Auction Design Using Bayesian Methods*

Dong-Hyuk Kim†

November 22, 2009

Abstract

For choosing a reserve price, the previous literature estimates parameter values for a structural auction model and then uses the point estimates to infer the optimal reserve price. We find, however, that since the seller's payoff (expected revenue) function is generally not symmetric about the optimal reserve price, the seller can obtain a larger payoff by incorporating the payoff structure and parameter uncertainty into the decision procedure. To see this, consider a seller whose payoff increases slowly up to the optimal price, but then drops sharply thereafter. Then, the seller should avoid overestimation more than underestimation systematically taking into account the sampling error. For this purpose, we propose the Bayesian decision method that maximizes the seller's predictive payoff for a given sample by formally considering the payoff structure and parameter uncertainty. Monte Carlo experiments show that this Bayes rule is especially useful when the payoff is fairly asymmetric and there is a large amount of uncertainty (due to either small samples or flexible specifications).

1 Introduction

We introduce a Bayesian decision theoretic method to choose a reserve price maximizing future revenue for an auction with private values. Suppose the bidders' valuations for the auctioned item are independently distributed as the valuation distribution $F(\cdot|\theta^*)$ with density $f(\cdot|\theta^*)$ and that the seller's valuation is v_0. Riley and Samuelson (1981) show that a reserve price ρ^* solving $\rho^* = v_0 + \frac{1-F(\rho^*|\theta^*)}{f(\rho^*|\theta^*)}$ maximizes the seller's expected revenue. Let $\rho_R(\theta^*) := \rho^*$. Let $\widehat{\theta}$ be a consistent estimator for θ^*.

*I am greatly indebted to Keisuke Hirano, my advisor, for his guidance, patience, and encouragement.

†Department of Economics, University of Arizona. Email: donghyuk@email.arizona.edu

Paarsch (1997) proposes to choose $\rho_R(\theta^*)$ to approximate $\rho_R(\theta^*)$. This decision process is called a 'plug-in' rule.[1]

However, the plug-in rule is not an optimal decision, since it does not consider the payoff (or, expected revenue) structure and parameter uncertainty. Consider a seller whose payoff increases slowly up to $\rho_R(\theta^*)$, but then drops sharply thereafter. For a given sample, it must be either $\rho_R(\widehat{\theta}) > \rho_R(\theta^*)$ or $\rho_R(\widehat{\theta}) < \rho_R(\theta^*)$ due to sampling error. Then, the seller should prefer underestimation to overestimation. Thus, we can make a higher payoff by formally considering the payoff structure as well as parameter uncertainty (that is related to sampling error).

We propose the usage of a Bayes rule that maximizes the predictive payoff (posterior mean of payoff). Savage (1954) and Anscombe and Aumann (1963) show that a rational decision maker acts in this way. This approach coherently incorporates both parameter uncertainty and the payoff structure into the decision procedure. Moreover, it is also shown to be optimal under the Average (Bayes) risk principle.

A Monte Carlo study shows that our method produces higher payoff than the plug-in rule when the payoff is fairly asymmetric. This advantage appears to be greater for small samples. Moreover, when a more flexible model is employed, this advantage still holds even for fairly large samples, because there is more parameter uncertainty. This suggests that the Bayesian method may be especially beneficial when a nonparametric model is employed. Note that the Monte Carlo experiment is a frequentist evaluation of the preference of the decision rule.

2 Decision Theoretic Approach

2.1 Seller's Problem with Asymmetric Payoff

The seller has sold homogeneous items via auctions indexed by $t = 1, \ldots, T$. At auction t, each bidder $i = 1, \ldots, N_t$ bids $b_{i,t}$ after observing his valuation $v_{i,t}$ without knowing his rivals' valuations. The bidder with the highest bid obtains the item at the price equal to the second highest bid. We assume that $\{v_{i,t}\}$ are independent and identically distributed, drawn from some continuous distribution $F(\cdot|\theta^*)$, and that the seller does not know θ^* but may have prior beliefs $P(\theta)$ over the parameter

[1] Li, Perrigne, and Vuong (2003) propose a plug-in rule to choose a reserve price maximizing future revenue for auctions with affiliated private values (APV).

space Θ.[2] Let $z := (b_{1,t}, \ldots, b_{N_t,t})_{t=1}^T$ be the sample the seller observes, $M := \sum_{t=1}^T N_t$ be the sample size, and $N := N_{T+1}$ be the number of bidders of the future auction. Lastly, we assume that every bidder follows the unique symmetric equilibrium strategy, $\beta(v_{i,t}|\theta^*) = v_{i,t}$.[3] So, $b_{i,t} = v_{i,t}$ for all i and t.[4]

The seller wants to extract the largest expected revenue from the auction $T+1$ by choosing a reserve price $\rho \in \mathcal{A}$, the feasible action set. As noted in the previous section, if M is small and the payoff is asymmetric, the seller may get higher revenue by considering the payoff structure and parameter uncertainty. Here, we claim that payoff structure is not generally symmetric about the revenue maximizing reserve price and the degree of asymmetry depends on the number of bidders by providing some examples.

First, we consider the exponential distribution with arrival rate θ, for which the revenue maximizing reserve price is $\rho^* = 1/\theta^*$. We set $\theta^* = 1$. This distribution appears on the left panel of Figure 1. The right panel plots the payoff functions for $N = 3, 4, 5$. For these values of N, the payoff structure turns out to be asymmetric around ρ^*, and it gets more asymmetric as N increases. From this, we see that the payoff structure depends on N and is not generally symmetric.[5] Note that in the exponential distribution case, this asymmetry result is identical for all θ, since the density functions have the same shape with different scales.

Second, we try Beta distributions that have various shapes of densities depending on the parameter (α, β). The first row of Figure 2 represents the density function and payoff functions for $(\alpha^*, \beta^*) = (1, 3)$, for which $\rho^* = 0.25$. Similarly to the exponential case, the payoff functions are asymmetric, and in this case their structure depends on N. In addition, the density function of the Beta(2,2) distribution and its payoff functions are plotted on the second row of Figure 2. Note that the Beta(2,2) distribution puts more probability mass on higher values than the Beta(1,3) distribution. As a result, its revenue maximizing reserve price is also higher, i.e. $\rho^* \approx 0.42$. The right panels of Figure 2 show that for given N its payoff structure is more asymmetric than Beta(1,3).

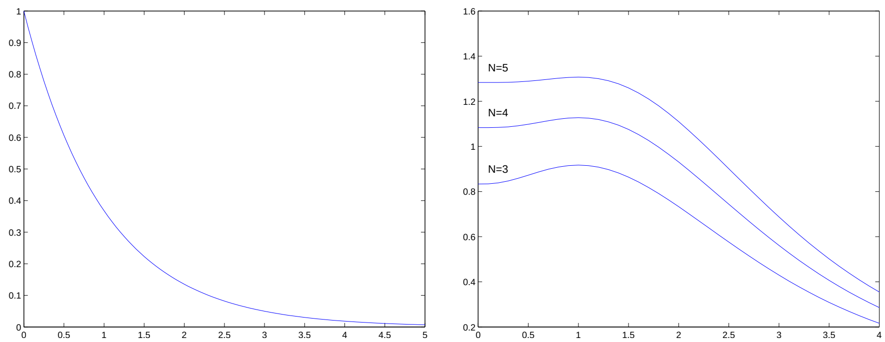
[2] This implies that the seller knows which parametric family the valuation distribution belongs to. So, we assume that the model is correctly specified.

[3] See Milgrom and Weber (1982).

[4] Though $\beta(v_{i,t}|\theta^*) = v_{i,t}$ is a weakly dominant strategy as long as the valuations are private information, we maintain the IID assumption since it makes the likelihood simple.

[5] However, we cannot say anything about the relationship between N and the degree of payoff asymmetry, since the payoff could become symmetric for large N.

Figure 1: Density Function and Asymmetry of Payoff: 1. Exponential(1)



In these examples, the payoff structures increase gradually up to the maximum point and decrease sharply, and the degree of asymmetry depends on the number of bidders. It turns out that this phenomenon is quite general, since when we examine many other parametric distributions, we have similar asymmetry patterns as in Figure 1 and 2.[6]

2.2 Bayes Action as an Optimal Decision Rule

We propose a Bayesian decision theoretic method for the auction design problem. Savage (1954) and Anscombe and Aumann (1963) show that a rational decision maker with a preference relation satisfying some axioms selects an action maximizing the posterior mean of the payoff. Such an action is called the Bayes action. Let $P(\theta|z)$ be the posterior and $\Pi(\theta, \rho, N)$ be the payoff when θ is the true parameter, ρ is the reserve price, and N is the number of bidders. Then, the Bayes action is given by

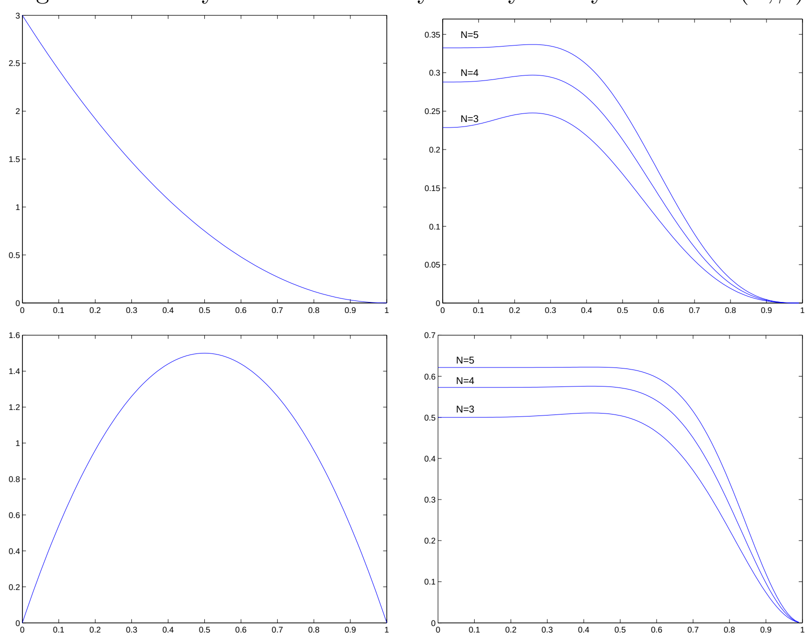
$$\rho_B(z, N) := \arg\max_{\rho \in \mathcal{A}} \int_{\Theta} \Pi(\theta, \rho, N) dP(\theta|z) \tag{1}$$

Note that the posterior distribution formally quantifies all parameter uncertainty remaining after considering the sample information. We also see that (1) also systematically takes into account the payoff structure determined by N.

We can compute each component of (1). The seller knows which parametric family the valuation

[6]We examine the beta, gamma, Pareto, and Weibull distribution with various parameter values. They are all similar to Figure 1 and 2.

Figure 2: Density Function and Asymmetry of Payoff: 2. Beta(α, β)



distribution belongs to. Then, under the IPV paradigm, the payoff function is given by

$$\Pi(\theta, \rho, N) = N \left\{ \rho(1 - F(\rho|\theta))F(\rho|\theta)^{N-1} + \int_{\rho}^{\bar{v}} y(N-1)(1-F(y|\theta))F(y|\theta)^{N-2} f(y|\theta)dy \right\} + F(\rho|\theta)^N v_0 \quad (2)$$

as shown by Milgrom and Weber (1982). In addition, the observed bids are distributed identically to valuations since $b_{i,t} = v_{i,t}$ for every i and t. Therefore, the individual likelihood function is simply $f(b_{i,t}|\theta)$. Then, the posterior can be expressed as

$$P(\theta|z) \propto P(\theta) \prod_{t=1}^{T} \prod_{i=1}^{N_t} f(b_{i,t}|\theta) \quad (3)$$

Note that it is enough to compute the posterior up to the normalizing constant for our purpose, because multiplying the right hand side of (1) by some constant does not change the solution.

$\rho_B(z, N)$ is optimal with respect to the Average (Bayes) risk principle.[7] Let $d : \mathcal{Z} \to \mathcal{A}$ be a decision rule where $\mathcal{Z}$ is the sample space of data. Recall that the risk is given by $R(\theta, d) := E_\theta[L(\theta, d(Z))]$ for the loss $L := -\Pi$.[8] Then, the Bayes risk is $r(P, d) := \int_\Theta R(\theta, d)dP(\theta)$ for the prior distribution P.

[7]This is one of the most widely used frequentist decision principles. See Berger (1985).

[8]We suppress the dependence of the loss on N.

According to the Bayes risk principle the decision rule d^P minimizing $r(P,d)$ is said to be optimal. d^P is called a Bayes rule.[9] To see the equivalence between the Bayes rule and the Bayes action, we put the Bayes risk as follows, assuming the densities are absolutely continuous with respect to Lebesgue measure,

$$\begin{aligned} r(P,d) &= \int_\Theta R(\theta,d)p(\theta)d\theta \\ &= \int_\Theta \left\{ \int_\mathcal{Z} L(\theta,d(z))f(z|\theta)dz \right\} p(\theta)d\theta \\ &= \int_\mathcal{Z} \int_\Theta L(\theta,d(z))p(\theta)f(z|\theta)d\theta dz \\ &= \int_\mathcal{Z} \left\{ \int_\Theta L(\theta,d(z))cp(\theta|z)d\theta \right\} dz \end{aligned}$$

where c is the normalizing parameter and $p(\theta|z)$ is the posterior density. In order to minimize the Bayes risk r we can minimize the inner integral, the posterior mean of L for each $z \in \mathcal{Z}$. Hence, the optimality of the Bayes action follows. Furthermore, the Bayes rule is known to be admissible. Our decision rule $\rho_B(z,N)$ generally gives different answers than the plug-in rule $\rho_R(\hat{\theta})$. When $\Pi(\theta,\rho,N)$ is nonlinear in θ, $\Pi(E[\theta|z],\rho,N) \neq E[\Pi(\theta,\rho,N)|z]$ in general. Let $\hat{\theta} := E[\theta|z]$.[10] Then, the plug in method $\rho_R(\hat{\theta})$ maximizes the left hand side. Therefore, $\rho_B(z,N) \neq \rho_R(\hat{\theta})$.

The decision theoretic optimality concepts we employ in this paper must be distinguished from the ones commonly used in the auction literature. The decision problem is nontrivial when the state of nature θ^* is uncertain. A decision rule is said to be optimal under a decision principle if its way to pick an action dealing with the uncertainty meets the requirements of the principle. $\rho_R(\theta^*)$ is usually called the optimal reserve price. However, it is not even a decision rule we can employ since θ^* is unknown. Therefore, we do not follow the convention to call it optimal reserve price. In addition, when the plug-in method is implemented, an optimal estimation procedure for θ can be employed. For example, Li, Perrigne, and Vuong (2003) optimally estimated $\rho_R(\hat{\theta})$ with respect to mean squared error. This is the optimality of the nonparametric estimation procedure but not the optimality with respect to payoff.

[9]Notice that the Bayes rule is conceptually different from the Bayes action, because the former considers all possible data realizations while the latter depends only on the realized data.

[10]See Section 3.1 for more discussion of $\hat{\theta}$.

3 Monte Carlo

3.1 Implementation and Monte Carlo Design

Generally, the integral of (1) does not have a closed form expression. Therefore, once the posterior distribution is computed, we estimate the posterior mean of payoff using simulation. Let

$$\hat{\rho}_B(z, N) = \arg\max_{\rho \in [0,1]} \frac{1}{S} \sum_{s=1}^{S} \Pi(\theta^s, \rho, N) \tag{4}$$

where $\{\theta^s\}_{s=1}^S$ are random draws from the posterior distribution for large S.

For the plug-in method we adopt

$$\hat{\theta}_B := \arg\min_{a \in \Theta} \int (\theta - a)^2 dP(\theta|z) \tag{5}$$

as the estimator for the true parameter θ^*. This estimator is a Bayes action that minimizes squared error loss. It turns out that $\hat{\theta}_B = E[\theta|z]$, the posterior mean of θ. Note that $\hat{\theta}_B$ is asymptotically equivalent to the maximum likelihood estimator and therefore it is consistent and asymptotically efficient under conventional regularity conditions.[11] Then, the plug-in rule chooses the reserve price $\rho_R(\hat{\theta}_B) = \hat{\rho}$ such that $\hat{\rho} = \frac{1-F(\hat{\rho}|\hat{\theta}_B)}{f(\hat{\rho}|\hat{\theta}_B)}$

For the Monte Carlo experiment, we simulate a sample of size M from $F(\cdot|\theta^*)$. Then, we compute $\rho_B(z, N)$ and $\rho_R(\hat{\theta}_B)$ and evaluate the true payoffs under two different decision frameworks. i.e. $\Pi(\theta^*, \rho_B(z, N), N)$ and $\Pi(\theta^*, \rho_R(\hat{\theta}_B), N)$. We repeat this procedure one thousand times to compare the overall performances of the two different methods for a fixed data generating process. We emphasize that this is a frequentist evaluation of the preference of the decision rule. On the other hand, a Bayesian is primarily interested in the performance of decision rule for the actually observed data and the Bayes action should be chosen under the conditional Bayes principle.[12]

3.2 Simple Parametric Model

We take the exponential distribution with $\theta^* = 1$ for the true valuation distribution. We employ the Gamma distribution with $(\alpha, \beta) = (1, 1)$ as the prior so that the prior mean equals θ^*. The advantage of using a Gamma prior is that it is conjugate for the parameter of an exponential distribution.

[11]See Theorem 8.3 of Lehmann and Casella (1998).

[12]See Berger (1985) Chapter 1 and 4 for details.

Therefore, it is convenient to estimate $\hat{\theta} = E[\theta|z]$ and to implement the plug-in method. Moreover, it is easy to generate $\{\theta^s\}_{s=1}^S$ for the Bayes action (4).

Table 1 presents the results of the Monte Carlo study. Since θ is only one dimensional, there is less parameter uncertainty than more complicated model for given sample information. This makes the difference between two decision methods disappear quickly as sample size grows. For this reason, to see their different behaviors, we try only small samples, $M = 10$ and 20. As for the number of bidders in the future auction, we try $N = 3, 4, 5$ for which the different payoff structures are plotted in Figure 1.

Column (1) and (2) provides the sample means and sample standard deviations of the reserve prices chosen by the Bayesian method and the plug-in method, respectively, for different M and N. We see that for each M and N, the Bayesian method selects smaller reserve prices than the plug-in method. This is because the Bayesian method takes into account the fact that overestimation results in greater loss than underestimation due to the asymmetric payoff structure. We also observe that for each M the Bayesian method chooses smaller reserve prices for larger N. Recall that among these N's larger N induces more asymmetric payoff and overestimation is worse. Hence, the Bayesian method selects smaller reserve prices to maximize the future payoff. In contrast, the plug-in method chooses the reserve price only depending on the estimate $\hat{\theta}_B$. So, it does not choose different prices for different payoff structures. As a result, the former produces higher payoff than the latter on average. This is shown in column (3), which is the sample mean of the percentage gain of the Bayesian method over the plug-in method in terms of payoff.

In addition, we find that, when $M = 20$, both methods select the reserve prices closer to $\rho^* = 1$ on average with smaller sample standard deviations than when $M = 10$ and therefore the percentage gain for $M = 20$ is smaller.

For the plug-in method, this result is expected, because it is already known to be consistent. Though we do not make a rigorous statement, one can show that the Bayesian method is also consistent. Intuitively, if there were no parameter uncertainty and the posterior were degenerate at θ^*, both methods would give identical answers. Since we have less parameter uncertainty for larger samples, the discrepancy between the two methods vanishes as sample size grows. Therefore, it is also natural for $\rho_B(z, N)$ to approach ρ^*.

In this Monte Carlo study, the performance of the Bayesian decision method does not seem significantly better than the plug-in method even for quite small sample size, since the percentage gains

Table 1: Monte Carlo Results for Simple Parametric Model

Sample Size	# of bidders	$\rho_B(z, N)$	$\rho_R(\hat{\theta}_B)$	% gain over plug-in method
M	N	(1)	(2)	(3)
10	3	0.9961 (0.3085)	1.0979 (0.3392)	0.7081
	4	0.9405 (0.2911)	”	0.9013
	5	0.8850 (0.2730)	”	0.9736
20	3	0.9875 (0.2117)	1.0412 (0.2229)	0.1648
	4	0.9594 (0.2056)	”	0.2052
	5	0.9312 (0.1994)	”	0.2184

Note: The true revenue maximizing reserve price is given by $\rho^* = 1$

in column (3) for the case where $M = 10$ are less than one percent. The reason for this is that the model is so simple that the sample size ten is already large enough to get rid of much of the parameter uncertainty. Hence, the plug-in method performs quite well even for $M = 10$. However, if a more flexible model is employed, the Bayesian decision method would perform much better than the plug-in method for the same sample size. To see this, we repeat the Monte Carlo study with the model with higher dimensional parameter.

3.3 Flexible Model

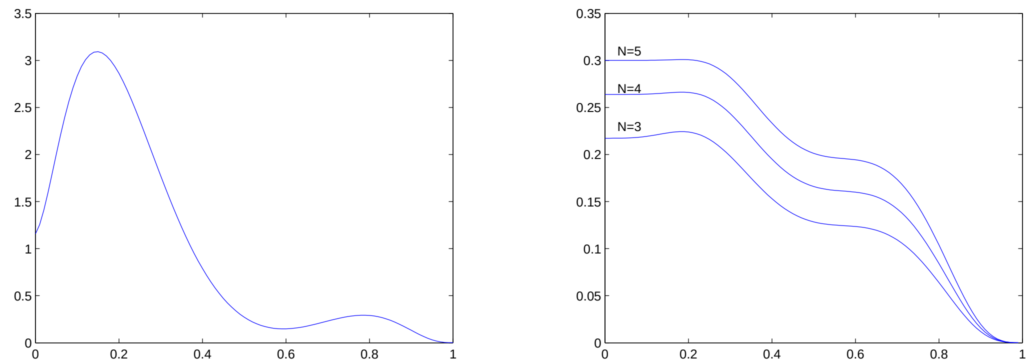
We assume that the density function of the valuation distribution has the form of the Bernstein density of Petrone (1999a,b) as follows

$$f(v|\theta) = \sum_{j=1}^{k} \theta_j \text{beta}(v|j, k-j+1) \tag{6}$$

where beta$(v|a, b)$ is the Beta density with parameter a and b and $\theta \in \Theta := \left\{\theta \in \Re_+^k \,|\, \sum_{j=1}^{k} \theta_k = 1\right\}$. Petrone (1999a, b) shows that if the prior for $(k, \theta_1, \ldots, \theta_k)$ has full support, the analysis is nonparametric in the sense that this prior has large support with respect to the weak convergence topology of the set of all distribution functions on $[0, 1]$.

For the Monte Carlo study, we fix $k = 15$ and employ the uniform prior over Θ which is now the

Figure 3: Density Function and Payoff Structures: Bernstein Density



fourteen dimensional simplex. Though this model is not fully nonparametric, it is flexible enough to exhibit fairly different results than the simple parametric model in the previous subsection.[13] We assume that the valuation distribution has the density function plotted on the left panel of Figure 3. Then, the payoff structures for $N = 3, 4, 5$ are appeared on the right panel of Figure 3. Similarly to the exponential distribution, the payoff structure is more asymmetric for larger N for these N's. We try the sample size $M = 20, 100$ to make a comparison with the simple parametric case in the previous subsection.

Table 2 summarizes the Monte Carlo results for the flexible model. The general features are similar to the simple parametric case in the previous subsection: the Bayesian method chooses smaller reserve prices for larger N, since it accounts for the different payoff asymmetry caused by different N. On average, the Bayesian method produces higher payoff than the plug-in method. As the sample size grows, both methods get close to the true revenue maximizing reserve price.

However, what distinguishes these results from the ones in Table 1 is the amount of parameter uncertainty. We compare the percentage gains of the Bayesian method over the plug-in method for the case of $M = 20$. Column (3) of Table 1 shows that the gain of the Bayesian method is about 0.2 percent when one dimensional model is employed. However, we see that the gain can be higher than five percent when the flexible one is used from Table 2. In addition, the gains of sample size $M = 20$ for the simple model roughly amounts to the ones of sample size $M = 100$ for the flexible model.

[13] Appendix B discusses the dependence of the flexibility of the Bernstein density on k.

Table 2: Monte Carlo Results for the Flexible Model

Sample Size M	# of bidders N	$\rho_B(z, N)$ (1)	$\rho_R(\hat{\theta}_B)$ (2)	% gain over plug-in method (3)
20	3	0.3686 (0.0521)	0.3846 (0.0517)	3.9583
	4	0.3610 (0.0516)	"	5.1035
	5	0.3538 (0.0510)	"	5.7830
100	3	0.2182 (0.0196)	0.2231 (0.0194)	0.3031
	4	0.2173 (0.0193)	"	0.2404
	5	0.2160 (0.0192)	"	0.1891

Note: The true revenue maximizing reserve price is given by $\rho^* = 0.18375$

Therefore, the Bayesian method is beneficial for wider range of sample sizes for the flexible models. Notice that the fourteen dimensional model is still fairly restrictive. Since auction theory provides only little information on the underlying distribution, nonparametric models may be preferred to avoid misspecification error. Then, the Bayesian method may be especially beneficial, since the parameter is now infinite dimensional.

We end this section by remarking on the posterior approximation method we employ for the flexible model. Since the posterior distribution does not have a closed form expression, we have to approximate it using a Markov Chain Monte Carlo (MCMC) simulation method. If we employed the nonparametric model Petrone (1999 a,b) develop, the algorithm she suggests in her papers would be a natural choice. However, since we fix k for the Monte Carlo study above, we consider the Metropolis-Hastings algorithm as a reasonable choice, because it is much faster than Petrone's algorithm and hence more desirable for the Monte Carlo study where the MCMC should be implemented repeatedly many times. Note that since θ can be seen as a probability distribution over k distinct points, each candidate of θ should satisfy the implied restrictions.[14] Hence, we incorporate the Metropolis-Hastings algorithm with the GHK sampler so that the proposal function generates only such θ's.[15] We explain the modified Metropolis-Hastings algorithm in Appendix A.2 in detail.

[14] All elements in θ are positive and sum up to one.

[15] For the GHK sampler, see Hajivassiliou and McFadden (1990) and Keane (1990).

4 Conclusion

We introduce a Bayesian decision theoretic method to choose the reserve price that maximizes the seller's payoff from the future auction. We show that the payoff is not generally symmetric about the true revenue maximizing reserve price. Then, when we have parameter uncertainty due to small sample size, we may earn higher payoff by considering the payoff structure and parameter uncertainty. The Bayesian method we propose formally incorporates these elements into the decision making procedure. On the other hand, the plug-in method computes the revenue maximizing reserve price regarding the estimated parameters as the true ones. Therefore, it considers neither the payoff structure nor parameter uncertainty. As a result, the Bayesian method can produce higher payoff than the plug-in method. This is supported by the Monte Carlo evidence.

The Monte Carlo study with the simple parametric model shows that the Bayesian method produces higher payoff than the plug-in method by choosing downwardly biased reserve prices, because the payoff structure given in the experiment implies that overestimation results in more loss than underestimation. The difference between two methods is highlighted when sample is small. However, it becomes negligible as sample size grows, since much of parameter uncertainty is removed by the sample information. The Monte Carlo results show that the plug-in method becomes a good approximation for the decision by the Bayesian method even for quite small samples, because the amount of parameter uncertainty is small when a simple parametric model is employed.

However, from the Monte Carlo study with the more flexible model, we find that for the same sample size the difference between two methods is larger when the complicated model is adopted than the simple model and also that the Bayesian method produces significantly higher payoff for wider range of sample sizes. This suggests that if a nonparametric model is employed, the usage of the Bayesian method may be beneficial even when the sample is fairly large. This is important because the auction theory is silent on the choice of parametric family and one might want to use the fully flexible model to avoid any misspecification error.

Appendix A: Metropolis Hastings Algorithm

We explain the Metropolis-Hastings algorithm that we employ for the approximation of the posterior. Let $q(\theta^{s-1}, \tilde{\theta})$ be a proposal density function to generate the candidate $\tilde{\theta}$ for the next random draw given the current draw θ^{s-1}. At the s-th step of the algorithm, we generate θ^s using the proposal function q, and calculate the acceptance rate

$$\alpha = \min\left\{\frac{p(\tilde{\theta}|z)}{p(\theta^{s-1}|z)}\frac{q(\tilde{\theta},\theta^{s-1})}{q(\theta^{s-1},\tilde{\theta})}, 1\right\} \tag{7}$$

provided that the denominator is nonzero. If it happens to be zero, we set $\alpha = 1$. Then, we let

$$\theta^s := \begin{cases} \tilde{\theta} & \text{with probability } \alpha \\ \theta^{s-1} & \text{with probability } 1-\alpha \end{cases} \tag{8}$$

The sequence of $\{\theta^s\}_{s=1}^S$ has a distribution that converges to the posterior distribution under general conditions as the number of simulation steps S goes to infinity.

Appendix A.1: proposal function ratio for parametric model

To run the Metropolis-Hastings algorithm, we need to specify q. For the simple parametric model, we generate the parameters for Beta distribution. So, $\Theta = \Re_+^2$. Let $\delta := \log\theta$. Then, we draw $\tilde{\delta}$ from $N(\delta^{s-1}, \sigma^2 I)$. Then, (7) is simplified since by symmetry of normal distribution and change of variable, it can be shown that

$$\frac{q(\tilde{\theta},\theta^{s-1})}{q(\theta^{s-1},\tilde{\theta})} = \frac{\tilde{\theta}_1\tilde{\theta}_2}{\theta_1^{s-1}\theta_2^{s-1}} \tag{9}$$

Appendix A.2: proposal function ratio for flexible model

For the flexible model (6), $\Theta = \left\{\theta \in \Re_+^k | \sum_{j=1}^k \theta_j = 1\right\}$. Then, the inequality restrictions on θ can be expressed as

$$B\theta \geq b \tag{10}$$

where $\theta = (\theta_1, \ldots, \theta_{k-1})'$, $b = (-1, 0, \ldots, 0)$, and

$$B = \begin{pmatrix} -1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \\ & & \cdots & & \\ & & & & 1 \end{pmatrix}_{(k-1)\times(k-1)}$$

provided that $\theta_1 \geq 0$. Then, this is equivalent to $\sum_{j=1}^{k-1} \theta_j \leq 1$ and $\theta_j \geq 0$ for $j = 1, \ldots, k-1$. Note that we let θ be the first $k-1$ elements of the parameter vector, since $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$.

Suppose that the currunt draw is θ^s and the candidate for the next step is determined by

$$\tilde{\theta} = \theta^s + u$$

where $u \sim N(0, \sigma^2 I_{k-1}) 1(B\tilde{\theta} \geq b)$ for the scale parameter $\sigma^2 > 0$. Hence, we need to be able to draw u from the truncated normal distribution.

Hajivassiliou and McFadden (1990) and Keane (1990) independently developed a method that draws a random vector from a truncated normal distribution, which is called the GHK sampler.[16] Furthermore, Geweke (1995) extended the GHK sampler to the situation where linear inequality contraints are imposed on the random vector. Using Geweke (1995), we can generate θ that satisfies the linear inequality constraints above.

Since $\tilde{\theta} = \theta^s + u$,

$$Bu \geq b^* := b - B\theta^s \tag{11}$$

Then,

$$Bu \sim N(0, \sigma_u^2 BB') 1(Bu \geq b^*) \tag{12}$$

Now, we apply the GHK sampler to (12). Let C be the Cholesky decomposition of $\sigma_u^2 BB'$. Then, it is easy to see that

$$Bu = C\varepsilon \text{ with } \varepsilon \sim N(0, I_{k-1}) 1(\varepsilon \geq \underline{\varepsilon}) \tag{13}$$

where for each $j = \{1, 2, \cdots, k-1\}$,

$$\underline{\varepsilon}_j = \frac{b_j^* - \sum_{i=1}^{j-1} c_{j,i} \varepsilon_i}{c_{j,j}} \tag{14}$$

where b_j^* is the jth element of b^* and $c_{i,j}$ is (i,j)th element of the matrix C. Then, the problem boils down to generating a random number ε_j from the truncated standard normal distribution with lower limit $\underline{\varepsilon}_j$ recursively from $j = 1$ to $k-1$[17]. Once we draw a random vector ε, the candidate $\tilde{\theta}$ is determined by

$$\tilde{\theta} = \theta^s + B^{-1} C\varepsilon \tag{15}$$

[16] See also Keane (1993), Keane (1994), and Geweke, Keane, and Runkle (1995).

[17] For $j = 1$, $\underline{\varepsilon}_1 = b_1^* / c_{1,1}$

Next, we need to evaluate $q(\tilde{\theta}, \theta^s)$. First, we know that

$$p(\varepsilon_j|\varepsilon_1, \cdots, \varepsilon_{j-1}) = \frac{\phi(\varepsilon_j)}{1 - \Phi(\underline{\varepsilon}_j)} \tag{16}$$

Then, the joint density function of the vector ε is given by

$$p(\varepsilon_1, \cdots, \varepsilon_p) = \prod_{j=1}^{p} \frac{\phi(\varepsilon_j)}{1 - \Phi(\underline{\varepsilon}_j)} \tag{17}$$

Since $\varepsilon = C^{-1}Bu$, we can evaluate the joint density function of u as follows,

$$p(u_1, \cdots, u_p) = \prod_{j=1}^{p} \frac{\phi(\varepsilon_j)}{1 - \Phi(\underline{\varepsilon}_j)} |C^{-1}B| \tag{18}$$

Then, since the candidate $\tilde{\theta} = \theta^{(s)} + u$,

$$q(\theta^{(s)}, \tilde{\theta}) = p(u_1, \cdots, u_p) \tag{19}$$

Finally, we need to evaluate $q(\tilde{\theta}, \theta^s)$ as follows. Note that $\theta^s = \tilde{\theta} - u$. Then, it turns out that $Bu \leq b^{**} := -(b - B\tilde{\theta})$. Since $u \sim N(0, \sigma^2 I_{k-1})1(Bu \leq b^{**})$, $Bu \sim N(0, \sigma^2 BB')1(Bu \leq b^{**})$. Then, $Bu = C\varepsilon$ for $\varepsilon \sim N(0, I_{k-1})1(\varepsilon \leq \bar{\varepsilon})$ where for each $j = \{1, 2, \cdots, k-1\}$, $\bar{\varepsilon}_j := \frac{b_j^{**} - \sum_{i=1}^{j-1} c_{j,i}\varepsilon_i}{c_{j,j}}$.

Then, the conditional density of ε_j is given by

$$p(\varepsilon_j|\varepsilon_1, \cdots, \varepsilon_{j-1}) = \frac{\phi(\varepsilon_j)}{\Phi(\bar{\varepsilon}_j)}$$

Then, the joint distribution of ε can be written as

$$p(\varepsilon_1, \cdots, \varepsilon_p) = \prod_{j=1}^{p} \frac{\phi(\varepsilon_j)}{\Phi(\bar{\varepsilon}_j)}$$

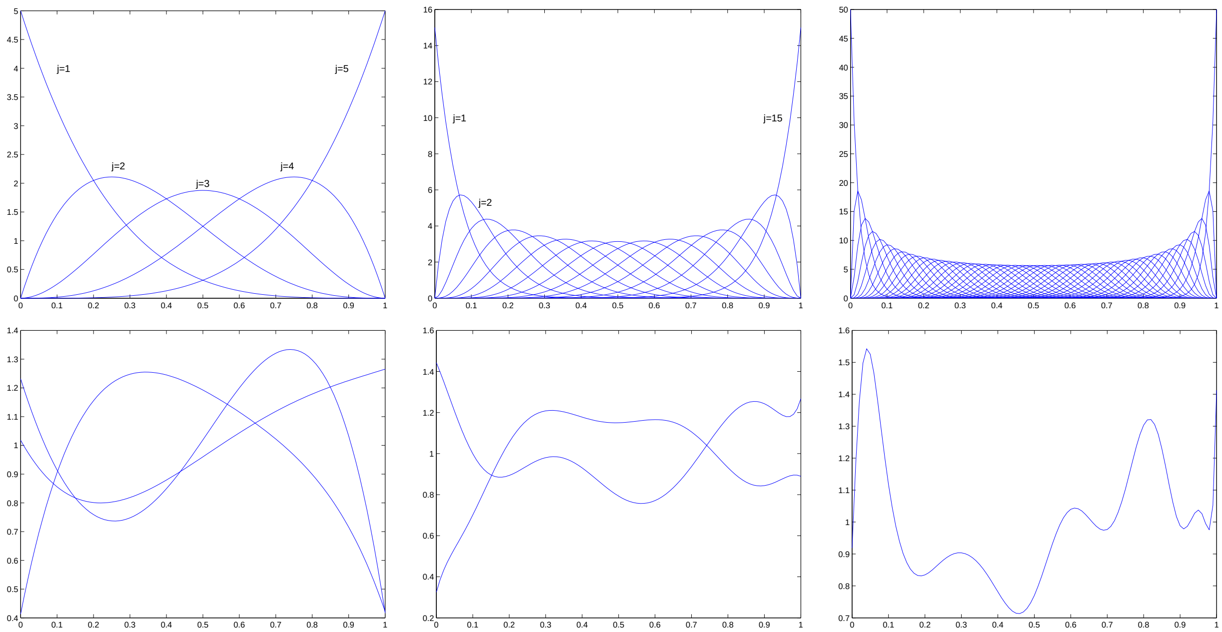
since $\varepsilon = C^{-1}Bu$,

$$p(u_1, \cdots, u_{k-1}) = \prod_{j=1}^{k-1} \frac{\phi(\varepsilon_j)}{\Phi(\bar{\varepsilon}_j)} |C^{-1}B| \tag{20}$$

that is actually the value of the proposal function $q(\tilde{\theta}, \theta^s)$. Therefore, deviding (20) by (19), we have

$$\frac{q(\tilde{\theta}, \theta^s)}{q(\theta^s, \tilde{\theta})} = \prod_{j=1}^{p} \frac{1 - \Phi(\underline{\varepsilon}_j)}{\Phi(\bar{\varepsilon}_j)} \tag{21}$$

Figure 4: Flexibility of Bernstein Densities



Appendix B. Bernstein Densities with different k

It might be useful to understand how flexible (6) is for different k's. For any $j \in \{1, \dots, k\}$ with a given k, the mode of the beta distribution with parameters j and $k - j + 1$ is $\frac{j-1}{k-1}$. Hence, (6) is a weighted average of beta densities for which the modes are the equally spaced k grid points $\left\{0, \frac{1}{k-1}, \dots, \frac{k-2}{k-1}, 1\right\}$ with the weights given by θ. It turns out that we can approximate any density function defined on the unit interval using (6) for some k. We describe the flexibility for different k visually in Figure 4. The first panels of Figure 4 illustrate the beta densities used for the Bernstein densities in (6) for $k = 5, 15$, and 50. The second panels are some examples of the Bernstein densities for each k.[18] Figure 4 provides a rough idea on the flexibility implied by different k's.

[18]To construct these examples, we draw θ from the uniform distribution over Θ for each k.

References

[1] Anscombe, F.J., and R. J. Aumann (1963), "A Definition of Subjective Probability," *Annals of Mathematical Statistics,* **34**, 199-205.

[2] Athey, S., and Haile, P. A. (2007), "Nonparametric Approaches to Auctions," *Handbook of Econometrics,*.**6A**, Elsevier.

[3] Athey, S., and Haile, P. A. (2006), "Empirical Models of Auctions," *Cowles Foundation Discussion Paper No. 1562.*

[4] Berger, J. O. (1985), "Statistical Decision Theory and Bayesian Analysis," Springer

[5] Geweke, J. (1995), "Bayesian Inference for Linear Models Subject to Linear Inequality Constraints" *Working Paper.*

[6] Geweke, J., Kean, M., and Runkle, D. (1995), "Recursively Simulating Multinomial Multiperiod Probit Probabilities", *Proceedings of the American Statistical Association, Business and Economic Statistics Section* 1–6

[7] Hajivassiliou, V., and McFadden, D. (1998), "The Method of Simulated Scores for the Estimation of LDV Models," *Econometrica,* **66**: 863 − 896.

[8] Keane, M. (1990), "Four Essays in Empirical Macro and Labor Economics" Ph.D. dissertation, Brown University.

[9] Keane, M. (1993), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica,* **62** 95–116.

[10] Keane, M. (1994), "Simulation Estimation for Panel Data Models with Limited Dependent Variables," In G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), handbook of Statistics vol. 11, 545 – 71. Amsterdam: Elsevier Science Publishers.

[11] Krishna, V. (2002) "Auction Theory," *Academic Press.*

[12] Laffont, J. and Voung, Q. (1996) "Structural Analysis of Auction Data," *American Economic Review, Papers and Proceedings,* **86**, : 414–420.

[13] Lehmann, E. L., Casella, G. (1998) "Theory of Point Estimation," *Springer, 2nd ed.*

[14] Li,T., Perrigne, I., and Vuong, Q., (2002) "Structural Estimation of the Affiliated Private Value Auction Model," *RAND Journal of Ecoomics,* **33**, : 171–193.

[15] Li,T., Perrigne, I., and Vuong, Q., (2003) "Semiparametric Estimation of the Optimal Reserve Price in First-Price Auctions," *Journal of Business & Economic Statistics,* **21**(1), : 53–64.

[16] Milgrom, P. R., and Weber, R. J. (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica,* **50**, 1089–1122.

[17] Newey, N. K., and McFadden, D. (1994) "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics,* **4**, 2111 – 2245

[18] Paarsch, H. J. (1997), "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbian Timber Sales," *Journal of Econometrics,* **78**, 333–357.

[19] Petrone, S. (1999a), "Random Bernstein Polynomials" *Scandinavian Journal of Statistics,* **26**, 373 – 393.

[20] Petrone, S. (1999b), "Bayesian Density Estimation Using Bernstein Polynomials" *Canadian Journal of Statistics,* **27**, 105 – 126.

[21] Riley, J., and Samuelson, W. (1981), "Optimal Auctions," *American Economic Review,* **71**, 381–392.

[22] Roberts, G. and Rosenthal, J. (2001), "Optimal Scaling for Various Metropolis-Hastings Algorithms," **Working paper**.

[23] Savage, L. J. (1954), "Foundations of Statistics," Wiley, New York, reissued in 1972 by Dover.