Pooling Information from Auctions with Varying Number of Bidders

Carolina Marquez, Kemal Guler, Hsiu-Khuern Tang, Bin Zhang
Intelligent Enterprise Systems Laboratory
HP Laboratories Palo Alto
HPL-2002-262
September 20th , 2002*

auctions, bidding strategies, nonparametric estimation, econometrics of auctions

First price auctions are among the most commonly used auction mechanisms, especially in procurement. Recent research activity in economics of auctions focuses on development of econometric methods to analyze the bid data from such auctions to estimate the underlying structural variables. Optimization of almost all auction-related decisions (reserve price, auction format, information rules) requires reliable estimates of these structural elements. Current nonparametric methods to estimate the distribution of bidders' private valuations in first price auctions make use of functionals of bid distributions. Currently known estimation methods require separate estimation of these functionals for each configuration of observable variates. In many situations the sizes of some subsamples are too small to obtain reliable estimates of the functionals of interest. In such cases, the practice is to discard such subsamples and work with the remaining observations to estimate the latent valuation distributions. This waste of information may have substantial impact on the reliability of the final estimates and the decisions based on such estimates. Especially in situations where the total size of available samples are small to start with, the need for methods and algorithms that make use of all available data is obvious. We propose a new fully nonparametric approach that makes full use of all available auction data. This paper describes the methods based on the new approach and presents Monte Carlo experimental evaluation of alternative methods of combining information from auctions with varying numbers of bidders.

* Internal Accession Date Only

Pooling Information from Auctions with Varying Number of Bidders

Carolina Marquez, Kemal Guler, Hsiu-Khuern Tang, Bin Zhang

Hewlett-Packard Laboratories

August 30, 2002

Section 1: Introduction

First price auctions are among the most commonly used auction mechanisms, especially in procurement. Recent research activity in economics of auctions focuses on development of econometric methods to analyze the bid data from such auctions to estimate the underlying structural variables. Optimization of almost all auction-related decisions (reserve price, auction format, information rules) requires reliable estimates of these structural elements.

Current methods to estimate the distribution of bidders' private valuations in first price auctions make use of functionals of bid distributions. These functionals are estimated using data on bids and other observable variates. Currently known estimation methods require separate estimation of these functionals for each configuration of observable variates. For example, the available sample is broken into subsamples based on the number of bidders in the auction. In many situations the sizes of some subsamples are too small to obtain reliable estimates of the functionals of interest. In such cases, the practice is to discard such subsamples and work with the remaining observations to estimate the latent valuation distributions. This waste of information may have substantial impact on the reliability of the final estimates and the decisions based on such estimates. Especially in situations where the total size of available samples are small to start with, the need for methods and algorithms that make use of all available data is obvious.

Increasingly, auctions are an integral part of a company's strategy for procurement and excess disposal. Auctions can produce substantial cost savings by lowering transaction and negotiation costs, and reducing uncertainty. In order to remain competitive in this arena it is important to establish methods for making optimal auction decisions. These decisions range from a seller setting the optimal reserve price to a bidder estimating the characteristics of his competitors. Underlying all auction decisions is the joint distribution of bidder values. With this characterization of the auction market structural variables can be estimated and bidder behavior can be forecast.

The goal is to use observed bid data to estimate the distribution of values. By imposing a family of distributions, the equilibrium strategy can be estimated directly or through maximum likelihood methods[1]. This is the traditional method, however for many distributions the equilibrium solution does not exist in a closed form or is too complicated to calculate. Recent research has developed an indirect estimation method to deal with these cases[2]. The general idea is to use the inverse bid function to convert the observed bids into corresponding values. The value distribution can then be estimated from these pseudo-values. This method is also useful because it lends itself to nonparametric estimation. This is an attractive quality because nonparametric estimation allows estimation methods and recommendations to be generalized to a variety of products and auction markets. The downside is that accurate nonparametric estimation requires large samples, which are not often available.

A related empirical issue is the role of the structural variables of the auction. The data used for estimation is gathered over a series of auctions. These auctions may be for similar markets and comparable products but they do not necessarily share the same structural variables. The theoretical framework used to derive the inverse bid function, however, assumes that the set of structural variables is constant for all auctions. As a result empirical analysis must be restricted to sub samples of the data where the set of structural variables is fixed. This further reduces the amount of data available for estimation and creates two sources of information loss: loss of small sub samples and loss of correlation between sub samples.

In this paper we develop a estimation methods that address these issues. The focus is on an estimation method that incorporates data for auctions with a varying set of structural variables rather than restricting analysis to sub samples of the data. In section 2 we setup the model. In section 3 we develop the experiments. In section 4 we discuss the results. Section 5 concludes.

Section 2: Model

The key component of the model is the bid function. For a given set of structural variables the bid function represents the optimal bid corresponding to all values of a specified value distribution, $F(v_1, v_2, \ldots v_n)$. To simplify the analysis we let the number of bidders be the sole structural variable. In addition we assume the bid function is identical for all bidders.

$$b_i = s(v_i, F, n)$$

While this is the most direct way to represent bidder behavior, in empirical analysis bids are observed. What is needed for analysis is a function relating bidder values to the distribution of posted bids, $G(b_1, b_2, \ldots b_n)$.

$$v_i = s^{-1}(b_i, G, n)$$

Both functions are derived from the bidder's optimization problem. A bidder's objective is to set a bid that maximizes expected utility. Under the assumptions of risk neutrality and linear utility we maximize the following function

(1) $$E(U_i) = (v_i - b_i)P(b_i \geq b_{i\neq j}) .$$

The probability of setting the winning bid is naturally a function of the bid distribution. However under equilibrium the probability of setting the winning bid is equal to the

[1] Important direct estimation method papers.
[2] Reference to indirect estimation papers.

probability of drawing the highest value. Therefore we can rewrite this probability in terms of the distribution of bidder values as well. The model is developed for two assumptions: Independent Private Values (IPV) and Affiliated Private Values (APV). In addition we assume that values are exogenous to the structural variables.

Independent Private Value Model

In the IPV case we assume that bidders have independent and identical marginal value distributions, $F(v_1, v_2, ..., v_n) = F^n(v)$. For a given value, v_i, the probability of setting the winning bid is given by $F^{n-1}(v_i)$. The first order condition of the bidder's optimization problem can be written as

(2) $F(v_i)s'(v_i) = [v_i - s(v_i, n)](n-1)f(v_i)$.

The solution yields the bid function for the IPV case.

(3) $$b_i = s(v_i, n) = v_i - \frac{1}{[F(v_i)]^{n-1}} \int_{\underline{v}}^{v_i} [F(u)]^{n-1} du$$

Under equilibrium, we claim $F(v_i) = F(s^{-1}(b_i, n)) = G(b_i, n)$ and $g(b_i, n) = \frac{f(v_i)}{s'(v_i, n)}$.

Rewriting the differential equation in (2) yields the inverse bid function.

(4) $$v_i = b_i + \frac{1}{n-1} \frac{G(b_i)}{g(b_i)}$$

The bid distribution and density can be estimated nonparametrically from a set of observed bids. Here L is the number of auctions in the dataset and n is the number of bidders.

$$G(b) = \frac{1}{nL} \sum_{l=1}^{L} \sum_{i=1}^{n} 1(b_{il} \le b)$$

$$g(b) = \frac{1}{nLh_g} \sum_{l=1}^{L} \sum_{i=1}^{n} K(\frac{b - b_{il}}{h_g})$$

While the value distribution is assumed exogenous of the number of bidders, the same cannot be assumed for the bid distribution. The dependence on structural variables is passed onto the bid distribution through the bid function. As a result, the inverse bid function can only be applied to auctions with the same number of bidders. So while the inverse bid function is the most straightforward tool for estimating values, these calculations must be restricted to sub samples with a given number of bidders. In addition the nonparametric equations for estimating the bidder density and distribution require a large number of observations[3]. As a result, small sub samples of the dataset are lost. Even for sub-samples with a sufficient number of observations there is information loss from ignoring the correlation between sub-samples with different number of bidders. While the structural variables may vary, auctions that serve the same market share information that can be used to refine estimation. The hypothesis is that better estimates can be obtained by a method that captures these sources of information loss. The challenge is to find a way of combining data from auctions with a varying number of

[3] The necessary number of observations is determined by the characteristics of the dataset and the nature of estimation.

bidders. Since the value space is independent of the number of bidders it is a good candidate for this role.

Any sub sample that has a sufficient number of observations can be converted to pseudo values by individual calculation on each sub sample. Since the value distribution is independent of the structural variables, the pseudo-values from each sub sample can be pooled together to calculate a single empirical distribution. This process combines the data from auctions with a varying number of bidders, but it does not address either of the information loss issues. The small sub samples have been dropped, and each sub sample has been analyzed independently, ignoring any correlation. Assume there exists a sub sample with sufficient observations for estimation by the inverse bid function. The estimated pseudo-values can be used to calculate an empirical distribution. While estimation is only conducted for the given sub sample, the estimated distribution is an acceptable representation of the general value distribution. As a result it can be applied to auctions with any number of bidders. In particular, the estimated value distribution can be used in the direct bid function to create a set of sample bids that correspond to any other sub sample[4]. The set of estimated pseudo-values and calculated bids form a correspondence that can be viewed as a numerical bid function. At this point, the observed bids from the second sub sample can be mapped onto the sample bids to extract value estimates through linear inversion. The new set of estimated pseudo-values can be pooled directly with the original set since values are independent of the number of bidders. The empirical distribution of this expanded set of values now incorporates data from auctions with two separate numbers of bidders. This distribution can be used to update the value distribution estimate, and the process can be repeated for all sub samples in the dataset. In the end, the final value distribution is calculated using the entire data set as a single unit[5]. This is the general framework of the IPV experiments performed.

Affiliated Private Values Model

Most auction markets do not conform to the Independent Private Value assumption, thus the Affiliated Private Value model is more realistic and can provide more useful recommendations. In this model it is not appropriate to simplify the joint value distribution since the marginal distributions are correlated. As a result the conditional value distribution is the main determinant of the bidding decision. Under equilibrium, the probability of setting the winning bid is given by $F_{y_1|v_1}(v_i \mid v_i)$ where $y_i = \max_{j \neq i} v_j$. As in the IPV case the bid function is obtained by solving the following differential equation.

$$F_{y_1|v_1}(v_i \mid v_i)s'(v_i) = [v_i - s(v_i)]f_{y_1|v_1}(v_i \mid v_i)$$

The APV bid function is expressed as

(6) $$b_i = v_i - \int_{\underline{v}}^{v_i} L(\alpha \mid v_i)d\alpha \text{, with } L(\alpha \mid v_i) = \exp[-\int_{\alpha}^{v_i} \frac{f_{y_1|v_1}(u \mid u)}{F_{y_1|v_1}(u \mid u)} du].$$

[4] Numerical integration is used to calculate bids from the direct bid function.

[5] For certain analysis it may be useful to estimate the value density as well. Kernel estimation with a varying number of bidders is complicated due to bandwidth selection and has not been implemented in this project.

The inverse bid function is obtained by setting $G_{B_1|b_1}(b_i \mid b_i) = F_{y_1|v_1}(s^{-1}(b_i) \mid s^{-1}(b_i))$ and $B_1 = s(y_1)$ to get

(7) $$v_i = b_i + \frac{G_{B_1|b_1}(b_i \mid b_i)}{g_{B_1|b_1}(b_i \mid b_i)}.$$

One of the challenges of the APV implementation is estimating the conditional distribution and density. For the inverse bid function it is necessary to calculate $G_{B_1|b_1}(b_i \mid b_i) / g_{B_1|b_1}(b_i \mid b_i)$. This can be estimated by the joint analog where

$$G_{B_1,b_1}(B \mid b) = \frac{1}{Lh_G} \sum_{l=1}^{L} \frac{1}{n} \sum_{i=1}^{n} 1(B_{il} \le B) K_G(\frac{b - b_{il}}{h_G}), \text{ and}$$

$$g_{B_1,b_1}(B \mid b) = \frac{1}{Lh_g^2} \sum_{l=1}^{L} \frac{1}{n} \sum_{i=1}^{n} K_g(\frac{B - B_{il}}{h_g}, \frac{b - b_{il}}{h_g}).$$

These formulas can also be used to calculate $f_{y_1|v_1}(u \mid u) / F_{y_1|v_1}(u \mid u)$ in the direct bid function. However once the pseudo-values have been calculated, the conditional distribution itself is of interest. To estimate the conditional distribution and density, the joint formulas above are divided by the probability of observing the conditioning factor.

$$g_{b_1}(b) = \frac{1}{nLh_g} \sum_{l=1}^{L} \sum_{i=1}^{n} K_g(\frac{b - b_{il}}{h_g})$$

Both the conditional bid distribution and the conditional value distribution rely heavily on the structural variables of the auction. There is no longer a common space in which the data from auctions with a varying number of bidders can be combined. However it is still possible to develop a process that parallels the estimation method laid out in the IPV case. Even though the conditional value distribution depends on the number of bidders in the auction, the exogeneity assumption still holds for the joint distribution. A set of pseudo-values estimated from a given sub sample could just as easily occur in an auction with a different set of bidders. As in the IPV estimation, the first step is to use the inverse bid function to estimate pseudo-values on a large enough sub sample. Those values can be re-sampled to create a set of values that conforms to the dimensions of another sub sample. The direct bid function can now be applied to the re-sampled set of values to calculate a set of sample bids. Once again as in the IPV case, the observed data can be mapped onto the sample bids to obtain the new pseudo-value estimates. For the next sub sample values can be re-sampled from both the original and new sets of pseudo-values.

While this method parallels the IPV estimation rather closely, there are two important distinctions that weaken the estimation. First, re-sampling is only possible for sub samples with fewer numbers of bidders. If the initial sub sample has five bidders. Then all possible combinations can be stacked to obtain re-sampled value for four, three, and two bidders. But straightforward re-sampling is not possible for six or seven bidders. All sub samples with number of bidders greater than the initial sub sample must be dropped. The APV estimation method also has a greater degree of correlation loss than the IPV method. Since a separate conditional distribution must be calculated for each sub sample, only information from auctions with a greater number of bidders is used to enhance the estimation. The data set is not used as a single unit as in the IPV case.

Section 3: Experiments

The purpose of the proposed estimation methods is to convert bid data into pseudo-values that can be used to characterize the auction market. In order to evaluate the performance of each experiment, the results are compared to the known distribution used to generate the data[6]. In this paper bidder values are drawn randomly from a beta distribution representing the characteristics of the market[7]. The data used is drawn from a beta distribution with a mean of 60 and a standard deviation of 1.2. Using the bid function and the known value distribution, bids are calculated for each value. This dataset represents the data one could observe from a series of auctions. A variety of sample size distributions are used in order to test for robustness to clustering and sample composition. For a given data set, the sample size distribution refers to the number of auctions in each sub sample. Normally, this follows a lognormal distribution on the number of bidders. Experiments are run on a variety of samples chosen by varying the parameters of the log normal distribution. The frequency of auctions is chosen from the lognormal for 2 through 6 bidders in the IPV case and for 2 through 10 bidders in APV estimation[8]. Bid data is calculated separately for each auction. The data used to test APV estimation is injected with correlation through a standard stretching algorithm. An independent vector of values is drawn from the beta distribution. A common affiliation is added to each independent value. Here we have use one half the mean of the drawn values. Finally, each sub sample is classified as either a large sample or a small sample. A large sample is considered large enough for accurate nonparametric estimation. In these experiments a sample is considered large enough for estimation if it has at least 40 observations[9]. The experiments vary in their treatment of large and small samples.

First each dataset is analyzed using the conventional method. The inverse bid function is applied to each large sub sample separately and the small sub samples are dropped. In the IPV case all pseudo-values are pooled to calculate a single value distribution. In the APV case a separate conditional value distribution is calculated for each sub sample without pooling. The next experiment addresses the issue of information loss due to dropping small samples. The data from small sub samples cannot be used directly in nonparametric estimation, however estimates for these samples can be obtained by using the pseudo-values from the large sub samples. The large sub samples are still treated individually, but now sampling methods are applied to the small sub samples. In the IPV case the pseudo-values from the large samples are pooled to calculate a single value distribution. Using the direct bid function and the large sample estimates, sample bids are calculated for a given small sub sample. Pseudo-values for that sample are obtained by matching the observed bids to the sample bids. The new

---

[6] Using actual bid data is less effective because the theoretical distribution is unknown therefore goodness of fit cannot be assessed just by comparison.

[7] The beta distribution is chosen for its flexibility. A variety of bidder behaviors can be characterized by adjusting the parameters. Affiliation is incorporated into the samples used to test the APV methods through a standard stretching algorithm.

[8] Larger samples are used in the APV case in order to make better comparisons. In the APV case the total size of the dataset is also larger.

[9] The experiments were run for several thresholds: 50, 45, 40, 35, 25, 20. The qualitative results were similar for all thresholds. Goodness of fit decreased as the threshold fell, due to increased error in the initial sample.

pseudo-values are added to the pool and the procedure is repeated for the next small sub sample. In the APV case the large sample results are not pooled together. The pseudo-values from each large sub sample are re-sampled separately to create all combinations of value data conforming to the number of bidders in a given small sub sample. The re-sampled data is used to calculate sample bids for the small sample. Matching extracts the value estimates. Using the direct bid function in this way allows the small sub samples to be incorporated into estimation, and captures a portion of the shared information between large sub samples and small sub samples. However the majority of the analysis is conducted for each sub sample separately.

The final two experiments explore different ways of incorporating correlation between sub samples. First, the inverse bid function is applied to the largest sub sample. This sub sample is chosen to initialize the pool because it is likely to produce the most accurate individual estimate due to its sample size. This is important because the initial sub sample has the greatest influence on subsequent estimation. Next, the pseudo-values from the initial sub sample are used in the direct bid function to calculate a set of sample bids corresponding to the next largest sub sample. Pseudo-values are obtained by matching the observed bids to the sample bids, and added to the pool. The newly expanded pool is then used to calculate pseudo-values for the next largest sub sample and the process continues for the entire dataset. At the end of estimation, the pool of values contains data from each sub sample and captures correlation between the sub samples. In the IPV case the final experiment explores the degree to which information sharing can improve estimation. If there is truly a gain from information sharing, the estimates should improve with each loop through the data. In the IPV case, it is possible to test this hypothesis by iterating back to the largest sub sample and repeating the estimation process until the value estimates fail to improve.

Looping back through the dataset is difficult in the APV case because the sampling procedure relies on the pooled data being of higher dimension than the sub sample in question. However there are several refinements that can be made. A big issue in the APV procedure is the size of the sub sample used to initialize the pool. Using the largest sub sample means that sub samples with a greater number of bidders than the initial set, but fewer total observations, are dropped from the analysis. If these sub samples are large enough they can be estimated separately, but perhaps a procedure that includes such sub samples benefits from a greater degree of information sharing. Ideally, the initial sub sample would be the one with the largest number of bidders. However, nonparametric estimation requires sufficiently large samples. An alternative to using the largest sub sample is to use the sub sample with the largest number of bidders that satisfies a reasonable observation threshold[10]. All experiments are illustrated in figure 1.

<u>Section 4: Results</u>

Each experiment is run several times to reduce the influence of any one dataset. For each run, the goodness of fit is measured by the maximum distance between the

[10] If analysis is restricted to a given number of bidders, the initial sub sample can be chosen as the one that will yield the largest set of re-sampled pseudo-values. This enhances estimation since nonparametric methods improve with an increase in sample size.

theoretical value distribution chosen at the beginning of the experiment[11] and the empirical distribution calculated from the generated bids. The Wilcox rank sum test is used to determine whether the goodness of fit differs significantly between any two experiments[12].

Independent Private Value Model

The results for the IPV model are in Table 1. In all samples there is an improvement in estimation from just adding the small samples. The p-value on the rank test for the first two experiments is close to 0. Incorporating the correlation between sub samples is an even greater improvement. Sampling through the entire dataset yields an average improvement of 30%. As expected, estimation for all experiments improves as the sample size increases. For datasets with close to 100 auctions, the average error using the conventional method is approximately 0.13. Using the sampling method this same error can be obtained for datasets with only 20 auctions. For those datasets with 100 auctions the sampling method produces an error of around 0.08. In the IPV case, the sampling method is very effective at extracting information from small datasets that otherwise could only be obtained from unreasonably large samples. Even for large samples, where very few observations are dropped by conventional methods, significant improvement is obtained by incorporating the correlation between sub samples and thereby treating the dataset as a single unit. Since the results demonstrate that sharing information between sub samples is an important refinement to estimation, one would expect that further iteration would continue to improve estimation. However the rank test results for the fourth experiment do not support this hypothesis.

Affiliated Private Value Model

Table 2 contains the results for the APV case. Here results vary with each sub sample and rely more heavily on the composition of the data than the IPV results. In general, by sampling the large sub samples good results can be obtained for small sub samples that would otherwise have been dropped. The estimation error for these sub samples is 0.1 on average[13]. However estimating for all sub samples through re-sampling is not an appropriate estimation method. In almost all cases the results for experiment 3 are either equivalent to or worse than those for experiment 2.

Sampling from the largest possible number of bidders as opposed to sampling from the largest sub sample produces the worst results[14]. In experiment 4 the initial sub

---

[11] Since the estimation output in the APV case is a conditional distribution, the theoretical distribution used to evaluate goodness of fit is simulated to obtain a conditional distribution for number of bidders in question.

[12] In the IPV case, the Kolmogorov-Smirnov statistic can be used evaluate goodness of fit as well. Here we compare two vectors of values by using the ranksum() command in MATLAB. This command performs a two-tailed rank test.

[13] For experiment 2 there are quite a few two-bidder sub sample that have estimation error of around 0.2, which is not very strong. However careful examination reveals that these sub samples are considered large enough for estimation so re-sampling has not been applied.

[14] The only exception is Sample 9 where the improvement is minimal. The p-value for the rank test between experiment 2 and 3 is less than 25% for at least one sub sample and the average maximum

sample is the one with the largest number of bidders that satisfies the given threshold of 40 observations. By using this criteria instead of the largest sub sample, we are selecting a sub sample with inferior estimation power and passing error on to the subsequent sub samples instead of capturing useful information. The realm of estimation is expanded at the cost of lost information. The re-sampling employed in the APV case magnifies the importance of selecting an appropriate initial sub sample. However, the results indicate that estimating each large sub sample separately is preferable, regardless of the sub sample chosen to initialize the pool.

While the re-sampling method is a good first step and can safely be used to estimate pseudo-values for small samples, a different method must be developed to capture the correlation between sub samples. The re-sampled data can grow to be quite large since re-sampling stacks all combinations of previously estimated values. It is likely that the estimation results obtained in experiment 3 and 4 are being driven by the re-sampled data. Useful information in the observed samples is dwarfed by the artificial samples. As estimation continues for each sub sample, the "bad information" is perpetrated and amplified through re-sampling. Due to this contamination, better results are obtained by estimating large sub samples individually and capturing only the information they can provide on their own. Another issue to consider is the estimation noise that is introduce by the sampling method. In order to use the direct bid function in estimation it is necessary to calculate numerical integrals and perform several numerical linear inversions. These calculations introduce error that is not present when using the inverse bid function alone. In the IPV case the gain from capturing correlation between sub samples outweighs the estimation noise, but it could be that the APV results are substantially affected by this error. One last observation on the APV results is that the estimates do not improve as sample size increases. This goes against very basic expectations and suggests that the data used for estimation is affiliated in such a way that estimation is poor for even the best method. Greater attention in future work should be placed on how the data used for experimentation is generated.

Section 5: Conclusion

Traditional auction estimation methods are based on the direct bid function and require imposing a family of distribution on the data. More recent research has relaxed this restriction by developing indirect estimation methods that use the inverse bid function to calculate pseudo-values from the observed data. However even with these methods, the analysis must be restricted to sub samples of the data that share a given set of structural variables. This leads to two sources of information loss: loss of small samples that are inadequate for estimation, and loss of correlation between sub samples that differ in their structural variables. Treating the dataset as a single unit, rather than separate sub-samples, minimizes the information loss and yields better estimates. Results from previously estimated sub samples can be used to refine estimation and compensate for insufficient sample size. In this paper we develop an estimation method that uses the value space to combine data from various sub samples.

---

distance from re-sampling for all sub samples is less than that from re-sampling for only the small sub samples. In Sample 9 improvement is likely obtained simply by the sheer magnitude of data that is generated from re-sampling six large sub samples.

Under the IPV assumption, this estimation method performs very well. Pooling over sub samples improves results 30% over the standard indirect estimation method. In addition, the results obtained using the standard method can be replicated by the sampling method with data from just 20% of the auctions. While this method allows useful information to be passed on to all sub samples and thereby help refine estimation, error can also be passed on through the same channels. Therefore it is important to be cautious in setting the initial sub sample and orchestrating the order of iteration.

In the APV case the sampling method produces relatively strong results for small sub samples that would otherwise be dropped by the standard method. However it does not adequately capture the correlation between sub samples. Further development is needed. Special attention must be placed on how the data is generated, estimation noise, and over-sampling. It would also be useful to address the issue of sampling up to a higher dimension.

There are many directions for further development in both the IPV and APV case. A significant difference between the IPV method and the APV method is in the output. The results reported in this paper for the APV model are based on conditional distribution calculated separately for each sub sample. Perhaps if a single joint distribution were estimated, as in the IPV case, we would observe positive information sharing between the sub samples. In both cases, density estimation may produce additional information. One challenge in density estimation is setting the kernel bandwidth for data with varying dimensions. In general, customizing the kernel estimates to specific markets or decisions may improve estimation. Another variation is to explore more efficient ways of pooling the data, perhaps incorporating the estimation error of a particular sub sample. Estimating bidder values through these pooling methods yield very promising results. With additional refinement and implementation towards application, pooling methods can contribute greatly to developing optimal auction strategies.

References

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

Klemperer, Paul (1999). "Auction Theory: A Guide to the Literature." *Journal of Economic Surveys*, 13(3), pp. 227-286.

Laffont, Jean-Jacques (1997). "Game Theory and Empirical Economics: the Case of Auction Data." *European Economic Review*, 41, pp. 1-35.

Laffont, Jean-Jacques and Quang Vuong (1996). "Structural Econometric Models of Strategic Behavior." *AEA Papers and Proceedings*, 86(2), pp. 414-420.

Li, Tong, Isabelle Perrigne, and Quang Vuong (2001). "Semiparametric Estimation of the Optimal Reserve Price in First-Price Auctions." *Journal of Business Economics & Statistics (forthcoming)*.

Perrigne, Isabelle and Quang Vuong (1999). "Structural Econometrics of First-Price Auctions: A Survey of Methods." *Canadian Journal of Agricultural Economics,* 47, pp. 203-223.

Figure 1A: Experiment 1

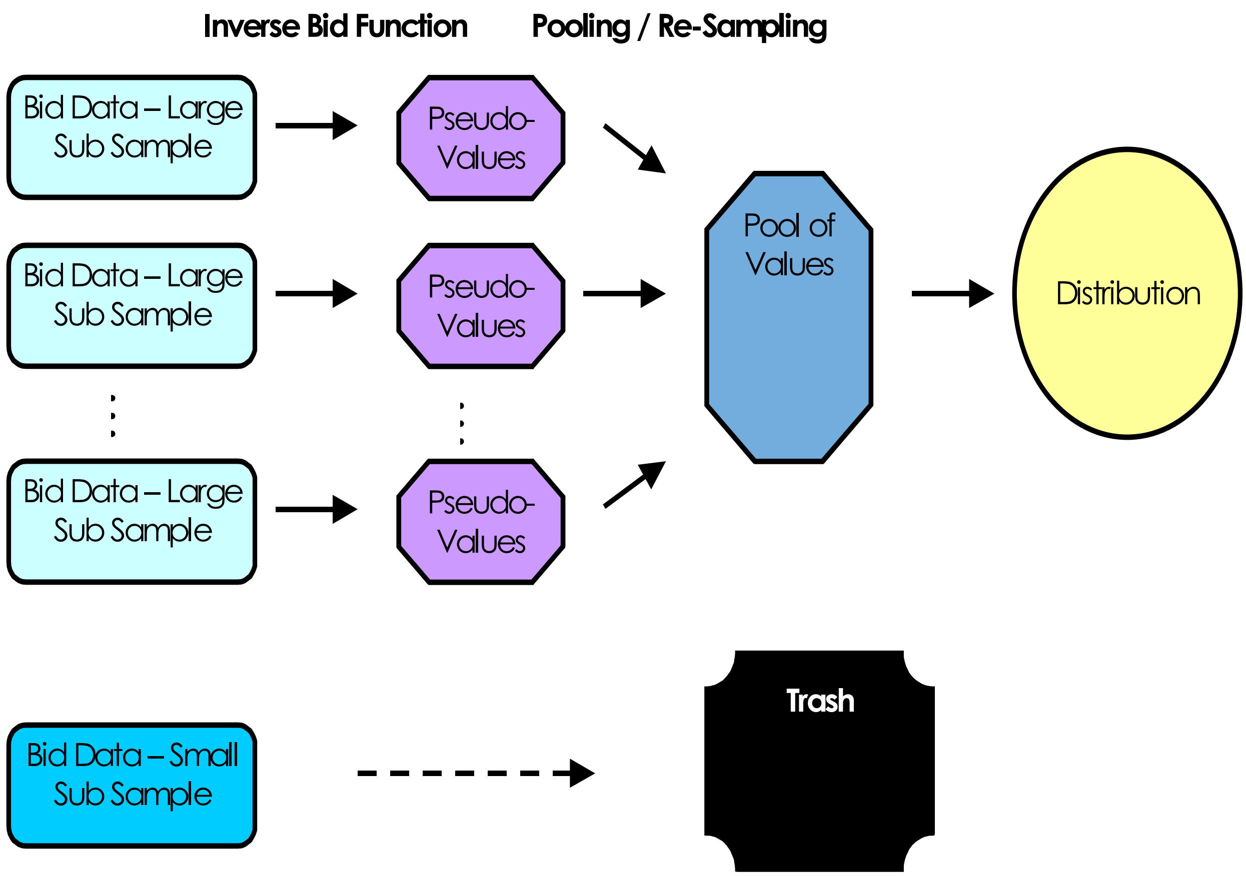

Figure 1B: Experiment 2

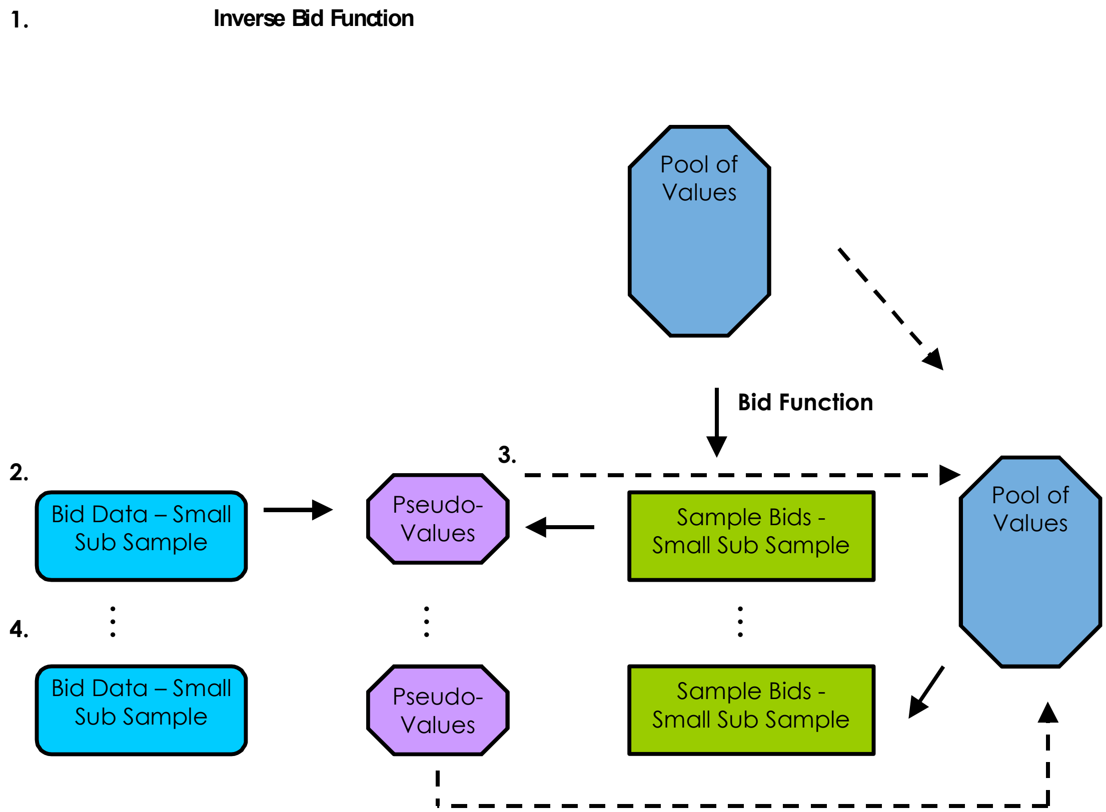

Figure 1C: Experiment 3

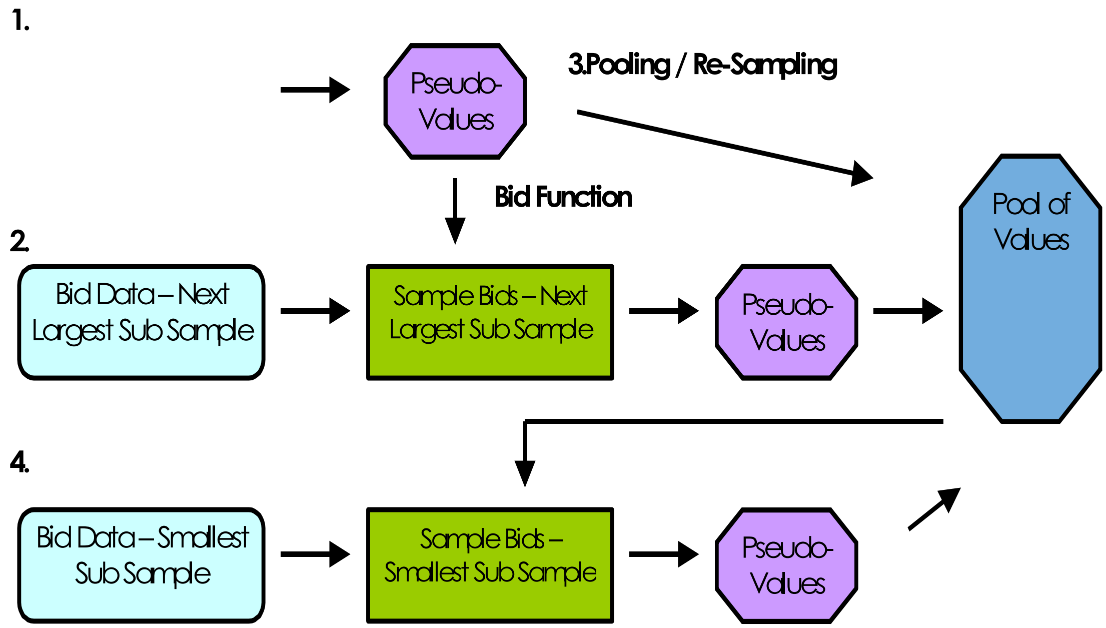

Figure 1D: Experiment 4 – IPV

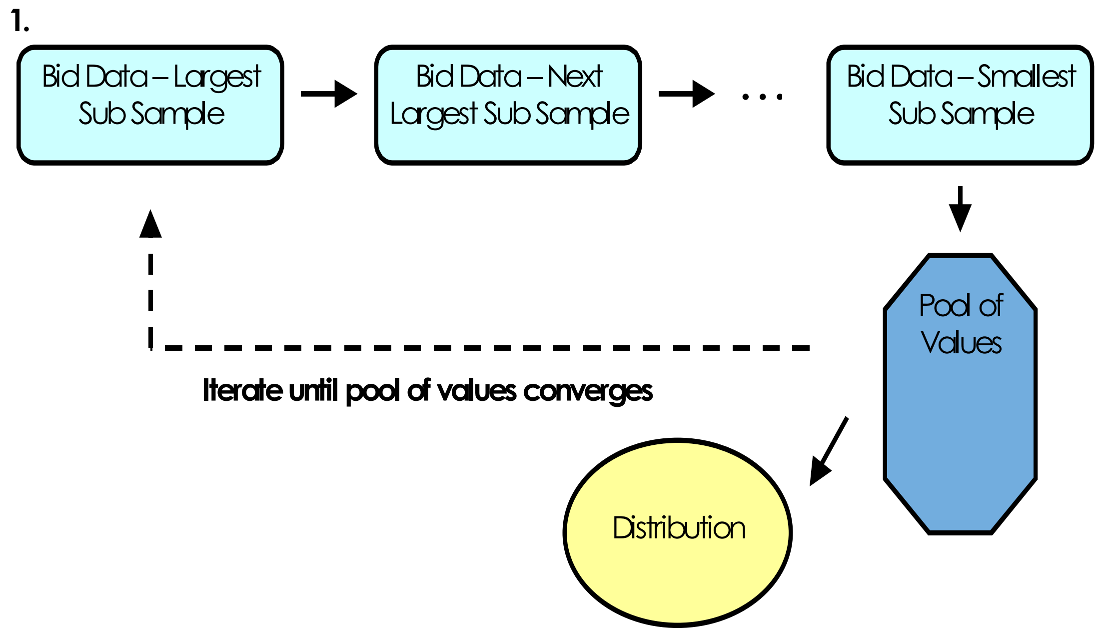

Figure 1E: Experiment 4 – APV

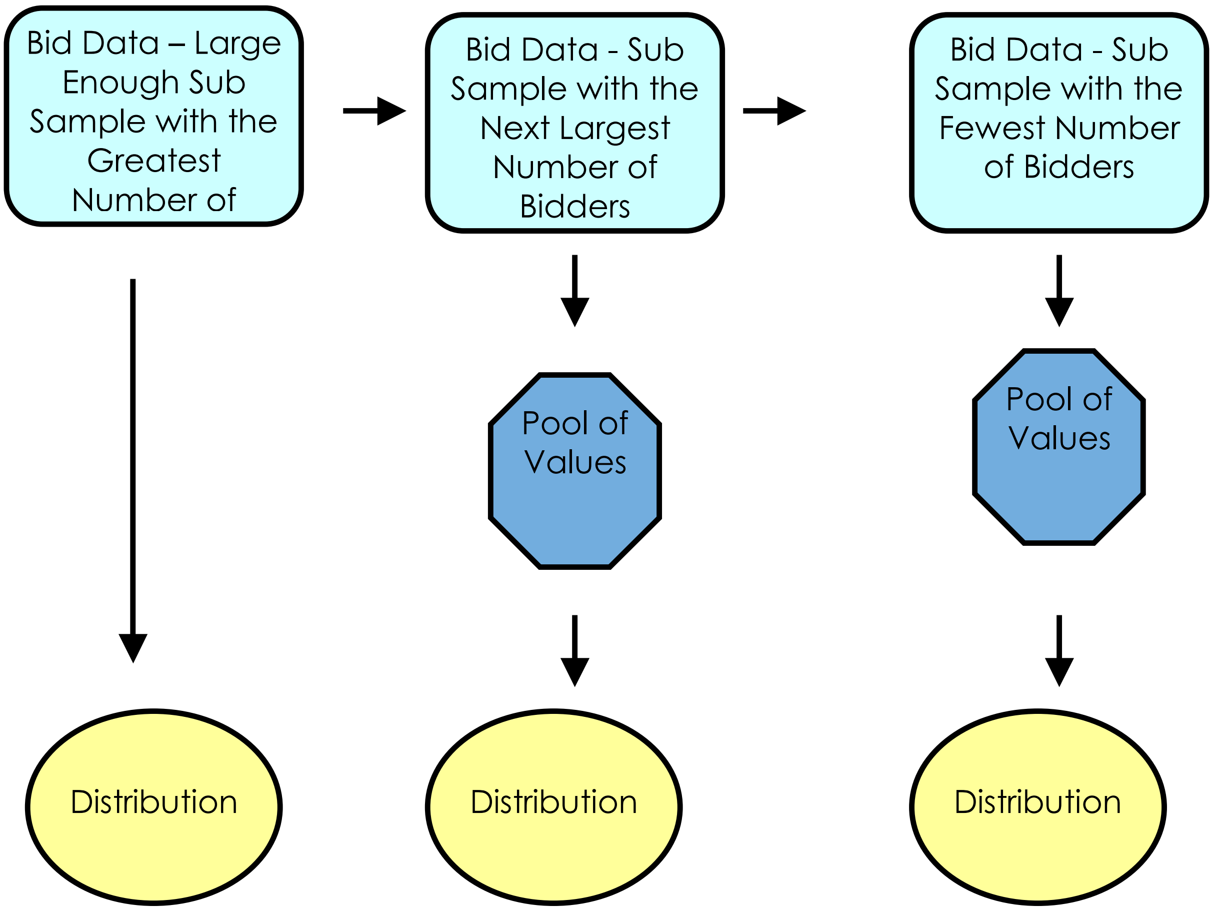

Table 1: IPV Results

Maximum Distance Between Theoretical and Empirical Distribution

(Averaged over 100 Runs)

Number of Bidders	2	3	4	5	6	No. Auctions	No. Observ.
Number of Auctions - Sample 1		6	12	2		20	76
1. Drop Small Samples	0.167						
	(0.042)						
2. Sample for Small Sample Estimates	0.134						
	(0.042)						
3. Sample for all Estimates - One Loop	*Same as Experiment 2*						
4. Sample for all Estimates - Iterate to Convergence	0.131						
	(0.039)						
Rank Test (p-value)							
H0: 1 = 2	0.00						
H0: 2 = 3							
H0: 3 = 4	0.66						
Number of Auctions - Sample 2		15	29	4		48	181
1. Drop Small Samples	0.147						
	(0.026)						
2. Sample for Small Sample Estimates	0.134						
	(0.024)						
3. Sample for all Estimates - One Loop	0.108						
	(0.027)						
4. Sample for all Estimates - Iterate to Convergence	0.108						
	(0.026)						
Rank Test (p-value)							
H0: 1 = 2	0.00						
H0: 2 = 3	0.00						
H0: 3 = 4	0.75						
Number of Auctions - Sample 3		31	59	8		98	369
1. Drop Small Samples	0.129						
	(0.016)						
2. Sample for Small Sample Estimates	*No Small Samples*						
3. Sample for all Estimates - One Loop	0.089						
	(0.020)						
4. Sample for all Estimates - Iterate to Convergence	0.088						
	(0.019)						

Table 1: IPV Results (cont.)							
Maximum Distance Between Theoretical and Empirical Distribution							
(Averaged over 100 Runs)							
Number of Bidders	2	3	4	5	6	No. Auctions	No. Observ.
Number of Auctions - Sample 4	1	7	6	3	1	18	68
1. Drop Small Samples	*No Large Samples*						
2. Sample for Small Sample Estimates	*No Large Samples*						
3. Sample for all Estimates - One Loop	0.139						
	(0.044)						
4. Sample for all Estimates - Iterate to Convergence	0.137						
	(0.039)						
Rank Test (p-value)							
H0: 1 = 2							
H0: 2 = 3							
H0: 3 = 4	0.84						
Number of Auctions - Sample 5	2	19	16	8	3	48	183
1. Drop Small Samples	0.148						
	(0.025)						
2. Sample for Small Sample Estimates	0.136						
	(0.025)						
3. Sample for all Estimates - One Loop	0.098						
	(0.032)						
4. Sample for all Estimates - Iterate to Convergence	0.098						
	(0.028)						
Rank Test (p-value)							
H0: 1 = 2	0.00						
H0: 2 = 3	0.00						
H0: 3 = 4	0.93						
Number of Auctions - Sample 6	4	37	32	16	6	95	363
1. Drop Small Samples	0.134						
	(0.018)						
2. Sample for Small Sample Estimates	0.119						
	(0.016)						
3. Sample for all Estimates - One Loop	0.074						
	(0.020)						
4. Sample for all Estimates - Iterate to Convergence	0.076						
	(0.018)						

Table 1: IPV Results (cont.)							
Maximum Distance Between Theoretical and Empirical Distribution							
(Averaged over 100 Runs)							
Number of Bidders	2	3	4	5	6	No. Auctions	No. Observ.
Number of Auctions - Sample 7	3	6	4	3	2	18	67
1. Drop Small Samples	*No Large Samples*						
2. Sample for Small Sample Estimates	*No Large Samples*						
3. Sample for all Estimates - One Loop	0.136						
	(0.045)						
4. Sample for all Estimates - Iterate to Convergence	0.135						
	(0.035)						
Rank Test (p-value)							
H0: 1 = 2							
H0: 2 = 3							
H0: 3 = 4	0.77						
Number of Auctions - Sample 8	8	15	11	7	4	45	164
1. Drop Small Samples	0.167						
	(0.029)						
2. Sample for Small Sample Estimates	0.113						
	(0.025)						
3. Sample for all Estimates - One Loop	0.102						
	(0.028)						
4. Sample for all Estimates - Iterate to Convergence	0.102						
	(0.024)						
Rank Test (p-value)							
H0: 1 = 2	0.00						
H0: 2 = 3	0.00						
H0: 3 = 4	0.86						
Number of Auctions - Sample 9	17	29	22	14	8	90	327
1. Drop Small Samples	0.132						
	(0.020)						
2. Sample for Small Sample Estimates	0.123						
	(0.019)						
3. Sample for all Estimates - One Loop	0.074						
	(0.023)						
4. Sample for all Estimates - Iterate to Convergence	0.079						
	(0.021)						

Figure 2: IPV Estimation

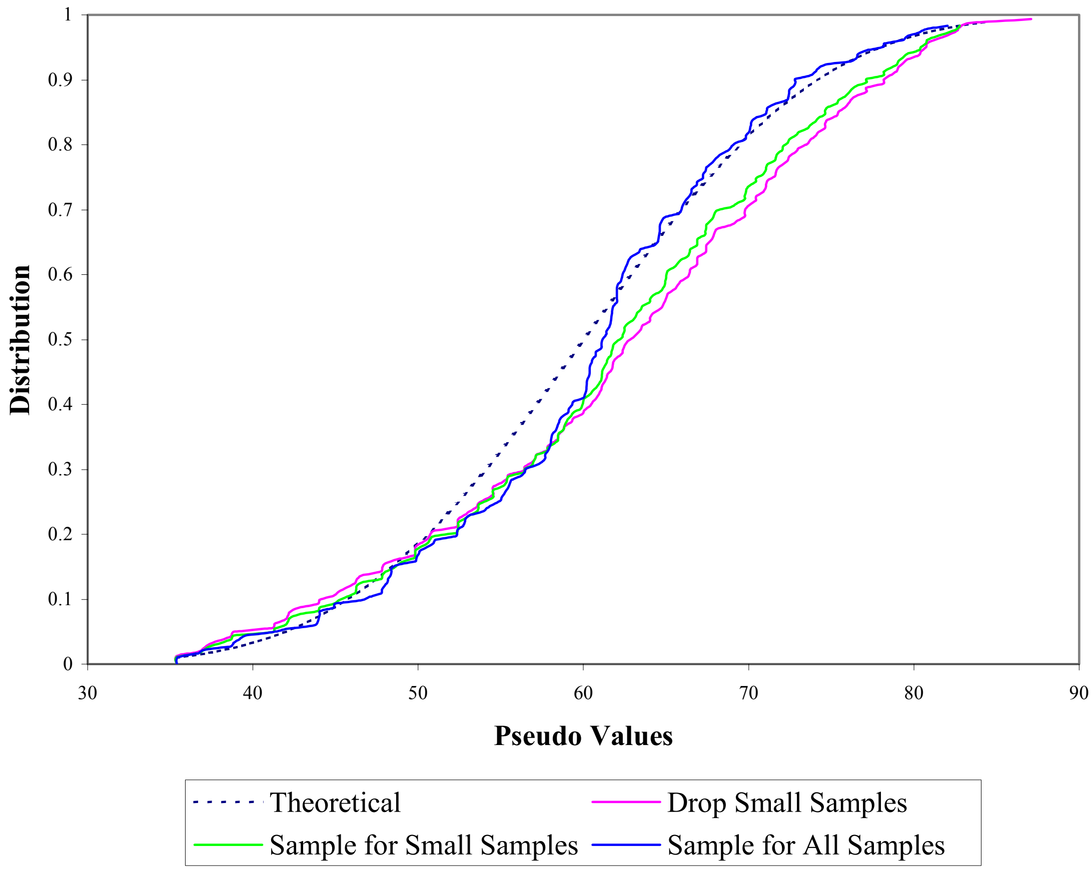

Table 2: APV Results								
Maximum Distance Between Theoretical and Empirical Distribution								
(Averaged over 10 Runs)								
							No.	No.
Number of Bidders	2	3	4	5	6	7	Auctions	Observ.
Number of Auctions - Sample 1		15	29	4			48	181
1. Drop Small Samples		0.200	0.237					
		(0.084)	(0.064)					
2. Sample for Small Sample Estimates		0.198	0.238					
		(0.086)	(0.067)					
3. Sample for all Estimates - starting with Largest Sub Sample		0.205	0.236					
		(0.034)	(0.062)					
4. Sample for all Estimates - starting with Largest Possible Number of Bidders		0.218	0.242					
		(0.043)	(0.065)					
Rank Test (p-value)								
H0: 1 = 2		0.94	0.82					
H0: 2 = 3		0.88	0.94					
H0: 3 = 4		0.26	0.94					
Number of Auctions - Sample 2		31	59	8			98	369
1. Drop Small Samples		0.218	0.234	0.210				
		(0.052)	(0.038)	(0.072)				
2. Sample for Small Sample Estimates		0.213	0.236	0.221				
		(0.047)	(0.041)	(0.089)				
3. Sample for all Estimates - starting with Largest Sub Sample		0.214	0.228					
		(0.035)	(0.041)					
4. Sample for all Estimates - starting with Largest Possible Number of Bidders		0.243	0.254	0.215				
		(0.087)	(0.101)	(0.075)				
Rank Test (p-value)								
H0: 1 = 2		0.88	0.94	1.00				
H0: 2 = 3		0.82	0.71					
H0: 3 = 4		0.88	1.00					
Number of Auctions - Sample 3		61	118	15	1		195	736
1. Drop Small Samples		0.203	0.197	0.254				
		(0.027)	(0.043)	(0.082)				
2. Sample for Small Sample Estimates		0.204	0.179	0.250				
		(0.038)	(0.038)	(0.085)				
3. Sample for all Estimates - starting with Largest Sub Sample		0.174	0.199					
		(0.037)	(0.053)					

Table 2: APV Results (cont.)								
Maximum Distance Between Theoretical and Empirical Distribution								
(Averaged over 10 Runs)								
							No.	No.
Number of Bidders	2	3	4	5	6	7	Auctions	Observ.
Number of Auctions - Sample 4	2	19	16	8	3	1	49	190
1. Drop Small Samples		0.211	0.226	0.171				
		(0.058)	(0.078)	(0.069)				
2. Sample for Small Sample Estimates	0.099	0.212	0.226	0.173				
	(0.028)	(0.058)	(0.076)	(0.088)				
3. Sample for all Estimates - starting with Largest Sub Sample	0.181	0.254	0.234					
	(0.041)	(0.040)	(0.095)					
4. Sample for all Estimates - starting with Largest Possible Number of Bidders	0.159	0.199	0.211	0.164				
	(0.035)	(0.056)	(0.045)	(0.077)				
Rank Test (p-value)								
H0: 1 = 2		0.88	0.88	0.88				
H0: 2 = 3	0.00	0.10	0.94					
H0: 3 = 4	0.23	0.03	0.76					
Number of Auctions - Sample 5	4	37	32	16	6	2	97	377
1. Drop Small Samples		0.223	0.203	0.197				
		(0.070)	(0.063)	(0.064)				
2. Sample for Small Sample Estimates	0.093	0.219	0.206	0.205				
	(0.029)	(0.073)	(0.054)	(0.070)				
3. Sample for all Estimates - starting with Largest Sub Sample	0.195	0.233	0.200					
	(0.046)	(0.047)	(0.053)					
4. Sample for all Estimates - starting with Largest Possible Number of Bidders	0.198	0.223	0.222	0.198				
	(0.045)	(0.053)	(0.064)	(0.070)				
Rank Test (p-value)								
H0: 1 = 2		0.94	0.60	0.94				
H0: 2 = 3	0.00	0.41	0.55					
H0: 3 = 4	0.88	0.60	0.71					
Number of Auctions - Sample 6	9	74	65	31	12	5	196	762
1. Drop Small Samples		0.223	0.221	0.240	0.249			
		(0.036)	(0.044)	(0.070)	(0.078)			
2. Sample for Small Sample Estimates	0.086	0.219	0.218	0.240	0.252			
	(0.019)	(0.033)	(0.046)	(0.068)	(0.074)			
3. Sample for all Estimates - starting with Largest Sub Sample	0.151	0.231	0.219					
	(0.031)	(0.034)	(0.038)					

Table 2: APV Results (cont.)								
Maximum Distance Between Theoretical and Empirical Distribution								
(Averaged over 10 Runs)								
							No.	No.
Number of Bidders	2	3	4	5	6	7	Auctions	Observ.
Number of Auctions - Sample 7	8	15	11	7	4	2	47	178
1. Drop Small Samples		0.242	0.219					
		(0.111)	(0.073)					
2. Sample for Small Sample Estimates	0.130	0.235	0.217					
	(0.048)	(0.100)	(0.072)					
3. Sample for all Estimates - starting with Largest Sub Sample	0.188	0.244						
	(0.087)	(0.117)						
4. Sample for all Estimates - starting with Largest Possible Number of Bidders	0.224	0.238	0.212					
	(0.060)	(0.069)	(0.064)					
Rank Test (p-value)								
H0: 1 = 2		0.88	0.94					
H0: 2 = 3	0.05	0.94						
H0: 3 = 4	0.17	0.88						
Number of Auctions - Sample 8	17	29	22	14	8	5	95	362
1. Drop Small Samples		0.216	0.193	0.270	0.231			
		(0.046)	(0.068)	(0.107)	(0.085)			
2. Sample for Small Sample Estimates	0.102	0.218	0.196	0.282	0.237			
	(0.032)	(0.064)	(0.067)	(0.107)	(0.090)			
3. Sample for all Estimates - starting with Largest Sub Sample	0.214	0.220	0.200					
	(0.062)	(0.060)	(0.075)					
4. Sample for all Estimates - starting with Largest Possible Number of Bidders	0.202	0.214	0.226	0.241	0.232			
	(0.042)	(0.050)	(0.091)	(0.096)	(0.101)			
Rank Test (p-value)								
H0: 1 = 2		0.76	1.00	0.76	0.82			
H0: 2 = 3	0.00	0.82	0.76					
H0: 3 = 4	0.76	0.71	0.45					
Number of Auctions - Sample 9	33	58	44	27	16	9	187	710
1. Drop Small Samples	0.212	0.218	0.203	0.218	0.190	0.198		
	(0.032)	(0.026)	(0.047)	(0.062)	(0.066)	(0.073)		
2. Sample for Small Sample Estimates	0.212	0.221	0.201	0.212	0.187	0.213		
	(0.040)	(0.047)	(0.048)	(0.050)	(0.070)	(0.079)		
3. Sample for all Estimates - starting with Largest Sub Sample	0.214	0.233	0.209					
	(0.039)	(0.024)	(0.042)					

Figure 3: APV Estimation