometrica 50, 1089-1122.

Paarsh H., 1992, "Deciding Between the Common and Private Value Paradigms in Empirical Model of Auctions", Journal of Econometrics, 51, 191-215.

Riley J. and M. Samuelson,1981, "Optimal Auctions", American Economic Review, 71, 381-92.

Shao, J. and D. Tu, 1995, "The Jackknife and Bootstrap", Springer-Verlag, New York.

Silverman, B.W, 1986, "Density Estimation for Statistics and Data Analysis", Chapman and Hall, London.

Vickrey W., 1961, "Counterspeculation, Auctions and Competitive Sealed Tenders", Journal of Finance, 16, 8-37.

Vuong Q., I. Perigne and E. Guerre (2000): "Optimal Nonparametric Estimation of First Price Auctions", Econometrica.

Wilson R., 1977, "A Bidding Model of Perfect Competition", Review of Economic Studies, 511-518.

Wilson R., 1992, "Strategic Analysis of auctions", Handbook of Game Theory, 227-277.