BAYESIAN ECONOMETRICS FOR AUCTION MODELS

by

Dong-Hyuk Kim

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2010
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Dong-Hyuk Kim entitled Bayesian Econometrics for Auction Models and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Keisuke Hirano  
Date: 15 June 2010

Gautam Gowrisankaran  
Date: 15 June 2010

Mo Xiao  
Date: 15 June 2010

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Keisuke Hirano  
Date: 15 June 2010
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED:  Dong-Hyuk Kim
I was very fortunate and honored to have an exciting intellectual journey with Prof. Keisuke Hirano through all my graduate career. His guidance, encouragement, and patience made this dissertation, my first step as an economist, possible. I would like to express my deepest gratitude to Prof. Hirano. He is a great role model as a researcher and an advisor. I wish to have his attitude towards the study of economics, and hope to be such a thoughtful mentor for my future students.

I would like to extend equal appreciation to my other committee members. Prof. Gautam Gowrisankaran observed the potential of the second chapter in my dissertation, when it was still preliminary, and encouraged me to develop it to a job market paper. Prof. Mo Xiao gave me a lot of useful advice on forming research ideas and putting the findings in the format of an academic paper. Moreover, their suggestions and questions lead me to strengthen the I.O aspects of my dissertation.

I am grateful to Prof. John Wooders for many useful comments on the theoretical setup for the auction mechanisms and questions about institutional aspects of the OCS wildcat sales. I also thank Prof. Jonah Gelbach for many advice on the econometric methodologies and interpretation of data and empirical findings. In addition, I thank my colleagues, Keoka Grayson, Jenny Hawkins, and Dr. Mary Lopez. They never hesitated to spend time on proofreading the draft.

I would not forget the help from the department staff, Lana Sooter, Michelle Piontek, and Tony Stevens. They were extremely diligent in providing all administrative support. The department of economics at University of Arizona provided generous financial support, computing facilities, and office space.

I would like to thank my family. My parents allowed us to live far way from them in the twilight years of their lives. They have missed us, particularly their grandchildren, for so many days and nights. My brother, Dongjun, and his wife, Soomi, supported my parents financially and emotionally, which should have been my responsibility. They were all supporting and encouraging me with their best wishes.

I especially thank my wife, Youngeun, for cheering me up and standing there for me through the good times and bad. I also thank my children, Eugene, Aaron, and Jason for giving me such a great joy. I regret that I did not spend as much time with them as I wished. But, I will be a really good daddy. I promise.
DEDICATION

To my wife, Youngeun,

and

three children, Eugene, Aaron, and Jason

with all my love.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. 8

LIST OF TABLES ................................................................. 9

ABSTRACT ................................................................. 10

CHAPTER 1 AUCTION DESIGN USING BAYESIAN METHODS .......... 12
  1.1 Introduction .......................................................... 12
  1.2 Decision Theoretic Approach ....................................... 14
    1.2.1 Seller’s Problem with Asymmetric Payoff .................. 14
    1.2.2 Bayes Action as an Optimal Decision Rule ................. 16
  1.3 Monte Carlo ....................................................... 18
    1.3.1 Implementation and Monte Carlo Design .................. 18
    1.3.2 Simple Parametric Model ................................... 19
    1.3.3 Flexible Model ............................................... 22
  1.4 Conclusion ......................................................... 24

CHAPTER 2 FLEXIBLE BAYESIAN ANALYSIS OF FIRST PRICE AUCTIONS USING SIMULATED LIKELIHOOD ........ 26
  2.1 Introduction ........................................................ 26
  2.2 Auction Models and Empirical Method ............................ 28
    2.2.1 First Price Auctions under the APV/IPV Paradigm ......... 28
    2.2.2 Bayes with Simulated Likelihood .......................... 29
    2.2.3 Valuation Density Specification ............................ 32
    2.2.4 Discussion ..................................................... 34
  2.3 Monte Carlo ....................................................... 36
  2.4 Estimation and Auction Design for OCS Wildcats ............... 40
    2.4.1 Data and Sample Space Discretization ...................... 40
    2.4.2 Valuation/Bid Density Estimation .......................... 42
    2.4.3 Auction Design with Bayes Rule ............................ 43
  2.5 Conclusion ......................................................... 46

CHAPTER 3 SIMPLE APPROXIMATIONS FOR BAYESIAN AUCTION DESIGN ................................................................. 48
  3.1 Introduction ........................................................ 48
  3.2 Auction Models and Bayesian Decision Framework ............... 49
### TABLE OF CONTENTS – Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Second Price Auctions: Regular Models</td>
<td>52</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Consistency</td>
<td>52</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Monte Carlo Study</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>First Price Auctions: Non-regular Models</td>
<td>59</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusion</td>
<td>65</td>
</tr>
<tr>
<td><strong>APPENDIX A</strong></td>
<td>APPENDIX TO CHAPTER 1</td>
<td>66</td>
</tr>
<tr>
<td>A.1</td>
<td>Detailed Discussion of the Simple Model</td>
<td>66</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Payoff for Exponential Model</td>
<td>66</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Posterior Analysis</td>
<td>67</td>
</tr>
<tr>
<td>A.2</td>
<td>The Adaptive Metropolis Algorithm with GHK sampler</td>
<td>68</td>
</tr>
<tr>
<td>A.3</td>
<td>Bernstein Densities with different $k$</td>
<td>71</td>
</tr>
<tr>
<td><strong>APPENDIX B</strong></td>
<td>APPENDIX TO CHAPTER 2</td>
<td>72</td>
</tr>
<tr>
<td>B.1</td>
<td>Simulation Algorithm</td>
<td>72</td>
</tr>
<tr>
<td>B.1.1</td>
<td>Sampling Valuations and Bidding Function Evaluation for IPV</td>
<td>72</td>
</tr>
<tr>
<td>B.1.2</td>
<td>Sampling for APV: $(\tilde{v}<em>{1,1}, \tilde{v}</em>{2,1}), \ldots, (\tilde{v}<em>{1,R}, \tilde{v}</em>{2,R}) \sim f(\cdot</td>
<td>\theta)$</td>
</tr>
<tr>
<td>B.1.3</td>
<td>Evaluation of Equilibrium Bidding Function for APV</td>
<td>73</td>
</tr>
<tr>
<td>B.1.4</td>
<td>Alternative Method to Estimate Likelihoods</td>
<td>74</td>
</tr>
<tr>
<td>B.2</td>
<td>Basis Functions and Prior Specification</td>
<td>75</td>
</tr>
<tr>
<td>B.2.1</td>
<td>Legendre Polynomials and Smoothness Control</td>
<td>75</td>
</tr>
<tr>
<td>B.2.2</td>
<td>Normalized B splines</td>
<td>75</td>
</tr>
<tr>
<td>B.2.3</td>
<td>Affiliation Restrictions with Normalized B splines</td>
<td>76</td>
</tr>
<tr>
<td>B.2.4</td>
<td>Tail Behavior</td>
<td>76</td>
</tr>
<tr>
<td>B.3</td>
<td>The Adaptive Metropolis Algorithm</td>
<td>77</td>
</tr>
<tr>
<td>B.3.1</td>
<td>Sampling $\psi$ under $A\psi &gt; 0$</td>
<td>77</td>
</tr>
<tr>
<td>B.4</td>
<td>More Discussion on Monte Carlo Studies</td>
<td>78</td>
</tr>
<tr>
<td>B.4.1</td>
<td>Implementation of BSL</td>
<td>78</td>
</tr>
<tr>
<td>B.4.2</td>
<td>Specification for GPV</td>
<td>80</td>
</tr>
<tr>
<td>B.4.3</td>
<td>Specification for Oracle GPV</td>
<td>81</td>
</tr>
<tr>
<td>B.4.4</td>
<td>Information Loss of GPV</td>
<td>82</td>
</tr>
<tr>
<td>B.5</td>
<td>Appendix to Wildcat Auction Analysis</td>
<td>83</td>
</tr>
<tr>
<td>B.5.1</td>
<td>Specification for $f(\cdot</td>
<td>\theta)$</td>
</tr>
<tr>
<td>B.5.2</td>
<td>Computation</td>
<td>83</td>
</tr>
<tr>
<td><strong>REFERENCES</strong></td>
<td></td>
<td>87</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Density Function and Asymmetry of Payoff: 1. Exponential(1) . . . . 15
1.2 Density Function and Asymmetry of Payoff: 2. Beta(α, β) . . . . . . 16
1.3 Density Function and Payoff Structures: Bernstein Density . . . . . 22

2.1 Mapping between Bid Density Space and Valuation Density Space 
(For the IPV) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
2.2 Nonincreasing Inverse Bidding Function for OCS wildcat auctions . 37
2.3 Monte Carlo Results for Log-Normal and Exponential . . . . . . . . 38
2.4 Monte Carlo Results for Nonsmooth and Wildcat-like . . . . . . . . 39
2.5 OCS Wildcat Data and Bid Space Discretization . . . . . . . . . . . 41
2.6 Predictive Densities for OCS wildcat auctions . . . . . . . . . . . . . 43
2.7 Predictive Revenue for OCS wildcat auctions . . . . . . . . . . . . . 45

3.1 Exponential Cost Density and the Implied Payoff Function . . . . . 56
3.2 Exact Bayes Rule vs Approximate Bayes Rule . . . . . . . . . . . . 57
3.3 Exact Bayes Rule vs Plug-in Rule . . . . . . . . . . . . . . . . . . . 58
3.4 Flexible Cost Density and Implied Payoff Structure . . . . . . . . . 59
3.5 Exact Bayes Rule vs. Approximate Bayes Rule (Flexible Density) . . 60
3.6 Exact Bayes Rule vs. Plug-in Rule (Flexible Density) . . . . . . . . 61
3.7 Exact Bayes Rule vs. Approximate Bayes Rule (Non-regular) . . . 62
3.8 Exact Bayes Rule vs. Plug-in Rule (Non-regular) . . . . . . . . . . . 63

A.1 Flexibility of Bernstein Densities . . . . . . . . . . . . . . . . . . . . 71

B.1 The AM outputs for the Log-Normal DGP . . . . . . . . . . . . . . 79
B.2 The AM outputs for the Exponential DGP . . . . . . . . . . . . . . 80
B.3 The AM outputs for the Nonsmooth DGP . . . . . . . . . . . . . . 81
B.4 The AM outputs for the OCS-like DGP . . . . . . . . . . . . . . . 82
B.5 The AM outputs for the OCS wildcat auction data analysis . . . . . . 84
LIST OF TABLES

1.1 Monte Carlo Results for the Simple Parametric Model . . . . . . . . . . 21
1.2 Monte Carlo Results for the Flexible Model . . . . . . . . . . . . . . 23
2.1 MISE Comparisons . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
B.1 Information Loss of GPV (%) . . . . . . . . . . . . . . . . . . . . . 83
This dissertation develops Bayesian methods to analyze data from auctions and produce policy recommendations for auction design. The essay, “Auction Design Using Bayesian Methods,” proposes a decision theoretic method to choose a reserve price in an auction using data from past auctions. Our method formally incorporates parameter uncertainty and the payoff structure into the decision procedure. When the sample size is modest, it produces higher expected revenue than the plug-in methods. Monte Carlo evidence for this is provided.

The second essay, “Flexible Bayesian Analysis of First Price Auctions Using Simulated Likelihood,” develops an empirical framework that fully exploits all the shape restrictions arising from economic theory: bidding monotonicity and density affiliation. We directly model the valuation density so that bidding monotonicity is automatically satisfied, and restrict the parameter space to rule out all the non-affiliated densities. Our method uses a simulated likelihood to allow for a very flexible specification, but the posterior analysis is exact for the chosen likelihood. Our method controls the smoothness and tail behavior of the valuation density and provides a decision theoretic framework for auction design. We reanalyze a dataset of auctions for drilling rights in the Outer Continental Shelf that has been widely used in past studies. Our approach gives significantly different policy prescriptions on the choice of reserve price than previous methods, suggesting the importance of the theoretical shape restrictions.

Lastly, in the essay, “Simple Approximation Methods for Bayesian Auction Design,” we propose simple approximation methods for Bayesian decision making in auction design problems. Asymptotic posterior distributions replace the true posteriors in the Bayesian decision framework, which are typically a Gaussian model (second price auction) or a shifted exponential model (first price auction). Our
method first approximates the posterior payoff using the limiting models and then maximizes the approximate posterior payoff. Both the approximate and exact Bayes rules converge to the true revenue maximizing reserve price under certain conditions. Monte Carlo studies show that my method closely approximates the exact procedure even for fairly small samples.
CHAPTER 1

AUCTION DESIGN USING BAYESIAN METHODS

1.1 Introduction

We introduce a decision theoretic method to choose a reserve price to maximize future revenue for an auction under the private value paradigm. We are especially concerned with situations where the sample from past sales is small. Riley and Samuelson (1981) show that if valuations are independently distributed with CDF \( F(\cdot | \theta^*) \) and PDF \( f(\cdot | \theta^*) \) and the seller’s valuation is \( v_0 \), the revenue maximizing reserve price is \( \rho_R(\theta^*) := \rho^* \), where \( \rho^* \) solves \( \rho^* = v_0 + \frac{1 - F(\rho^* | \theta^*)}{f(\rho^* | \theta^*)} \). Paarsch (1997) first estimates the parameter with the MLE \( \hat{\theta} \) and then proposes \( \rho_R(\hat{\theta}) \) as a consistent estimator of \( \rho_R(\theta^*) \). Therefore, \( \rho_R(\hat{\theta}) \) may be a good approximation of the true \( \rho_R(\theta^*) \) if the sample size is large enough. We call this method a ‘plug-in’ rule to distinguish it from the Bayesian method we propose.\(^1\)

If the sample is small, it may be important to take into account parameter uncertainty and the payoff structure.\(^2\) To see this, we give a simple example where the expected revenue is asymmetric about the true revenue maximizing reserve price point, i.e. it is gradually increasing up to \( \rho_R(\theta^*) \) and decreasing sharply thereafter. Suppose that we adopt the plug-in method. The estimate \( \rho_R(\hat{\theta}) \) may be either smaller or bigger than \( \rho_R(\theta^*) \) due to sampling error. Then, we see that, for the same magnitude of error, overestimation results in greater revenue loss than underestimation. Hence, a rational decision maker would take this into account and try to avoid overestimation. Since the magnitude of sampling error is related to the

\(^1\)Li, Perrigne, and Vuong (2003) extend this line of research to the affiliated private value (APV) case with nonparametric estimation. Their method can be viewed as a plug-in method because they estimated the underlying distribution first and derived the revenue maximizing reserve price regarding the estimate as the true one. In this case, \( \theta \) is infinite dimensional.

\(^2\)In this paper, the expected revenue is often called the payoff.
amount of parameter uncertainty, the decision maker can get a higher payoff by formally considering parameter uncertainty and the payoff structure.

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

We first focus on a simple parametric model for a second price sealed bid auction under the independent private value (IPV) paradigm, since the main components that distinguish our method from the plug-in methods are incorporation of parameter uncertainty and the asymmetric payoff structure, rather than flexibility of the model, auction format, or dependence among valuations. A Monte Carlo study shows that our method produces higher average payoff than the plug-in method when the payoff is asymmetric enough. The advantage of our method appears to be greater for smaller sample sizes. 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.

In the next section, we discuss the seller’s problem and the generality of asymmetric payoff motivating our approach. Then, we propose the Bayesian decision theoretic approach for the seller’s problem and discuss its properties. In Section 1.3, we implement two Monte Carlo studies, one with a simple parametric model and the other with a more flexible one. The former illustrates our methodology and compares it to plug-in methods in finite samples, and the latter shows that our method is beneficial for wider range of sample sizes for flexible models. Section 1.4 concludes. Finally, the appendices collect all computational details.
1.2 Decision Theoretic Approach

1.2.1 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 \( \theta^* \) but may have prior beliefs \( P(\theta) \) over the parameter space \( \Theta \).

Let \( z := (b_{1,t}, \ldots, b_{N_t,t})_T^T \) 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, \( \beta(v_{i,t}|\theta^*) = v_{i,t} \). So, \( b_{i,t} = v_{i,t} \) for all \( i \) and \( t \).

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 \( \theta \), for which the revenue maximizing reserve price is \( \rho^* = 1/\theta^* \). We set \( \theta^* = 1 \). This distribution appears on the left panel of Figure 1.1. 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 \( \rho^* \), and it gets more asymmetric as \( N \) increases. From this, we see that the payoff structure depends on \( N \) and is not generally symmetric.

---

3This implies that the seller knows which parametric family the valuation distribution belongs to. So, we assume that the model is correctly specified.

4See Milgrom and Weber (1982).

5Though \( \beta(v_{i,t}|\theta^*) = v_{i,t} \) is a weakly dominant strategy as long as the valuations are private information, we maintain the IID assumption since it makes the likelihood simple.

6However, we cannot say anything about the relationship between \( N \) and the degree of payoff
exponential distribution case, this asymmetry result is identical for all \( \theta \), 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 \((\alpha, \beta)\). The first row of Figure 1.2 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 1.2. 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 1.2 show that for given \(N\) its payoff structure is more asymmetric than Beta(1,3).

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 Figures 1.1 and 1.2.\(^7\)

\(^7\)We examine the beta, gamma, Pareto, and Weibull distribution with various parameter values. They are all similar to Figures 1.1 and 1.2.
1.2.2 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 $\theta$ is the true parameter, $\rho$ is the reserve price, and $N$ is the number of bidders. Then, the Bayes action is given by

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

(1.1)

Note that the posterior distribution formally quantifies all parameter uncertainty remaining after considering the sample information. We also see that (1.1) also systematically takes into account the payoff structure determined by $N$.

We can compute each component of (1.1). Suppose the seller’s valuation for the
auctioned item is zero. Then, under the IPV paradigm, the payoff function is given by

$$\Pi(\theta, \rho, N) = N \left( \rho (1 - F(\rho \| \theta)) F(\rho \| \theta)^{N-1} + \int_0^{\theta} y(N - 1)(1 - F(y \| \theta)) F(y \| \theta)^{N-2} f(y \| \theta) dy \right)$$

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

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

(1.3)

Note that it is enough to compute the posterior up to the normalizing constant for our purpose, because multiplying the right hand side of (1.1) by some constant does not change the solution.

\( \rho_B(z, N) \) is optimal with respect to the average (Bayes) risk principle, one of the most widely used frequentist decision principles.\(^8\) Let \( d : Z \rightarrow A \) be a decision rule where \( Z \) 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 \( L := -\Pi \).\(^9\) Then, the Bayes risk is \( r(P, d) := \int_\Theta R(\theta, d)dP(\theta) \) for the prior distribution \( P \). 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.\(^10\) 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,

$$r(P, d) = \int_\Theta R(\theta, d)p(\theta)d\theta$$

$$= \int_\Theta \left\{ \int_Z L(\theta, d(z))f(z \| \theta)dz \right\} p(\theta)d\theta$$

\(^8\)See Berger (1985).\(^9\)We suppress the dependence of the loss on \( N \).\(^10\)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.
\[
\int_z \int_{\Theta} L(\theta, d(z))p(\theta)f(z|\theta)d\theta dz \\
= \int_z \left\{ \int_{\Theta} L(\theta, d(z))cp(\theta|z)d\theta \right\} dz
\]

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 \( z \in Z \). 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 \( \theta \), \( \Pi(E[\theta|z], \rho, N) \neq E[\Pi(\theta, \rho, N)|z] \) in general. Let \( \hat{\theta} := E[\theta|z] \).

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 \( \theta^* \) 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 \( \theta^* \) 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 \( \theta \) can be employed. For example, one could estimate \( \hat{\theta} \) minimizing the mean squared error loss \( L(\theta, a) := (\theta - a)^2 \) and derive \( \rho_R(\hat{\theta}) \). This is the optimality of the estimation procedure but not the optimality with respect to the seller’s payoff.

1.3 Monte Carlo

1.3.1 Implementation and Monte Carlo Design

Generally, the integral of (1.1) does not have a closed form expression. Therefore, once the posterior distribution is computed, we estimate the posterior mean of payoff

\(^{11}\)See Section 3.1 for more discussion of \( \hat{\theta} \).
using simulation. Let

$$\hat{\rho}_B(z, N) = \arg \max_{\rho \in [0,1]} \frac{1}{S} \sum_{s=1}^{S} \Pi(\theta^s, \rho, N)$$  \hspace{1cm} (1.4)$$

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

For the plug-in method we adopt

$$\hat{\theta}_B := \arg \min_{a \in \Theta} \int (\theta - a)^2 dP(\theta|z)$$  \hspace{1cm} (1.5)$$
as the estimator for the true parameter $\theta^*$. 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 $\theta$. 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.\textsuperscript{12} 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.\textsuperscript{13}

1.3.2 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

\textsuperscript{12}See Theorem 8.3 of Lehmann and Casella (1998).

\textsuperscript{13}See Chapter 1 and 4 of Berger (1985).
mean equals $\theta^*$. The advantage of using a Gamma prior is that it is conjugate for the parameter of an exponential distribution. Therefore, it is convenient to estimate $\hat{\theta} = E[\theta|z]$ and to implement the plug-in method. Moreover, it turns out that the posterior payoff has a closed form expression, so the implementation of the Bayes action (1.4) is also straightforward. See Appendix A.1.

Table 1.1 presents the results of the Monte Carlo study. Since $\theta$ 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, $M = 5$ 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 1.1.

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 $M = 5$ 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
Table 1.1: Monte Carlo Results for the Simple Parametric Model

<table>
<thead>
<tr>
<th>Sample Size</th>
<th># of bidders</th>
<th>$\rho_B(z, N)$</th>
<th>$\rho_R(\hat{\theta}_B)$</th>
<th>% gain over plug-in method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$N$</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3</td>
<td>0.8426 (0.1824)</td>
<td>1.0033 (0.3726)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.8008 (0.1960)</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>0.7594 (0.2049)</td>
<td>&quot;</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>3</td>
<td>0.9114 (0.1147)</td>
<td>0.9971 (0.2086)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.8984 (0.1208)</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>0.8848 (0.1264)</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

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

be consistent. In Chapter 3 we show that the Bayesian method is also consistent. Intuitively, if there were no parameter uncertainty and the posterior were degenerate at $\theta^*$, 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 $\rho^*$.

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 in column (3) for the case where $M = 5$ are only about 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 $M = 5$. 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.
1.3.3 Flexible Model

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

\[ f(v|\theta) = \sum_{j=1}^{k} \theta_j \beta(v|j, k - j + 1) \]  \hspace{1cm} (1.6)

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

For the Monte Carlo study, we fix \( k = 15 \) and employ the uniform prior over \( \Theta \) which is now the 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.\(^{14}\) We assume that the valuation distribution has the density function plotted on the left panel of Figure 1.3. Then, the payoff structures for \( N = 3, 4, 5 \) are appeared on the right panel of Figure 1.3.

\(^{14}\)Appendix A.3 discusses the dependence of the flexibility of the Bernstein density on \( k \).
Table 1.2: Monte Carlo Results for the Flexible Model

<table>
<thead>
<tr>
<th>Sample Size $M$</th>
<th># of bidders $N$</th>
<th>$\rho_B(z, N)$ (1)</th>
<th>$\rho_R(\hat{\theta}_B)$ (2)</th>
<th>% gain over plug-in method (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3</td>
<td>0.3488 (0.0542)</td>
<td>0.3658 (0.0554)</td>
<td>4.5920</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.3406 (0.0536)</td>
<td>“</td>
<td>5.8263</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.3327 (0.0529)</td>
<td>“</td>
<td>6.5040</td>
</tr>
<tr>
<td>75</td>
<td>3</td>
<td>0.2266 (0.0241)</td>
<td>0.2299 (0.0252)</td>
<td>0.3298</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2250 (0.0237)</td>
<td>“</td>
<td>0.3277</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.2235 (0.0233)</td>
<td>“</td>
<td>0.2857</td>
</tr>
</tbody>
</table>

Note: The true revenue maximizing reserve price is given by $\rho^* = 0.1840$

Similarly to the exponential distribution, the payoff structure is more asymmetric for larger $N$ for these $N$’s. We try the sample size $M = 20, 75$ to make a comparison with the simple parametric case in the previous subsection.

Table 1.2 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 1.1 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 1.1 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 1.2. In addition, the gains of sample size $M = 20$ for the simple model roughly amounts to the ones of sample size $M = 75$ for the flexible model. 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 is not informative on the
choice of the parametric family for the valuation 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 (1999a) and Petrone (1999b) develop, the algorithm she suggests in her papers would be a natural choice. However, since we use a fixed, but high dimensional \( \theta \) for the Monte Carlo study above, we employ the adaptive Metropolis algorithm of Haario, Saksman, and Tamminen (2001). Since \( \theta \) can be seen as a probability distribution over \( k \) distinct points, each candidate of \( \theta \) should satisfy the implied restrictions.\(^{15}\) Hence, we incorporate the Adaptive Metropolis algorithm with the GHK sampler so that the proposal function generates only such \( \theta \)'s.\(^{16}\) We explain the modified Adaptive Metropolis algorithm in Appendix A.2 in detail.

1.4 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.

\(^{15}\)All elements in \( \theta \) are positive and sum up to one.

\(^{16}\)For the GHK sampler, see Hajivassiliou and McFadden (1998) and Keane (1990).
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.
2.1 Introduction

Auctions are a frequently used market institution for allocating a variety of economic resources, and account for a significant portion of the U.S. economy. For example, since 1954 the U.S. government has leased approximately thirty thousand tracts for offshore oil and gas development in areas of the Gulf of Mexico. Each of these Outer Continental Shelf (OCS) auctions has a sale price which ranges from hundreds of thousands to several million dollars.\footnote{1} Reflecting the economic significance of auctions, many economists have extensively investigated this mechanism. In particular, auction theory has characterized equilibrium bidding behavior and optimal auction design for a given valuation distribution, while empirical research has focused on estimating the valuation distribution and using these estimates for auction design, e.g., the estimation of optimal reserve prices.

For first price sealed bid auctions with independent private values (IPV), the pioneers in the empirical auction literature specify the valuation distribution using strong parametric assumptions.\footnote{2} However, a flexible specification is often preferred because inference strongly relies on the shape of valuation distribution. For this reason Guerre, Perrigne, and Vuong (2000) indirectly recover the valuation density by using the estimated bid distribution. Li, Perrigne, and Vuong (2002) generalize this
method to auctions with affiliated private value (APV). These indirect approaches have been widely used because they are fully flexible and computationally simple.

However, for the estimation of bid distribution, the indirect methods do not fully exploit the shape restrictions arising from economic theory: bidding monotonicity and density affiliation. Thus, these methods may rely on an estimated bid density that is either nonaffiliated or associated with a nonincreasing inverse bidding function. This problem is empirically important for the datasets of interest to researchers and policymakers. For instance, Li, Perrigne, and Vuong (2003) propose an optimal reserve price for the OCS auctions using a bid sample of 217 observations. We find that their policy recommendations are based on an estimated bid density that cannot be generated by an equilibrium. Note that recently Henderson, List, Millimet, Parmeter, and Price (2008) develop a kernel approach to impose bidding monotonicity, but do not discuss affiliated private values.

Motivated by this, we develop a Bayesian framework that satisfies both bidding monotonicity and density affiliation to provide more precise inference. Specifically, we directly parametrize the valuation density so that bidding monotonicity is satisfied for every parameter value and put zero prior weight on every density that violates density affiliation. For a reasonably flexible analysis we employ a series representation for the valuation density. To handle such a rich specification, we use a simulated likelihood. In particular, we employ a multinomial likelihood defined on a finely discretized sample space. Then, as Flury and Shephard (2008) discuss, we can obtain an exact posterior even with a finite number of simulation draws for each candidate parameter value.\(^3\) Note that the simulation error would typically cause loss of efficiency for other simulation methods.

We revisit a sample of OCS auctions that Li, Perrigne, and Vuong (2003) have analyzed. Our methodology simulates the posterior distribution of the valuation density for the OCS wildcat auctions. The resulting bid density estimate fits the data very well. We select a reserve price to maximize the seller’s future revenue

\(^3\) See Andrieu, Doucet, and Holenstein (2007a), Andrieu, Doucet, and Roberts (2007b). The information loss from the discretization can be minimal, if we use small bins.
using the Bayesian decision method introduced in Chapter 1. We find that a reserve price of $462 per acre is optimal given our likelihood and prior. According to our counterfactual analysis, this price increases the predictive revenue for each tract by $262,414 relative to the actual reserve price of $15. In addition, our recommended reserve price is much larger than the value of $273 obtained by Li, Perrigne, and Vuong (2003), suggesting the importance of the additional theoretical shape restrictions.

2.2 Auction Models and Empirical Method

This section defines the APV/IPV auctions, develops our empirical methodology, and discusses how it differs from the indirect approaches.

2.2.1 First Price Auctions under the APV/IPV Paradigm

Consider $N \geq 2$ risk neutral bidders in an auction with a reserve price $\rho$. The bids are collected simultaneously and the bidder with the highest bid obtains the auctioned item at the price equal to his own bid. Let $(v_1, \ldots, v_N) \in \mathbb{R}_+^N$ be a vector of valuations drawn from a joint distribution $F$ assumed to be absolutely continuous with density $f$. Each bidder $i = 1, \ldots, N$, after observing his own valuation $v_i$, bids $b_i$ to maximize his expected utility, $(v_i - b_i) \Pr (b_i > \max_{j \neq i} b_j)$. The number of bidders $N$, reserve price $\rho$, and the valuation distribution $F$ are common knowledge among the bidders.

The auction is said to be under the APV paradigm if the valuations $v_1, \ldots, v_N$ are affiliated. Informally, affiliation implies that if some elements of $(v_1, \ldots, v_N)$ are large, others are more likely to be also large. (See Milgrom and Weber (1982) for a rigorous treatment; also below we discuss affiliation further.) Hence, each bidder gets some information on the distribution of other bidders’ valuation from his own valuation and takes it into account to bid optimally. We assume that $F$ is exchangeable so that every bidder is ex-ante identical. This auction mechanism implies a symmetric game of incomplete information among the bidders. For this
game, Milgrom and Weber (1982) derive a symmetric Bayesian Nash equilibrium with a strictly increasing bidding function. Let \( f_{y_i|v_i}(\cdot|\cdot) \) be the conditional density of \( y_i := \max_{j \neq i} v_j \) given \( v_i \). If \( v \geq \rho \), the equilibrium bidding function is given by

\[
\beta(v|\rho, F) := v - \int_{\rho}^{v} \exp \left\{ - \int_{\alpha}^{v} \frac{f_{y_1|v_1}(u|\alpha)}{f_{y_1|v_1}(u|v_1)} du \right\} d\alpha
\]

(2.1)

Otherwise, any value strictly less than \( \rho \) is optimal. Note that the IPV paradigm is a special case of the APV paradigm in which \( v_1, \ldots, v_N \) are independent. Under the IPV, Equation (2.1) simplifies to

\[
\beta(v|\rho, F) := v - \int_{\rho}^{v} \left\{ \frac{F(\alpha)}{F(v)} \right\}^{N-1} d\alpha
\]

(2.2)

where \( F(\cdot) \) is the marginal distribution of an individual valuation, with a slight abuse of notation.⁴

Suppose we observe a random sample of \( T \) auctions with the common valuation distribution \( F \) and a zero reserve price. Let \( \beta(\cdot|F) := \beta(\cdot|0, F) \). We assume that for each auction, the \( N \) bidders follow the equilibrium bidding function \( \beta(\cdot|F) \). Let \( z := \{(b_{1,t}, \ldots, b_{N,t})\}_{t=1}^{T} \) denote the data set. From these data we want to learn \( F \).

2.2.2 Bayes with Simulated Likelihood

For this purpose, we model the valuation density using a flexible specification \( f(\cdot|\theta) \) and rule out all \( \theta \)'s with a nonaffiliated \( f(\cdot|\theta) \). (For the IPV case, we model only the marginal of the valuation density.) We take a Bayesian approach for the following reasons. First, we can formally control smoothness and tail behavior reflecting our prior beliefs. Thus, our data analysis, while flexible, would not result in an unreasonably noisy or thick-tailed density. Second, for the choice of reserve price, we can use Bayes rule that may produce higher revenue by formally considering the

⁴Under the IPV paradigm with the exchangeability assumption, the joint valuation distribution is the product of \( N \) identical marginals. i.e., \( F(v_1, \ldots, v_N) = \prod_{i=1}^{N} F(v_i) \). This abuse of notation should not lead to any confusion.
structure of payoffs and parameter uncertainty. Third, we can handle the restricted parameter space (due to affiliation) simply by putting zero prior over the excluded area of the parameter space.

However, the likelihood is challenging to evaluate. Consider the likelihood

$$
\prod_{t=1}^{T} \left\{ \frac{f(\beta^{-1}(b_{1,t}|\theta), \ldots, \beta^{-1}(b_{N,t}|\theta))}{\prod_{i=1}^{N} \beta'(\beta^{-1}(b_{i,t}|\theta)|\theta)} \right\} \cdot 1\{\max(z) \leq \bar{b}(\theta)\}
$$

(2.3)

where $\bar{b}(\theta) := \lim_{v \to \infty} \beta(v|\theta)$ and $1\{A\} = 1$ if $A$ is true, otherwise, zero. Since (2.3) lacks a closed form expression for our specification, we would need to numerically evaluate $\beta^{-1}(\cdot|\theta)$ at each data point for many parameter values. This is impractical because even for a single evaluation of $\beta^{-1}(\cdot|\theta)$, we need to evaluate the triple integrals in (2.1) repeatedly. For this reason, we simulate the equilibrium bids to construct a likelihood; generate random values from $f(\cdot|\theta)$, evaluate (2.1) at each drawn value, and then estimate the likelihood using these simulated bids.

When the likelihood is simulated, the Bayesian method is even more useful. Andrieu, Doucet, and Holenstein (2007a) and Andrieu, Doucet, and Roberts (2007b) show that using an unbiased simulated likelihood a Markov Chain Monte Carlo (MCMC) algorithm converges to the exact posterior distribution. The idea is simple. Let $p(\theta)$ and $p(z|\theta)$ denote a prior and a conditional density of data $z$ given $\theta$, and let $p_u(z|\theta)$ be an unbiased estimator for $p(z|\theta)$ constructed by a uniform random vector $u$. $E[p_u(z|\theta)] = \int p_u(z|\theta)p(u)du = p(z|\theta)$ and $\int p_u(z|\theta)p(u)du = \int p_u(z|\theta)du$. Thus, $\int p_u(z|\theta)du = p(z|\theta)$ and hence $p_u(z|\theta)$ can be regarded as a joint density of $(u, z)$, say, $p(u, z|\theta)$. Suppose we apply an MCMC algorithm to $p(\theta)p_u(z|\theta)$ instead of $p(\theta)p(z|\theta)$ drawing a new $u$ at each iteration. Then, this algorithm can be seen as an sampling scheme generating $(u_1, \theta_1), \ldots, (u_S, \theta_S)$ from

$$
p(u, \theta|z) \propto p(\theta)p(u, z|\theta) = p(\theta)p_u(z|\theta)
$$

Therefore, $\theta_1, \ldots, \theta_S$ are draws from the exact posterior $p(\theta|z)$. This method is

---

5See Chapter 1 for more detail. We also use Bayes rule to choose a reserve price to maximize the seller’s predictive revenue for the OCS wildcat auctions in section 4.
called the Bayes with Simulated Likelihood (BSL) in this paper.\footnote{Flury and Shephard (2008) introduce this method to econometrics. Fernandez-Villarverde, Rudio-Ramirez, and Santos (2006) and Fernandez-Villaverde and Rudio-Ramirez (2007) have also employed simulation to form a likelihood.}

To construct an unbiased simulated likelihood, we discretize the sample space into $D$ bins.\footnote{Note that, for an APV paradigm, one datum is an observed auction, or its bid profile $(b_{1,t}, \ldots, b_{N,t}) \in z$. But, for an IPV model, a single bid, $b_{i,t}$, can be used as a sample point, since all the observed bids are independently and identically distributed.} Let $y_d$ denote the number of sample points in the $d$-th bin. In addition, let $\pi_d(\theta)$ be the probability of the $d$-th bin under $\theta$. Then, the discretized sample space with the implied histogram $y := (y_1, \ldots, y_D)$ leads to a multinomial likelihood given by

$$L(\theta|y) = \prod_{d=1}^{D} \{\pi_d(\theta)\}^{y_d}$$

(2.4)

Let $\tilde{\pi}_d(\theta)$ be the fraction of simulation draws belonging to the $d$-th bin. Then, we estimate (2.4) unbiasedly using

$$\tilde{L}(\theta|y) := \prod_{d=1}^{D} \{\tilde{\pi}_d(\theta)\}^{y_d}$$

(2.5)

This approach is very easy to compute and fairly flexible.\footnote{Note that it might seem to be natural to estimate the theoretical bid density using a kernel method or a method of series over the simulated auctions to construct likelihoods. However, these methods are biased and, moreover, the latter would not be computationally simple, if flexible.} For a given discretization, we obtain the exact posterior by using a new simulation draw at each iteration. The information loss of using $y$ rather than $z$ would be minimal if the bins are small.

An alternative method would be the maximum likelihood estimator (MLE) to maximize (2.3) or its simulation estimate. However, as Hirano and Porter (2003) show, the MLE is not efficient because the bid data has parameter dependent support, $[0, \bar{b}(\theta)]$. Moreover, maximizing (2.3) under many constraints (due to affiliation) over a high dimensional parameter space would be much more involved. Lastly, simulation typically inflates the asymptotic variance of the MLE and also the number of simulation draws needs to grow to infinity unlike BSL.
2.2.3 Valuation Density Specification

We discuss our choice for $f(\cdot|\theta)$ for the IPV and APV, separately. The simpler goes first.

**IPV case**

We use the specification of Verdinelli and Wasserman (1998). Let $\tilde{F}$ be a simple parametric distribution indexed by $\mu$ with density $\tilde{f}$, $h$ be a flexible density on $[0, 1]$ with parameter $\psi$, and $\theta := (\mu, \psi)$. The specification is given by

$$f(v|\theta) = \tilde{f}(v|\mu)h(\tilde{F}(v|\mu)|\psi)$$

(2.6)

Note that (2.6) can be seen as a derivative of a flexible CDF, $H(\tilde{F}(v|\mu)|\psi)$.

We construct $h(\cdot|\psi)$ as follows. Let $\{\phi_i\}$ denote a sequence of functions on $[0, 1]$ such that we can approximate any function on $[0, 1]$ using their linear combination $\sum_{i \in I} \psi_i \phi_i$ for some index $I$ and some real numbers $\{\psi_i\}_{i \in I}$. Polynomials, splines, or Fourier functions can construct such $\{\phi_i\}$. Then,

$$h(x|\psi) \propto \exp\left(\sum_{i \in I} \psi_i \phi_i(x)\right) \cdot 1 \{x \in [0, 1]\}$$

(2.7)

approximates a density on $[0, 1]$. This approach is a method of series. Note that the seminonparametric method of Gallant and Nychka (1987) is an example of series estimation in econometrics.

We employ this particular specification for the following reasons. First, it is a parsimonious way to specify a flexible model. To see this, take the logarithm on each side of (2.6). Then, $\log f(v|\theta) = \log \tilde{f}(v|\mu) + \psi_1 \phi_1(\tilde{F}(v|\mu)) + \psi_2 \phi_2(\tilde{F}(v|\mu)) + \cdots$. Thus, we approximate the valuation density first using $\log \tilde{f}$ and then explain the difference between the true density and $\log \tilde{f}$ using the additional terms. Obviously, for a given accuracy, if $\tilde{f}$ is already a good approximation to the density of valuations, we would not need many additional terms. Hence, our model may not need to have a large number of components. In this case, a computational advantage also follows.
In practice, one may choose \( \tilde{f} \) to reflect beliefs about the form of the true valuation density. For example, if the empirical distribution of the bid sample resembles an exponential distribution, one could use an exponential with hazard rate \( \mu \), believing the valuations may be similarly distributed. Second, when it extends to the APV, exchangeability and density affiliation can be simply characterized as follows.\(^9\)

**APV case**

We focus on the two bidder case \((N = 2)\) for clarity. Then, (2.6) extends to

\[
f(v_1, v_2|\theta) = \tilde{f}(v_1|\mu)\tilde{f}(v_2|\mu)h\left(\tilde{F}(v_1|\mu), \tilde{F}(v_2|\mu)|\psi\right)
\]

and (2.7) becomes

\[
h(x_1, x_2|\psi) \propto \exp \left(\sum_{i \in I}\sum_{j \in I} \psi_{i,j}\phi_i(x_1)\phi_j(x_2)\right) \cdot 1\{ (x_1, x_2) \in [0,1]^2 \}
\]

Recall that (2.8) must be affiliated and exchangeable.\(^{10}\) The affiliation (or equivalently density log-supermodular) holds if and only if

\[
\frac{\partial^2}{\partial x_1 \partial x_2} \sum_{i \in I}\sum_{j \in I} \psi_{i,j}\phi_i(x_1)\phi_j(x_2) \geq 0
\]

for every \((x_1, x_2) \in [0,1] \times [0,1]\), infinitely many constraints. We use normalized B splines to construct \(\{\phi\}\), because, then, (2.10) reduces to a finite number of linear inequality constraints. (See Beresteau (2007).) In addition, the exchangeability restriction is equivalent to \(\psi_{i,j} = \psi_{j,i}\) for all \(i, j \in I\). Let \(\psi\) be a vector of \(\psi_{i,j}\) with

\(^9\) Though the normal mixture studied by Ferguson (1973), Escobar (1994), and Escobar and West (1995) are popularly used, it is not appropriate for our purpose because it is hard to extend to the APV. In particular, there is no convenient way to impose density affiliation.

\(^{10}\) One might want to use a copula to account for a correlation structure. In this case, the nonparametric (Bernstein) copula of Sancetta and Satchell (2004) should be employed, because we want the specification to be flexible. However, there is no convenient way to impose affiliation. Even if there is, the specification would be needlessly complicated, because we need to use another flexible specification for the marginal density separately. Only when we use the Bernstein density of Petrone (1999b) and Petrone (1999a) for the marginal, there is a simplification, which, however, would actually lead back to (2.8) with a slightly different \(h\) than (2.9).
\( i \geq j \). Then, there is a matrix \( A \) such that (2.8) is affiliated and exchangeable if and only if

\[
A \psi \geq 0
\]  

(2.11)

See Appendix for details of the matrix \( A \). Now, we simply put zero prior weight on the set of \( \psi \)'s violating (2.11) to impose the affiliation.

**Outline of Simulation Algorithm**

For the IPV case, Guerre, Perrigne, and Vuong (2000) show that random bids are also independent. Thus, we can model the marginal density \( f(\cdot | \theta) \) only. We estimate the bin probabilities using the simulation algorithm given by

1. draw \( \tilde{v}_1, \ldots, \tilde{v}_R \) from \( f(\cdot | \theta) \).

2. compute \( \tilde{b}_1, \ldots, \tilde{b}_R \) by evaluating (2.2) at each \( \tilde{v}_r, r = 1, \ldots, R \).

3. Then, \( \hat{\pi}_d(\theta) := R^{-1} \sum_{r=1}^{R} 1 \{ \tilde{b}_r \in d\text{-th bin} \} \)

For the APV case, we use similar algorithm. But, we need to jointly draw \( N(= 2) \) dimensional valuations and to discretize \( N \) dimensional sample space. Appendix A provides the simulation algorithm for both the IPV and APV models.

**2.2.4 Discussion**

Guerre, Perrigne, and Vuong (2000) first develop the nonparametric indirect method for the IPV case. They observe that the inverse of the bidding function can be expressed as

\[
\beta^{-1}(b) = b + \frac{G(b)}{(N - 1)g(b)}
\]  

(2.12)
where \( G \) and \( g \) are the marginal bid distribution and its density.\(^{11}\) Thus, if we knew \( G \) and \( g \), we could uncover \( v_{i,t} \) for each \( b_{i,t} \) in \( z \). Observing this, Guerre, Perrigne, and Vuong first estimate the bid distribution, say \( \hat{G} \) and \( \hat{g} \), from the observed bids and then estimate the valuation density over \( \hat{v}_{i,t} := b_{i,t} + \frac{\hat{G}(b_{i,t})}{(N-1)\hat{g}(b_{i,t})} \). This method is very useful, because it is fully flexible and computationally simple.

However, since they estimate \( \hat{g} \) without imposing bidding monotonicity, the estimated inverse bidding function may not be increasing. To see the implication of this, consider Figure 2.1. \( P \) denotes the set of all densities on \( \mathbb{R}_+ \), which can be seen as a valuation density space (left panel). Let \( M \subset P \) be the set of bid densities constructing a strictly monotone inverse bidding function. Guerre, Perrigne, and Vuong show that the equilibrium creates a one to one mapping between \( M \) and \( P \). That is, any \( g \in M \) can be rationalized as an equilibrium by some valuation density \( f \in P \). But, a bid density outside \( M \), say \( g'' \), cannot be an equilibrium outcome. Note that (2.12) would connect \( g'' \) with a density in \( P \), say \( f' \). But, the equilibrium links \( f' \) and some other bid density \( g' \in M \).

Li, Perrigne, and Vuong (2002) extend this indirect method to the APV. But, they exploit neither bidding monotonicity nor density affiliation for the estimation of bid distribution. Li, Perrigne, and Vuong (2003) estimate the optimal reserve price exclusively based on the estimated bid distribution for the OCS wildcat auctions. We find that the estimated bidding function for this sample violates monotonicity. See Figure 2.2. This suggests that the estimated bid density would be fairly different from the true one. Hence, inference based on this estimated bid density would not be accurate, and policy recommendations may not be reliable.

The idea of this paper is to rule out all the bid densities associated with a nonincreasing inverse bidding function, e.g., \( M^c \). We find that it is very difficult to nonparametrically estimate the bid density while imposing the monotonicity and affiliation. Thus, we take a direct approach and use a simulated likelihood to allow

\(^{11}\) Its derivation is simple. The expected utility of bidder \( i \) bidding \( b_i \) can be expressed by \( (v_i - b_i)F(\beta^{-1}(b_i))^{N^{-1}} \). The first order condition is \( -F(\beta^{-1}(b_i))^{N^{-1}} + (v_i - b_i)(N - 1)F(\beta^{-1}(b_i))^{N^{-2}}f(\beta^{-1}(b_i))/\beta'(\beta^{-1}(b_i)) = 0 \). From this, we obtain \( v_i = b_i + \frac{F(\beta^{-1}(b_i))}{(N-1)f(\beta^{-1}(b_i))/\beta'(\beta^{-1}(b_i))} \). Note that \( G(b) = F(\beta^{-1}(b)) \) and \( g(b) = f(\beta^{-1}(b))/\beta'(\beta^{-1}(b)) \).
for a flexible specification. Since our method satisfies all the shape conditions, it would provide more precise inference. However, we consider that our method complements the indirect approach because it is not fully nonparametric and imposes more computational costs. Our method may be particularly useful for small samples, because the contribution of the additional theoretical shape restrictions would be significant.

2.3 Monte Carlo

We compare our method (BSL) with Guerre, Perrigne, and Vuong (2000) (GPV). For BSL we employ Legendre polynomials for \( \{ \phi \} \) in (2.7) and use priors on parameters to control smoothness of the valuation density.\(^{12}\) We report the predictive density estimate \( E[f(\cdot|\theta)|y] \).\(^{13}\) For GPV we use the rule of thumb proposed by Guerre,

\(^{12}\)See Appendix B for Legendre polynomials.

\(^{13}\) Alternatively, we could use \( f(\cdot|\hat{\theta}_B) \) with \( \hat{\theta}_B = E[\theta|y] \). We find this gives only slightly different estimates.
Perrigne, and Vuong (2000) to choose bandwidths. In addition, since one can use different bandwidths, we also use the bandwidths minimizing the mean integrated squared error (MISE).\footnote{The MISE is a precision measure of the density estimate. Let $\hat{f}_z$ be an estimate for the true density $f_0$. Then, $MISE(\hat{f}_z) = \int E_z (\hat{f}_z(x) - f_0(x))^2 dx$ which is decomposed into $\int V_z (\hat{f}_z(x)) dx + \int (E_z \hat{f}_z(x) - f_0(x))^2 dx$. Thus, the MISE is small, only when both variance and bias are small.} We call this procedure ‘Oracle GPV.’ Note that Oracle GPV is infeasible in the real world because the true valuation density is unknown. For a fixed valuation density, we employ 1,000 Monte Carlo replications. For each replication, we generate a new sample of size $T \times N = 200 \times 2$ (similar to the OCS wildcat data) and run BSL, GPV, and Oracle GPV.

We try four different valuation densities. First, Guerre, Perrigne, and Vuong (2000), for their Monte Carlo study, employ a truncated log-normal with parameter 0 and 1 with support $[0.055, 2.5]$. We rescale it so that its support is $[0, 1]$. We label it ‘Log-Normal.’ Second, we use a truncated exponential distribution with mean 1/6 and support $[0, 1]$, denoted by ‘Exponential.’ Figure 2.3 summarizes the results for Log-Normal and Exponential. Each panel plots the true valuation density (plain), point-wise 5-th, 95-th percentiles, and point-wise mean (dotted).
Figure 2.3: Monte Carlo Results for Log-Normal and Exponential

along with a typical density estimate (dashed). Apparently, BSL is more precise than GPV, providing much narrower 90% frequency bands and behaving better near the boundaries. Strikingly, Table 2.1 shows that BSL has even smaller MISE than Oracle GPV. That is, GPV cannot be more precise than BSL for these valuation densities.

However, this comparison may seem to be a little unfair because the true valuation densities are very smooth and BSL imposes this smoothness, whereas GPV is designed to work well for a wide range of density functions. Thus, we introduce another valuation density for which GPV might be more accurate. We use the typical GPV estimate under Log-Normal (dashed-line on Figure 2.3(a)) for the third data generating process (DGP). We call this ‘Nonsmooth.’ GPV might perform better, because BSL “incorrectly” imposes more smoothness, BSL would not be sufficiently flexible, and Nonsmooth is actually from GPV. The Monte Carlo results in Figure 2.4 show that BSL is still superior to GPV, but very similar to Oracle GPV. In fact, MISE of Oracle GPV is slightly smaller than BSL (Table 2.1), indicating that for
Figure 2.4: Monte Carlo Results for Nonsmooth and Wildcat-like

Table 2.1: MISE Comparisons

<table>
<thead>
<tr>
<th>DGP's</th>
<th>GPV</th>
<th>Oracle GPV</th>
<th>BSL</th>
<th>BSL × 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Normal</td>
<td>0.23805</td>
<td>0.02955</td>
<td>0.01034</td>
<td>4.344</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.95726</td>
<td>0.15037</td>
<td>0.02430</td>
<td>2.539</td>
</tr>
<tr>
<td>Nonsmooth</td>
<td>0.29562</td>
<td>0.14500</td>
<td>0.16283</td>
<td>55.080</td>
</tr>
<tr>
<td>Wildcat-like</td>
<td>1.04260</td>
<td>0.20158</td>
<td>0.06406</td>
<td>6.144</td>
</tr>
</tbody>
</table>

some bandwidths GPV can be more precise.

The fourth DGP is the marginal valuation density estimate of the OCS wildcat auctions (Figure 2.6(b)). Overall, the Monte Carlo results are very similar to the Exponential case. To make a connection to the next section, we compare the expected revenues for the policy implications on reserve prices under GPV and BSL. For GPV we choose a reserve price maximizing the seller’s revenue for the estimated bid distribution following Li, Perrigne, and Voung (2003) and for BSL we maximize

15 Though this estimate comes from the APV model, we conduct this Monte Carlo under the IPV for simplicity.
16 Figure 2.4 second row displays the experiment results. See also Table 2.1 and Table 2.2.
the posterior mean of the seller’s revenue following Kim (2008). We find that BSL produces 18.30% higher revenue than GPV for the Wildcat-like distribution. For other experiments, BSL gains revenue increase by 5.52% for Log-Normal, 8.43% for Exponential, and 3.43% for Nonsmooth. Appendix D provide more detailed discussions of the Monte Carlo studies in this section.

2.4 Estimation and Auction Design for OCS Wildcats

We apply our methodology to the OCS wildcat auctions to simulate the posterior of the valuation distribution. We provide the valuation/bid density estimates and choose a reserve price maximizing the seller’s future expected revenue using the decision theoretic method introduced by Kim (2008).

2.4.1 Data and Sample Space Discretization

We discuss the OCS wildcat data briefly. The sample consists of 217 auctions, each having two bids in 1972 dollars. It contains the auctions between 1954 and 1969 among the sales held by the U.S. federal government to sell its mineral right on oil and gas on offshore lands off the Texas and Louisiana coasts.

Figure 2.5(a) is the marginal histogram of the sample, which roughly resembles an exponential density. The sample mean and standard deviation are 1.458 and 2.557 (×100 dollars for each). Panel (b) shows how the bid pairs \((b_1, b_2)\) with \(b_1 > b_2\) are scattered. Most bids are condensed around the origin, while the sample has a long tail. The sample correlation is 0.412, but some negative correlation pattern is observed in the tail area. To examine closely, panel (c) and (d) enlarge the parts for \(b_1 < 9.3\) and \(b_1 < 2.8\), respectively.

We follow the assumptions that Li, Perrigne, and Vuong (2000) and Li, Perrigne, and Vuong (2003) maintain for this sample: nonbinding reserve price, no dynamic considerations, auction homogeneity, and the symmetric APV paradigm. They ar-

\footnote{For a thorough data description, see Porter (1995), McAfee and Daniel (1992), Hendricks and H (1992), Hendricks, Pinkse, and Porter (2003), Li, Perrigne, and Vuong (2000), and Li, Perrigne, and Vuong (2000). The dataset is publically available at \url{http://capcp.psu.edu/index.html}}
Figure 2.5: OCS Wildcat Data and Bid Space Discretization

The assumptions have been taken by other researchers who use the same sample, there has been a disagreement on the APV hypothesis. For example, Li, Perrigne, and Vuong (2000), Li, Perrigne, and Vuong (2003) and Campo, Perrigne, and Vuong (2003) take the APV paradigm, while Hendricks, Pinkse, and Porter (2003) support the pure common value (PCV) paradigm.

Lastly, to obtain an unbiased simulated likelihood, we discretize the sample space as indicated in Figure 2.5. In panel (b) (and (c) and (d), too), for any two adjacent

\[^{18}\text{For example, they claim that the actual reserve price $15 is too low to be an effective screening device and the game induced by the auction mechanism can be seen as symmetric, because all bidders have equal opportunity to access the same information on the auctioned tract.}\]
squared bins on the horizontal axis, the right one is 80% larger than the left one. For each squared bin, we put an equally sized bin right above it until the remaining trapezoid cannot contain another squared bin. We target to construct 30 nonempty bins.

2.4.2 Valuation/Bid Density Estimation

We model the valuation density using (2.8) with (2.9). In particular, to impose the affiliation restriction, we employ normalized B splines to construct \( \{ \phi_i \} \) with 91 components. Since the bids are exponential-like distributed, we take an exponential with hazard rate \( \mu \) for \( \tilde{f}(\cdot | \mu) \) in (2.8). We control the tail behavior of the valuation density. (See 4.4.) The Adaptive Metropolis algorithm of Haario, Saksman, and Tamminen (2001) simulates the posterior.\(^{19}\) Appendix B provides all the details for the specification and implementation.

To obtain a valuation density estimate that preserves density affiliation, we compute the posterior mean of log density, i.e., \( E[\log f(v_1, v_2 | \theta) | y] \). Then, we use the exponential of this as our valuation density estimate. This posterior mean is consistently estimated by the AM output. In order to check the goodness-of-fit, it may be useful to construct a bid density estimate. We approximate the bid density using simulation; generate many auctions for each parameter value from the AM output \( \theta_1, \ldots, \theta_S \) and run a kernel method over these simulated bids. \( \hat{g}(\cdot | \theta_s) \). Then, we use the average of these kernel estimates as our bid density estimate.

Figure 2.6 summarizes the estimation results. Panel (a) plots the marginal bid density estimate (solid) and a 90% credible band (dotted) along with the posterior distribution of bid density (shaded). This bid density estimate fits the data fairly well and its narrow credible band indicates high precision. But, it does not closely fit the tail because we control it. Nevertheless, panel (c), which shows 1,000 level curves of the joint density estimate, explains some negative correlation pattern over the tail area. Similarly, panel (b) plots the posterior distribution of marginal valuation density. Since a bidder bids less than his valuation, the valuation densities are

\(^{19}\)See Appendix B for the Adaptive Metropolis algorithm.
spread out toward large values. The credible band is also narrow, which implies that the posterior is very condensed around its mean. Panel (d) is the 1,000 level curve contour of joint predictive valuation density.

2.4.3 Auction Design with Bayes Rule

This subsection uses the posterior distribution of the structural parameters to compute revenue-maximizing reserve prices. Riley and Samuelson (1981) show that when the seller’s valuation is \( v_0 \), a reserve price \( \rho_R(\theta) := \rho^* \) solving \( \rho^* = v_0 + \frac{1 - F(\rho^*|\theta)}{f(\rho^*|\theta)} \) maximizes the seller’s expected revenue (payoff). Paarsch (1997) estimates an optimal reserve price by \( \rho_R(\hat{\theta}) \) with a consistent estimate \( \hat{\theta} \). Similarly, Li, Perrigne, and
Vuong (2003) nonparametrically estimate the bid distribution and then derive an optimal price for this estimated bid distribution. This decision procedure is called a ‘plug-in’ rule because it makes a decision regarding an estimate as true.

However, the plug-in rule does not consider the payoff structure and parameter uncertainty. Consider a seller whose payoff increases slowly up to the optimal reserve price, but then drops sharply thereafter. For a given sample, plug-in rule must either overestimate or underestimate the true optimal price 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).

Chapter 1 introduces a Bayesian decision theoretic framework to auction design. Let $\Pi(\theta, \rho)$ denote the seller’s payoff for reserve price $\rho$ under $\theta$ and $A$ be a set of all feasible reserve prices. The Bayes rule selects a Bayes action defined by

$$\rho_B(y) := \arg\max_{\rho \in A} \int \Pi(\theta, \rho)p(\theta|y)d\theta$$ (2.13)

Observe that the posterior systematically quantifies parameter uncertainty and the Bayes rule considers the average payoff structure that is weighted by the posterior. Chapter 1 discusses the optimality of (2.13) as a decision rule. We revisit the auction design problem for the OCS wildcat auctions using the Bayes rule. Assume $v_0 = 0$ for simplicity. Then,

$$\Pi(\theta, \rho) = E\left[\beta(v_{(2)}|\rho, \theta) \cdot 1\left(v_{(2)} > \rho\right) | \theta\right]$$ (2.14)

where $v_{(1)} < v_{(2)}$. We consistently estimate (2.14) using Monte Carlo for each $\theta$ and $\rho$. Then, we approximate (2.13) using the MCMC output.

We find $\hat{\rho}_B(y) = 4.62$. Thus, the posterior expected revenue-maximizing reserve price is $462 per acre. We evaluate the predictive revenue at this price value; $\int \Pi(\theta, \hat{\rho}_B(y))p(\theta|y)d\theta = 294.24$. Hence, the seller’s predictive revenue for a typical tract of 5,000 acres would be $1,471,197 (= 294.24 \times 5,000)$. We find that the predictive revenue of $15, the actual price, is $1,208,783. Therefore, our choice
for the reserve price would increase the revenue by $262,414. This revenue gain is significant, since hundreds of tracts are offered annually.

We also find that $\rho_R(\hat{\theta}_B) = 4.47$ and its predictive revenue is $1,470,872$; the plug-in rule turns out to well approximate the Bayes rule. It must be the case that either the payoff is about symmetric or parameter uncertainty is negligible. In our case, the former is the answer. Figure 2.7 plots the predictive payoff (plain) along with its 90% credible band (dashed). The predictive revenue curve is roughly symmetric about its maximum, while parameter uncertainty is still large (wide credible band). Note that the predictive revenue for zero reserve price is very close to the sample mean of winning bids, 2.243, indicating the accuracy of our analysis. Note that our policy prescriptions are dramatically different from the estimate $273$ of Li, Perrigne, and Vuong (2003). This suggests that the contribution of the additional shape restrictions is significant.

Finally, though many researchers have pointed out that the actual price of $15$ is too low to be an effective screening device the U.S government still employs a very low reserve price. Our finding in this subsection supports the opinion of the
previous researchers by suggesting an even higher reserve price.

2.5 Conclusion

Our methodology formally considers the additional theoretical shape restrictions arising from economic theory, such as bidding monotonicity and density affiliation. These are not fully exploited by the current nonparametric methods. We directly parametrize the valuation density allowing for a flexible specification so that bidding monotonicity is satisfied. We also restrict the parameter space so that the posterior selects only affiliated densities. We simulate the likelihood to handle such a rich specification. Since our method exploits more information, it can give more precise inference especially when the sample is small. We reanalyze the sample from the OCS auctions that Li, Perrigne, and Vuong (2003) investigate. While the density estimate fits the data very well, our choice of reserve price is drastically different from the previous results. This indicates that the additional theoretical shape restrictions play an significant role.

We develop our framework under the APV assumptions. Li, Perrigne, and Vuong (2003) argues that these assumptions are quite reasonable for the sample from the OCS wildcat auctions. However, in order to provide more convincing policy recommendations, it would be useful to consider more features of the OCS auctions, such as endogenous entry and multi-unit auctions. That is, there are many auctions with more than or less than just two bidders. Not all the potential bidders participate in every auction, while reserve prices are very low. If bidders make entry decisions first, the theoretical optimal reserve price will be dramatically different.\textsuperscript{20} Relatedly, the government holds many auctions at a given date, creating another strategic situation among the bidders. Hence, it is necessary to consider these issues in order to provide more convincing policy analysis. We plan to address these issues extending our method developed in this paper. Some economists have argued that the pure common value better approximates the OCS wildcat auctions than the APV. From

\footnote{\textsuperscript{20}See Levin and L (1994) and Moreno and John (2008).}
the policymaker’s point of view, it would be important to obtain a policy advice considering the risk under all possible alternative models. Hence, we also plan to develop a decision method that incorporates alternative models into the decision process.
CHAPTER 3

SIMPLE APPROXIMATIONS FOR BAYESIAN AUCTION DESIGN

3.1 Introduction

Auction design has a pragmatic importance in reaching the objective of the social planner (e.g., revenue maximization) who runs an auction to allocate economic resources. Naturally, many economists have explored this subject theoretically and empirically. Auction theory suggests optimal auction design for a given distribution of bidders’ valuations or costs. Empirical studies uncover this underlying distribution from observed bids and produce policy recommendations using the point estimates for important auctions such as timber sales and oil and gas leases. See Paarsch (1997) and Li, Perrigne, and Vuong (2003). Recently, Kim (2008) proposes the use of the Bayesian decision method for choosing a reserve price. The Bayes rule formally considers the payoff structure and parameter uncertainty, thereby producing higher average payoffs than the previous plug-in rule that uses the point estimate as if it is the true parameter value.

However, a formal Bayesian analysis may not always be feasible or practical. In general, prior elicitation has been recognized as a hard problem for many applications and posterior simulation can be computationally extensive for highly parametrized models. Researchers are often forced to employ computationally practical specifications for both prior and likelihood. Thus, it would be useful if we can closely approximate the Bayes rule of the original problem by a simple alternative.

Motivated by this, we propose simple approximate methods for Bayesian decision making in auction design problems.¹ We approximate the posterior using large sample results, and choose a reserve price to maximize the approximate posterior

¹Hirano and Porter (2009) develop approximately optimal treatment rules for similar motivations.
payoff.

We discuss first and second price auctions separately because the associated statistical models have different limits. First, we employ a Gaussian model to approximate the posterior for second price auctions because the associated statistical models typically satisfy the usual local asymptotic normality (LAN) conditions. In particular, we use the sampling distribution of the maximum likelihood estimator (MLE) as the Bernstein-von Mises theorem suggests. Both the exact and approximate Bayes rules are shown to be consistent to the true revenue maximizing reserve price. Second, For first price auctions, the implied statistical models violate the LAN conditions because the support of bid data depends on the parameter of interest. Hirano and Porter (2003) show that for such non-regular cases statistical models weakly converges to a shifted exponential model. Thus, we approximate the posterior using a shifted exponential model. A Monte Carlo study shows that our approach closely mimics the exact Bayes rule even for fairly small samples for both regular and non-regular cases, but the plug-in rule can be substantially different, especially for first price auctions.

As indicated above, our approach is based on the previous studies investigating large sample properties of the posterior analysis. Most papers in the literature focus on the asymptotic behavior of various Bayesian estimators and test statistics. See Ghosh, Ghosal, and Samanta (1992) and references therein. However, our approach differs from these studies because we are directly concerned about the auction design problem of the social planner with a particular objective such as revenue maximization for future auctions.

3.2 Auction Models and Bayesian Decision Framework

We consider procurement auctions for simple exposition. A social planner assigns a project to one of $m$ risk neutral, expected profit maximizing potential bidders. Each bidder $i$ observes his own cost $c_i \in \mathbb{R}_+$ and bid $b_i$ without knowing his rivals’ costs. Specifically, the bid density function for first price auctions with exponential cost density has a closed form expression. This simplifies the Monte Carlo experiments in section 4.
The costs $c_1, \ldots, c_m$ are identically, independently distributed as $F_\theta$ with density $f_\theta$. Let $\rho$ denote the publicly announced reserve price. The bidder with the lowest bid $b_{(1)} \leq \rho$, undertakes the project. In a first price auction, the winner gets paid his own bid. In a second price auction, he receives $\min \{ \rho, b_{(2)} \}$ where $b_{(2)}$ is the second lowest bid. If no bid is lower than $\rho$, the social planner fails to assign bearing social cost $c_0$.

We assume that every bidder $i$ with $c_i \leq \rho$ follows the Bayesian Nash equilibrium strategy,

$$
\beta_I(c|\theta, \rho) = c + \int_c^\rho \left( \frac{1 - F_\theta(y)}{1 - F_\theta(c)} \right)^{m-1} dy
$$

(3.1)

for first price auctions and

$$
\beta_{II}(c|\theta, \rho) = c
$$

(3.2)

for second price auctions.\footnote{See McAfee and McMillan (1992) for proofs among many others. They derive these results for the usual high price auctions, but the proofs have analogues.} For first price auctions, the support for the implied bid distribution depends on the parameter value, $b(\theta, \rho) := \beta_I(0|\theta, \rho)$. Because of this, the implied statistical models are non-regular. The social planner’s payoff (expected profit) can be written as

$$
\Pi(\theta, \rho) = \int_0^\rho (c_0 - \beta_I(c|\theta, \rho)) m (1 - F_\theta(c))^{m-1} f_\theta(c) dc
$$

(3.3)

which is common for both auction formats by the revenue equivalence principle.\footnote{We assume that the social planner’s objective to maximize the expected profit, but we can use other objective functions that represent the social planner’s interest.}

Let $\rho_R(\theta) := \arg \max_\rho \Pi(\theta, \rho)$. Note $\rho_R(\theta)$ solves the first order condition $c_0 = \rho_R(\theta) + F(\rho_R(\theta)|\theta)/f(\rho_R(\theta)|\theta)$. The social planner would choose $\rho_R(\theta)$ regardless of the shape of (3.3) if he knew $\theta$.

We are interested in the decision problem of the social planner who is uncertain about $\theta$. We assume that he observes bid sample $z_n$ of size $n$ from past homogeneous
and uncorrelated $T$ auctions with a non-binding reserve price and $m$ bidders. ($n = mT$.) He now wishes to maximize the future payoff using $z_n$. For this problem, Paarsch (1997) and Li, Perrigne, and Vuong (2003) propose to choose $\rho_R(\hat{\theta})$ with $\hat{\theta} \xrightarrow{P} \theta_0$. But, as Kim (2008) points out, this plug-in rule uses $\hat{\theta}$ as if it is $\theta_0$. That is, it does not consider the structure of (3.3) and uncertainty about $\theta$, which can be important for the social planner’s decision making. The payoff depends on $\theta$, and is typically asymmetric about $\rho_R(\theta)$. In particular, consider a payoff structure increasing sharply up to $\rho_R(\theta)$ and decreasing slowly thereafter as in Figures 3.1 and 3.4. For this payoff structure, the social planner should avoid underestimation (choosing a reserve price smaller than $\rho_R(\theta)$) more than overestimation, formally considering the degree of asymmetry and the uncertainty about $\theta$.

Kim (2008) employs the Bayesian decision framework to systematically consider the payoff structure and parameter uncertainty. Let $P_{\theta|z_n}$ denote the posterior distribution and $\mathcal{A}$ be the set of all feasible reserve prices. Then, he proposes to choose the Bayes action,

$$
\rho_B(z_n) := \arg \max_{\rho \in \mathcal{A}} \int \Pi(\theta, \rho) dP_{\theta|z_n}(\theta)
$$

(3.4) represents a rational decision maker’s behavior in the sense of Savage (1954) and Anscombe and Aumann (1963). Moreover, the Bayes rule, a decision rule choosing (3.4) for every realization of data $z_n$, is shown to be optimal under the average risk principle, a widely used frequentist decision criteria. Kim (2008) provides Monte Carlo evidences that the Bayes rule produces higher average payoffs than the plug-in rule when the payoff is fairly asymmetric and there is a large amount of parameter uncertainty.

However, the exact posterior analysis would not always be feasible or practical. For example, prior elicitation has been recognized as a challenging problem for many applications. Even if we obtained a prior that well represents the decision maker’s subjective beliefs or the effect of prior on decision is almost negligible due to a large amount of data, the posterior simulation could be computationally very expensive.
For this reason, if an alternative decision procedure gives very similar answers to the exact Bayes rule, it would be useful whenever the exact Bayes rule is hard to implement.

3.3 Second Price Auctions: Regular Models

For data from second price auctions, we may regard \( z_n \) as draws from \( F_\theta \) due to (3.2). Thus, if \( F_\theta \) satisfies LAN, so does the associated statistical model. Then, the Bernstein-von Mises theorem suggests that we may approximate the posterior distribution using the sampling distribution of the MLE, which we denote by \( \hat{\theta}_n \). Let \( \Phi_{\mu,\Sigma} \) and \( \phi_{\mu,\Sigma} \) be a normal distribution and its density with mean \( \mu \) and covariance matrix \( \Sigma \). We propose to approximate (3.4) using

\[
\tilde{\rho}_B(z_n) := \arg \max_{\rho \in A} \int \Pi(\theta, \rho) d\Phi_{\hat{\theta}_n(nI_0)^{-1}}(\theta)
\]

where \( I_0 \) is the information matrix at \( \theta_0 \). In this section, we show that (3.4) and (3.5) are consistent, and provide a Monte Carlo evidence that the approximate Bayes rule closely mimics the exact rule for small samples.

3.3.1 Consistency

We first show that \( \rho_B \xrightarrow{p} \rho_0 \) based on the posterior consistency result of Schwartz (1965) and that \( \tilde{\rho}_B \xrightarrow{p} \rho_0 \) using the Bernstein-von Mises theorem. Let \( (\Omega, \mathcal{B}, P_\theta) \) be a probability space where \( \mathcal{B} \) is the Borel \( \sigma \)-algebra on \( \Omega \) and \( P_\theta \) is a probability measure indexed by \( \theta \in \Theta \), a compact subset of \( \mathbb{R}^k \). We have an \( n \) dimensional random vector \( Z_n : \Omega \to \mathbb{R}^n_+ \) for \( n > 0 \). We assume that \( Z_n \) is a collection of \( n \) independent random variables and each component is distributed as \( F_\theta \) under \( P_\theta \). Let \( P_\theta \) and \( P_{\theta|Z_n} \) be the prior and posterior probability measures on \( \Theta \).

**Theorem 3.3.1** (Schwartz (1965)) Suppose that there exists a uniformly consistent estimator \( \hat{\theta}(Z_n) \) such that, for every \( \varepsilon > 0 \), \( \sup_{\theta \in \Theta} P_{\theta}(\|\hat{\theta}(Z_n) - \theta\| \geq \varepsilon) \to 0 \) and also such that, for \( \eta > 0 \), \( P_\theta \left\{ \theta : \int f(z|\theta_0) \log \frac{f(z|\theta)}{f(z|\theta_0)} dz < \eta \right\} > 0 \). Then, for every
neighborhood $U$ of $\theta_0$, $P_{\theta|Z_n}(U) \xrightarrow{a.s} 1$ under the law determined by $\theta_0$.

Now, we assume that

A1. $\Pi$ is bounded and continuous over $\Theta \times A$.

A2. $A$ is a compact subset of $\mathbb{R}$.

A3. $\Pi(\theta_0, \rho)$ is uniquely maximized at $\rho_0 \in \text{Int} A$

A4. $\frac{\partial}{\partial \rho} \Pi(\theta, \rho)$ is uniformly bounded over $\Theta$ and for all $\rho \in A$.

Proposition 3.3.2 Under $A1$-$A4$ and the conditions for the Schwartz theorem, $\rho_B(Z_n) \xrightarrow{D} \rho_0$.

We find using de Finetti notation substantially simplifies the expectation expressions in the proofs below.\(^5\) That is, $GY$ denotes the expected value of the random quantity $Y$ under the probability distribution $G$.

**Proof** Let $U$ be a neighborhood of $\theta_0$ and $D_U := \{ w \in [\Omega : P_{\theta|Z_n(w)}(U) \rightarrow 1 \}$ where $w$ denotes a generic sample sequence. Then, A1 implies that, for each $\rho \in A$, the set $\{ \Pi(\theta, \rho) : \theta \in U \}$ is a neighborhood of $\Pi(\theta_0, \rho)$ and $P_{\theta|Z_n(w), w \in D_U} \Pi(\theta, \rho)$ is contained in this set for sufficiently large $n$. Under the conditions of Schwartz theorem, $Pr_{\theta_0}(D_U) = 1$ with an arbitrary $U$. Thus, $P_{\theta|Z_n} \Pi(\theta, \rho) \xrightarrow{a.s} \Pi(\theta_0, \rho) \forall \rho \in A$.

If this convergence is uniform with respect to $\rho$, the proof follows from theorem 2.1 of Newey and McFadden (1994) under A1-3. A4 implies that there exists $M < \infty$ such that, for all $\rho, \rho' \in A$, $|\Pi(\theta, \rho) - \Pi(\theta, \rho')| \leq |\rho - \rho'| M$ for every $\theta \in \Theta$.

So, $P_{\theta|Z_n} |\Pi(\theta, \rho) - \Pi(\theta, \rho')| \leq |\rho - \rho'| M$. Since $|P_{\theta|Z_r} \Pi(\theta, \rho) - P_{\theta|Z_r} \Pi(\theta, \rho')| \leq P_{\theta|Z_r} |\Pi(\theta, \rho) - \Pi(\theta, \rho')|$, we have $|P_{\theta|Z_r} \Pi(\theta, \rho) - P_{\theta|Z_r} \Pi(\theta, \rho')| \leq |\rho - \rho'| M$, satisfying the stochastic Lipschitz condition. Then, the uniform convergence follows by lemma 2.9 of Newey and McFadden (1994).

To state the Bernstein-von Mises theorem, we use the local alternative $h := \sqrt{n}(\theta - \theta_0)$ and let $P_{h|Z_n}$ denote the posterior distribution for $h$. Let $\Delta_{n,0} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{0}^{-1} s_{0,i}$ with $s_{0,i}$ the score function for the $i$-th observation at $\theta_0$.

\(^5\)See Section 1.4 of Pollard (2002) for more detail.
Theorem 3.3.3 (Bernstein (1917) and von Mises (1931)) Suppose that there exists a uniformly consistent estimator and $F_{\theta}$ is differentiable in quadratic mean at $\theta_0$ with a nonsingular information matrix $I_0$. Moreover, $P_{\theta}$ is absolutely continuous around $\theta_0$ with a continuous positive density at $\theta_0$. Then, the total variation distance between $P_{h|Z_n}$ and $\Phi_{\Delta_{n,0},I_0^{-1}}$ converges in probability to zero under $\theta_0$.

Theorem 3.3.3 leads us to use (3.5) to approximate (3.4). Note that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can replace $\Delta_{n,0}$ in the theorem because they are asymptotically equivalent. Since the total variation distance is invariant under location and scale changes, the theorem implies that the total variation distance between $\Phi_{\hat{\theta}_n,(nI_0)^{-1}}$ and $P_{|Z_n}$ also converges in probability to zero. Thus, we use $\Phi_{\hat{\theta}_n,(nI_0)^{-1}}$ in (3.5) to approximate the true posterior.

Proposition 3.3.4 Under $A1 \sim \mathcal{L}$ and the conditions for the Schwartz theorem and the Bernstein-von Mises theorem, $\tilde{P}_B(Z_n) \xrightarrow{P} \rho_0$.

Proof It is enough to show that the approximate posterior payoff $\Phi_{\hat{\theta}_n,(nI_0)^{-1}}\Pi(\theta, \rho) \xrightarrow{P} \Pi(0, \rho)$. Then, the result follows by the same logic as the proof of Proposition 3.2. We may assume $\Pi(\theta, \rho) < 1/2$ for every $\theta \in \Theta$ and $\rho \in \mathcal{A}$ under $A1$ without loss of generality. Now, consider the difference between the exact posterior payoff and its approximation. That is, for each $\rho \in \mathcal{A}$,

$$|P_{\theta|Z_n}\Pi(\theta, \rho) - \Phi_{\hat{\theta}_n,(nI_0)^{-1}}\Pi(\theta, \rho)| = |\int \Pi(\theta, \rho) \left\{P_{\theta|Z_n}(\theta) - \phi_{\hat{\theta}_n,(nI_0)^{-1}}(\theta)\right\} d\theta|$$

$$\leq \int \Pi(\theta, \rho)|p_{\theta|Z_n}(\theta) - \phi_{\hat{\theta}_n,(nI_0)^{-1}}(\theta)|d\theta$$

$$\leq \int \frac{1}{2} |p_{\theta|Z_n}(\theta) - \phi_{\hat{\theta}_n,(nI_0)^{-1}}(\theta)|d\theta$$

$$= \frac{1}{2} \int |p_{h|Z_n}(h) - \phi_{\sqrt{n}(\hat{\theta}_n - \theta_0),I_0^{-1}}(h)|dh$$

(change of variables)

---

6 $F_0$ is said to be differentiable in quadratic mean at $\theta_0$ if there exists a function $s : \mathbb{R}_+ \to \mathbb{R}^k$ such that $\int \left[dF_{\theta_0+h}^{1/2}(z) - dF_{\theta_0}^{1/2}(z) - \frac{1}{2}s(z)dF_{\theta_0}^{1/2}(z)\right]^2 = o(||h||^2)$ as $h \to 0$.

7 Note that the total variation distance $\|P_{h|Z_n} - \Phi_{\Delta_{n,0},I_0^{-1}}\| := \sup_{B \in \mathcal{B}} |P_{h|Z_n}(B) - \Phi_{\Delta_{n,0},I_0^{-1}}(B)| = \frac{1}{2} \int |p_{h|Z_n}(h) - \phi_{\Delta_{n,0},I_0^{-1}}(h)|dh$. For the proof of Theorem 3.3.3, see Van der Vaart (1998) 10.2.

8 See Van der Vaart (1998) 8.14 and Ch.10.
under $\theta_0$. Since $P_{\theta_0|Z_n}\Pi(\theta, \rho) \xrightarrow{p} \Pi(\theta_0, \rho)$, so does $\Phi_{\theta_0, (nI_0)^{-1}} \Pi(\theta, \rho)$.

Thus, the approximate Bayes rule would also give the correct answer for a large sample. Therefore, we would not lose too much from using it if we have a sufficiently large sample. However, the consistency by itself is not informative on small sample quality of the approximate Bayes rule (3.5). Hence, we conduct a Monte Carlo study in the next subsection.

3.3.2 Monte Carlo Study

We compare the exact Bayes rule (3.4), approximate Bayes rule (3.5), and also plug-in rule $\rho_R(\hat{\theta}_n)$ for small samples using a Monte Carlo study. We employ one thousand replications. For each replication, we generate a new sample of size $n = m \times T$ from a fixed $F_{\theta_0}$. Then, we implement the alternative decision rules. We set $m = 3$.

First, we consider a very simple model where costs are exponentially distributed, i.e.,

$$f_\theta(c) = \theta \exp(-\theta c) 1(c \geq 0), \quad \theta > 0$$

with $\theta_0 = 1$ and $c_0 = F^{-1}(0.95|\theta_0) \approx 2.9957$. Figure 3.1 displays (3.6) and the associated payoff function, which is highly asymmetric about $\rho_0 \approx 1.0726$. As discussed above, for this type of payoff structure, a rational decision maker would try to avoid choosing lower reserve price than $\rho_0$. Hence, we expect that the Bayes rule can choose fairly large reserve prices when there is a large amount of parameter uncertainty.

Figure 3.2 displays the Monte Carlo results for $T \in \{1, 2, 4, 10\}$. For each panel, the plain line and the dashed line are the kernel density estimates of the sampling
distributions of $\rho_B(Z_n)$ and $\tilde{\rho}_B(Z_n)$ under $\theta_0$, respectively.\footnote{To implement $\rho_B$, we use a Gamma prior with parameters $\alpha$ and $\beta$ such that prior mean is $\alpha/\beta$. Then, the posterior is also gamma with the updated parameters $\alpha + T$ and $\beta + Tb$ with sample mean $\bar{b}$. To compute $\tilde{\rho}_B$, we use a reparametrization $\delta := \log \theta$ so that its support coincides with the normal approximation of the posterior. Note that $\delta_{ML} = -\log \bar{b}$ and $\sqrt{T}(\delta_{ML} - \delta_0) \xrightarrow{d} \Phi_{0,1}$.} Even when the sample has only four auctions, $\tilde{\rho}_B$ behaves quite similarly to $\rho_B$, and they are almost identical for the ten auction case. Moreover, the distributions become more condensed around $\rho_0$ as sample size grows. Lastly, the distributions are bimodal for small samples like one or two auctions. This represents the social planner’s rational behavior in the presence of a large amount of parameter uncertainty. When $\theta$ is very uncertain, it is very likely to choose a reserve price much different from $\rho_R(\theta_0)$. Then, he should be very aversive to choosing a small reserve price because choosing a reserve price a little bit smaller than $\rho_R(\theta_0)$ can result in a much greater payoff loss than very large reserve prices. See Figure 3.1.

Similarly, Figure 3.3 compares the distributions of $\rho_B(Z_n)$ (solid line) and the plug-in rule $\rho_R(\hat{\theta}(Z_n))$ (dashed line). Since the plug-in rule does not consider the payoff structure and parameter uncertainty, it does not choose very high reserve prices like the (approximate) Bayes rule. But, as sample size grows, it gets quickly close to the exact Bayes rule because the model is very simple and the parameter
uncertainty disappears quickly.

Now, we consider a flexible model. Let \( \{ \psi_i \} \) denote a sequence of functions on \([0, 1]\) such that we can approximate any function on \([0, 1]\) using their linear combination \( \sum_j \theta_j \psi_j \) for some real numbers \( \{ \theta_i \} \). We employ the cost density

\[
f_\theta(c) = \exp \left( \sum_{j=1}^{k} \theta_j \psi_j(c) \right) / c(\theta)
\]

(3.7)

where \( c(\theta) = \int_0^1 \exp \left( \sum_{j=1}^{k} \theta_j \psi_j(u) \right) du \). We use the Legendre polynomial basis functions with \( k = 5 \) and \( \theta_0 \) that gives the density function and payoff function of
Observe that the payoff function is also highly asymmetric so that underestimation would result in much greater loss than overestimation. Since the model is flexibley parametrized, larger sample sizes would be required for a good approximation. Thus, we use $T \in \{5, 10, 20, 30\}$.

Figure 3.5 displays the kernel density estimates of the exact (solid) and approximate (dashed) Bayes rules. Since there are a large amount of parameter

Specifically, $\psi_j(u) := \sqrt{2j + 1}\tilde{\psi}_j(2u - 1)$ with $\tilde{\psi}_j(x) := \frac{d^j}{dx^j}((x^2 - 1)^j/(2^jj!))$. To implement (3.4), we use prior density $p_\theta(\theta) = \prod_{j=1}^{k} \phi_{0,\tau/2j}(\theta_j)$ for $\tau > 0$. Note that this prior controls the smoothness by squeezing out the variances of the coefficients for the noisy $\psi$’s. Then, we simulate the posterior using the Adaptive Metropolis algorithm of Haario, Saksman, and Tamminen (2001). For (3.5), we need to compute $I_0$. For our specification, it can be shown that $I_0 = \text{Cov}_{\theta_0}(\psi_1(c), \ldots, \psi_k(c))$. Thus, we can evaluate $I_0$ using random draws from (3.7).
uncertainty, the approximate method chooses quite different reserve prices than the exact Bayes rule for small samples such as five auctions and ten auctions. However, the discrepancy fades out as sample size grows, and finally both rules behave similarly for samples with thirty auctions. On the other hand, for this flexible model, the plug-in rule appears to be a poor approximation for the Bayes rule. See Figure 3.6.

3.4 First Price Auctions: Non-regular Models

The normal approximation works in the previous section because the statistical models for second price auctions typically satisfy standard regularity conditions. But, the observed bid data for first price auction models have a parameter dependent support, i.e., $b \geq b(\theta) > 0$, violating the differentiability in quadratic mean requirement. Thus, the normal approximation would not be valid. For this type of model, Hirano and Porter (2003) observe that the sequence of the likelihood ratio processes weakly converges to the likelihood ratio process of a shifted exponential model and the sequence of standardized posterior distributions weakly converges to the posterior associated with this limiting model under certain conditions.

The first price procurement auction model with the exponential cost density (3.6)
is a special case of their result. The bidding function (3.1) with \( \rho = \infty \) becomes 
\[ \beta(c|\theta) = c + \frac{1}{\theta(m-1)} \]. Since the bid is a monotone transformation of cost, using change of variables we obtain the implied bid density as follows.

\[ g_\theta(b) = \theta \exp\left(-\theta b + \frac{1}{m-1}\right) 1\left\{ b \geq \frac{1}{\hat{\theta}(m-1)} \right\} \]

It turns out that we can approximate the exact posterior density \( p_{\theta|z_n} \) using a shifted exponential density given by

\[ \tilde{p}_{\theta|z_n}(\theta) = \frac{n}{\hat{\theta}_n(m-1)} \exp\left\{ -\frac{n}{\hat{\theta}_n(m-1)} (\theta - \hat{\theta}_n) \right\} 1\left\{ \theta > \hat{\theta}_n \right\} \]  \hspace{1cm} (3.8)
with $\hat{\theta}_n := [b(1)(m - 1)]^{-1} \overset{P}{\to} \theta_0$. Then, we approximate the Bayes rule by

$$\tilde{\rho}_B(z) := \arg\max_{\rho \in \mathcal{A}} \int \Pi(\theta, \rho)\tilde{p}_{\theta|z_n}(\theta)d\theta$$

(3.9)

using (3.8) in place of the original posterior density. We will sketch the derivation of (3.8) below.

Note that $\hat{\theta}_n \leq \theta_0$ for every $z_n$ under $\theta_0$. Thus, $\theta_0$ is located at the rightmost boundary of the sampling distribution of $\hat{\theta}_n$. On the other hand, the approximate posterior (3.8) is a shifted exponential density with mean $\hat{\theta}_n(1 + (m - 1)/n)$ which is strictly greater than $\hat{\theta}_n$ for all $n > 0$ and consistent to $\theta_0$. Hence, the sampling distribution of $\hat{\theta}_n$ is substantially different from (3.8).
We conduct a similar Monte Carlo study to compare the exact rules and the approximate rules. Figure 3.7 plots the results of the Monte Carlo comparisons between the exact and approximate Bayes rules. We find that the approximation quality of $\tilde{\rho}_B$ is even better than the second price auction (regular model) case: even when we observe only one auction, $\tilde{\rho}_B$ closely mimics $\rho_B$.\textsuperscript{12} This is because the convergence rate for these non-regular models, $n_i$, is faster than the usual rate $\sqrt{n}$ of the regular models.

However, the plug-in rule does not well approximate the Bayes rule as appears

\textsuperscript{12} For the exact Bayesian analysis, we employ the Gamma prior with $(\alpha, \beta)$ that we use in the previous section. The resulting posterior is a truncated Gamma distribution with density $p(\theta|z) \propto \theta^{(\alpha+T)-1} \exp\left(-\left(\beta + T\bar{b}\right)\theta\right) 1\{\theta \geq [b(1)(m-1)]^{-1}\}$.
Figure 3.8: Exact Bayes Rule vs. Plug-in Rule (Non-regular)

In particular, as sample size grows, the shape of distribution of the plug-in rule gets very similar to the one of the Bayes rule, but it is shifted to the right. The bidders’ costs under the point estimate $\hat{\theta}$ are larger than the true costs on average because $\hat{\theta} \leq \theta_0$ for every $z_n$ under $\theta_0$. Therefore, $\rho_R(\hat{\theta}) \geq \rho_R(\theta_0)$ for every $z_n$ under $\theta_0$, which explains the fact that the distribution of $\rho_R(\hat{\theta})$ is shifted to the right. This suggests that the plug-in rule for first price auctions can choose substantially different reserve prices than the Bayes rule.

Finally, we sketch the derivation of (3.8). We use a local parameter $h := n(\theta - \theta_0)$
and derive the asymptotic posterior for $h$. Consider the likelihood ratio process

$$
\prod_{i=1}^{n} \frac{g_{\theta_0 + h/n}(b_i)}{g_{\theta_0}(b_i)} = \left( \frac{\theta_0 + \frac{h}{n}}{\theta_0} \right)^n \cdot \exp \left\{ -\left( \theta_0 + \frac{h}{n} \right) \cdot n \bar{b} \right\} \cdot 1 \left\{ \bar{b}_{(1)} \geq \frac{1}{(\theta_0 + \frac{h}{n})(m-1)} \right\} \\
= \left( \frac{n\theta_0 + h}{n\theta_0} \right)^n \cdot \exp(-\bar{b}h) \cdot 1 (-W_n < h)
$$

where $W_n := -n \left( \frac{1}{b_{(1)}(m-1)} - \theta_0 \right)$. As $n \to \infty$,

$$
\left( \frac{n\theta_0 + h}{n\theta_0} \right)^n = \exp \{ n \left[ \log(n\theta_0 + h) - \log(n\theta_0) \right] \} \to \exp \left\{ \frac{h}{\theta_0} \right\}
$$

$$
\exp(-\bar{b}h) = \exp \left\{ - \left( \bar{c} + \frac{1}{\theta_0(m-1)} \right) h \right\} \to \exp \left\{ - \left( \frac{1}{\theta_0} + \frac{1}{\theta_0(m-1)} \right) h \right\}
$$

and $W_n \xrightarrow{d} W \sim f_W(w) = \frac{1}{\theta_0(m-1)} \exp \left\{ -\frac{w}{\theta_0(m-1)} \right\} 1(w > 0)$. Therefore,

$$
\prod_{i=1}^{n} \frac{g_{\theta_0 + h/n}(b_i)}{g_{\theta_0}(b_i)} \xrightarrow{d} \exp \left\{ - \frac{h}{\theta_0(m-1)} \right\} 1(-W < h)
$$

Then, it can be shown that the sequence of posterior densities converges to a shifted exponential density as follows,

$$
p_{h|Z_n}(h) = \frac{p_\theta \left( \theta_0 + \frac{h}{n} \right) \prod_{i=1}^{n} \frac{g_{\theta_0 + h/n}(b_i)}{g_{\theta_0}(b_i)} \cdot d\theta}{\int p_\theta \left( \theta_0 + \frac{h}{n} \right) \prod_{i=1}^{n} \frac{g_{\theta_0 + h/n}(b_i)}{g_{\theta_0}(b_i)} \cdot d\theta} \xrightarrow{d} \frac{p_\theta(\theta_0) \exp \left\{ -\frac{h}{\theta_0(m-1)} \right\} 1(-W < h)}{p_\theta(\theta_0) \int \exp \left\{ -\frac{h}{\theta_0(m-1)} \right\} 1(-W < h) dh} = \frac{1}{\theta_0(m-1)} \exp \left\{ - \frac{h + W}{\theta_0(m-1)} \right\} 1(-W < h)
$$

which can be seen as a posterior density with one observation $W$ and a flat prior.

Now, we construct a sequence of approximate posterior densities weakly converging

\text{\footnotesize \textsuperscript{13} For each } w \in \mathbb{R}_+, \text{ Pr}_{\theta_0}(W_n < w) = Pr_{\theta_0} \left\{ -n \left( \frac{1}{b_{(1)}(m-1)} - \theta_0 \right) < w \right\} = Pr_{\theta_0} \left\{ b_{(1)} < \frac{n}{\theta_0(m-1)} \right\} = 1 - Pr_{\theta_0} \left\{ c > \frac{w}{\theta_0(n\theta_0-w)(m-1)} \right\}^n = 1 - \exp \left\{ -\frac{w}{\theta_0(m-1)} \right\}.\text{\footnotesize \textsuperscript{13}}
the same limit as above by replacing \( W \) by \( W_n \) and \( \theta_0 \) by \( \hat{\theta} \). That is,

\[
\tilde{p}_{n|Z_n}(h) := \frac{1}{\theta(m-1)} \exp \left\{ -\frac{h + W_n}{\theta(m-1)} \right\} 1(-W_n < h)
\]

\[
\frac{d}{\theta_0(m-1)} \exp \left\{ -\frac{h + W}{\theta_0(m-1)} \right\} 1(-W < h)
\]

Then, we obtain (3.8) using change of variables for \( h = n(\theta - \theta_0) \) and \( W_n := -n(\hat{\theta} - \theta_0) \).

3.5 Conclusion

We propose simple approximation methods for the Bayesian decision procedures for auction design. The approximate Bayes rule for second price auctions (regular statistical models) is shown to be consistent to the true revenue maximizing reserve price under certain conditions. The Monte Carlos study shows that our method have a good approximation quality for fairly small samples for both regular and non-regular cases, but the plug-in rule does not well approximate. Therefore, our method would be a simple and convenient alternative to the Bayesian decision framework which is useful for many policy problems. However, this paper has a limited scope in its analysis and applicability. We leave, as a future work, some important elements of this line of research such as developing a general approximation framework for first price auctions and exploring asymptotic Bayes risk properties of the approximate rules.
A.1 Detailed Discussion of the Simple Model

A.1.1 Payoff for Exponential Model

Recall that $F(v|\theta) = 1 - \exp(-\theta v)$ and $f(v|\theta) = \theta \exp(-\theta v)$ with $\theta \in \mathbb{R}_+$. Then, (1.2) is written as

$$
\Pi(\theta, \rho; N) = N \left\{ \rho \exp(-\theta \rho)(1 - e^{-\theta \rho})^{N-1} + \int_{\rho}^{\infty} y(N-1)e^{-\theta y}(1-e^{-\theta y})^{N-2}\theta e^{-\theta y}dy \right\}
$$

$$
= \rho N \exp(-\theta \rho) \sum_{k=0}^{N-1} \left( \begin{array}{c} N - 1 \\ k \end{array} \right) (-1)^k \exp(-k \theta \rho) 
$$

$$
+ \theta N (N-1) \rho \exp(-2\theta y) \sum_{k=0}^{N-2} \left( \begin{array}{c} N - 2 \\ k \end{array} \right) (-1)^k \exp(-k \theta y)dy 
$$

$$
= \rho N \sum_{k=0}^{N-1} \left( \begin{array}{c} N - 1 \\ k \end{array} \right) (-1)^k \exp(-(k+1) \theta \rho) 
$$

$$
+ \theta N (N-1) \sum_{k=0}^{N-2} \left( \begin{array}{c} N - 2 \\ k \end{array} \right) (-1)^k \int_{\rho}^{\infty} y \exp(-(k+2)\theta y)dy 
$$

Note that the binomial expansion, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$, is used for the second equality. Using $\int y \exp(cy)dy = \exp(cy)(cy - 1)/c^2$,

$$
\int_{\rho}^{\infty} y \exp(-(k+2)\theta y)dy = \left[ \frac{\exp(-(k+2)\theta y)}{(k+2)^2\theta^2} (- (k+2)\theta y - 1) \right]_{\rho}^{\infty}
$$

$$
= \rho \frac{\exp(-(k+2)\theta \rho)}{(k+2)\theta} + \frac{\exp(-(k+2)\theta \rho)}{(k+2)^2\theta^2}
$$
Plugging this back to the payoff function, we have

\[ \Pi(\theta, \rho; N) = \rho N \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \exp(-(k + 1)\theta \rho) + \theta N(N - 1) \sum_{k=0}^{N-2} \binom{N-2}{k} (-1)^k \left[ \frac{\exp(-(k + 2)\theta \rho)}{(k + 2) \theta} + \frac{\exp(-(k + 2)\theta \rho)}{(k + 2)^2 \theta^2} \right] \]

Finally, relabeling the summation indexes, we obtain

\[ \Pi(\theta, \rho; N) = \rho N \sum_{k=1}^{N} \binom{N-1}{k-1} (-1)^{k-1} \exp(-k\theta \rho) + \rho N(N - 1) \sum_{k=1}^{N-1} \binom{N-2}{k-1} (-1)^{k-1} \frac{\exp(-(k + 1)\theta \rho)}{k + 1} + N(N - 1) \sum_{k=1}^{N-1} \binom{N-2}{k-1} (-1)^{k-1} \frac{\exp(-(k + 1)\theta \rho)}{\theta(k + 1)^2} \]

A.1.2 Posterior Analysis

The Gamma prior density is given by

\[ p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 1} \exp(-\beta \theta) \quad (A.1) \]

for some positive \( \alpha \) and \( \beta \). Let \( \bar{v} := \frac{1}{T} \sum_t v_t \). The posterior is also a gamma distribution with updated parameters \( \tilde{\alpha} = \alpha + T \) and \( \tilde{\beta} = \beta + T\bar{v} \):

\[ p(\theta|\bar{v}) = \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \theta^{\tilde{\alpha} - 1} \exp\left(-\tilde{\beta} \theta\right) \quad (A.2) \]

Now, we simplify the posterior payoff as follows.

\[ \int \Pi(\theta, \rho; N)p(\theta|z)d\theta = \rho N \sum_{k=1}^{N} \binom{N-1}{k-1} (-1)^{k-1} \int \exp(-k\theta \rho) p(\theta|z)d\theta \]
\[
+ \rho N(N-1) \sum_{k=1}^{N-1} \binom{N-2}{k-1} \frac{(-1)^{k-1}}{k+1} \int \exp(-(k+1)\theta) p(\theta|z) d\theta \\
+ N(N-1) \sum_{k=1}^{N-1} \binom{N-2}{k-1} \frac{(-1)^{k-1}}{(k+1)^2} \int \theta^{-1} \exp(-(k+1)\theta) p(\theta|z) d\theta
\]

Note that the integrals can be simplified. For example, the first one is

\[
\int \exp(-k\theta) p(\theta|z) d\theta = \int \exp(-k\theta) \frac{\bar{\beta}^\alpha}{\Gamma(\bar{\alpha})} \theta^{\bar{\alpha}-1} \exp(-[\bar{\beta} + \bar{\beta}]\theta) d\theta = \left( \frac{\bar{\beta}}{k\rho + \bar{\beta}} \right)^\alpha
\]

Similarly for the other ones, we obtain the closed form expression

\[
\int \Pi(\theta, \rho; N)p(\theta|z) d\theta = \rho N \sum_{k=1}^{N} \binom{N-1}{k-1} (-1)^{k-1} \left( \frac{\bar{\beta}}{k\rho + \bar{\beta}} \right)^\alpha \\
+ \rho N(N-1) \sum_{k=1}^{N-1} \binom{N-2}{k-1} \frac{(-1)^{k-1}}{k+1} \left( \frac{\bar{\beta}}{(k+1)\rho + \bar{\beta}} \right)^\alpha \\
+ \frac{\bar{\beta} N(N-1)}{\bar{\alpha}-1} \sum_{k=1}^{N-1} \binom{N-2}{k-1} \frac{(-1)^{k-1}}{(k+1)^2} \left( \frac{\bar{\beta}}{(k+1)\rho + \bar{\beta}} \right)^{\alpha-1}
\]

A.2 The Adaptive Metropolis Algorithm with GHK sampler

For a given prior \(p(\theta)\), we can simulate the posterior using a Metropolis-Hastings algorithm. However, since choosing a good proposal density is hard for a high dimensional \(\theta\), we use an adaptive MCMC algorithm. In particular, we employ the Adaptive Metropolis (AM) algorithm of Haario, Saksman, and Tamminen (2001). Suppose \(\theta\) is \(d\) dimensional. Let \(I_d\) be the \(d \times d\) identity matrix and \(s_d := (2.38)^2/d\). At each \(s\)-th iteration, we draw a proposal \(\tilde{\theta}\) from \(N(\theta_{s-1}, \Omega_{s-1})\) where, for a small
number \( \varepsilon > 0 \) and a prespecified initial periods \( s_0 \),

\[
\Omega_{s-1} := \begin{cases} 
\Omega_0 & \text{if } s \leq s_0 \\
 s_d \text{Cov}(\theta_0, \theta_1, \ldots, \theta_{s-1}) + s_d \varepsilon I_d & \text{otherwise}
\end{cases}
\]

with \( \Omega_0 \) and \( \text{Cov} \) denoting some initial covariance and the sample covariance, respectively. Then, \( \theta_s := \tilde{\theta} \cdot 1 \{ u < \alpha \} + \theta_{s-1} \cdot 1 \{ u > \alpha \} \) with \( u \sim \text{Unif}[0,1] \) and

\[
\alpha := \min \left\{ \frac{p(\tilde{\theta}) \hat{L}(|\tilde{\theta}|)}{p(\theta_{s-1}) \hat{L}(|\theta_{s-1}|)}, 1 \right\}
\]

(Haario, Saksman, and Tamminen (2001) show that the AM algorithm converges to the correct posterior for any \( \theta_0 \) with \( p(\theta_0) > 0 \) provided that the posterior is bounded from above and has a bounded support. They also note that we can update \( \Omega_s \) for any increasing subset of \( \{ \theta_s \} \).

For the flexible model (1.6), \( \Theta = \left\{ \theta \in \mathbb{R}_+^k | \sum_{j=1}^k \theta_j = 1 \right\} \). Then, the inequality restrictions on \( \theta \) can be expressed as

\[
B \theta \geq b
\]

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

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

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 \( \theta \) be the first \( k-1 \) elements of the parameter vector, since \( \theta_k = 1 - \sum_{j=1}^{k-1} \theta_j \).

Suppose that the current draw is \( \theta^* \) and the candidate for the next step is
determined by
\[ \tilde{\theta} = \theta^* + u \]

where \( u \sim N(0, \Omega)1(B\tilde{\theta} \geq b) \). Hence, we need to be able to draw \( u \) from the truncated normal distribution.

Following Geweke (1995), we can generate \( \theta \) that satisfies the linear inequality constraints above. Since \( \tilde{\theta} = \theta^* + u \),

\[ Bu \geq b^* := b - B\theta^* \quad (A.5) \]

Then,

\[ Bu \sim N(0, B\Omega B')1(Bu \geq b^*) \quad (A.6) \]

Now, we apply the GHK sampler to (A.6). Let \( C \) be the Cholesky decomposition of \( B\Omega B' \). Then, it is easy to see that

\[ Bu = C\varepsilon \text{ with } \varepsilon \sim N(0, I_{k-1})1(\varepsilon \geq \varepsilon) \quad (A.7) \]

where for each \( j = \{1, 2, \ldots, k-1\} \),

\[ \xi_j = \frac{b_j^* - \sum_{i=1}^{j-1} c_{j,i}\varepsilon_i}{c_{j,j}} \quad (A.8) \]

where \( b_j^* \) is the \( j \)th 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 \( \varepsilon_j \) from the truncated standard normal distribution with lower limit \( \xi_j \) recursively from \( j = 1 \) to \( k - 1 \). Once we draw a random vector \( \varepsilon \), the candidate \( \tilde{\theta} \) is determined by

\[ \tilde{\theta} = \theta^* + B^{-1}C\varepsilon \quad (A.9) \]

\[ ^1\text{For } j = 1, \xi_1 = b_1^*/c_{1,1} \]
Figure A.1: Flexibility of Bernstein Densities

A.3 Bernstein Densities with different $k$

It might be useful to understand how flexible (1.6) is for different $k$’s. For any $j \in \{1, \cdots, k\}$ with a given $k$, the mode of the beta distribution with parameters $j$ and $k - j + 1$ is $\frac{j-1}{k-1}$. Hence, (1.6) is a weighted average of beta densities for which the modes are the equally spaced $k$ grid points $\{0, \frac{1}{k-1}, \cdots, \frac{k-2}{k-1}, 1\}$ with the weights given by $\theta$. It turns out that we can approximate any density function defined on the unit interval using (1.6) for some $k$. We describe the flexibility for different $k$ visually in Figure A.1. The first panels of Figure A.1 illustrate the beta densities used for the Bernstein densities in (1.6) for $k = 5, 15, \text{and} 50$. The second panels are some examples of the Bernstein densities for each $k$.$^2$ Figure A.1 provides a rough idea on the flexibility implied by different $k$’s.

---

$^2$To construct these examples, we draw $\theta$ from the uniform distribution over $\Theta$ for each $k$.  

APPENDIX TO CHAPTER 2

B.1 Simulation Algorithm

B.1.1 Sampling Valuations and Bidding Function Evaluation for IPV

The auction simulation algorithm consists of two steps as follows. $H$ denotes the CDF of $h$ and $x_{(r)}$ is the $r$-th smallest out of $x_1, x_2, \ldots$. Under given $\theta$,

1. Draw a random sample of size $R$, $\tilde{v}_1, \ldots, \tilde{v}_R$.

   Let \{\tilde{u}_r\}_{r=1}^R be uniform draws. Then, \( \tilde{v}_r := F^{-1}(\tilde{u}_r|\theta) = \tilde{F}^{-1}(H^{-1}(\tilde{u}_r|\psi)|\mu). \)

   To approximate $H^{-1}(\tilde{u}_r|\psi)$, we evaluate $H(\cdot|\psi)$ on each point in \{0, 0.01, \ldots, 0.99, 1\} and apply a monotonicity preserving interpolation.\(^1\) Most statistical softwares can evaluate $\tilde{F}^{-1}(\cdot|\mu)$ numerically.

2. Compute the equilibrium bids $\tilde{b}_1, \ldots, \tilde{b}_R$ using (2.2).

   Note that $F(\tilde{v}_r|\theta) = \tilde{u}_r$. Hence, a trapezoidal rule over \( \left((0,0), \left\{\left(\tilde{v}_{(r)}, \tilde{u}_{(r)}^{N-1}\right)\right\}_{r=1}^R\right) \) solves $\int_0^{\tilde{v}_{(r)}} F(\alpha|\theta)^{N-1} d\alpha$. That is,

   \[
   \int_0^{\tilde{v}_{(r)}} F(\alpha|\theta)^{N-1} d\alpha \approx \int_0^{\tilde{v}_{(r-1)}} F(\alpha|\theta)^{N-1} d\alpha + \frac{1}{2} \left(\tilde{u}_{(r-1)}^{N-1} + \tilde{u}_{(r)}^{N-1}\right) \left(\tilde{v}_{(r)} - \tilde{v}_{(r-1)}\right)
   \]

   starting from $\int_0^{\tilde{v}_{(1)}} F(\alpha|\theta)^{N-1} d\alpha \approx \frac{1}{2} \tilde{u}_{(1)}^{N-1} \tilde{v}_{(1)}$.

   This trapezoidal rule is fairly accurate. For the part where $F$ quickly increases, we may have many reference points because they are random sample from $F$. For the part where $F$ is almost flat, we may not have many reference points, because the probability density is small. However, the area to be integrated is almost a rectangle.

---

\(^1\) For example, a piecewise linear function interpolation or the piecewise cubic interpolation method of Fritsch and Carlson (1980).
B.1.2 Sampling for APV: \((\tilde{v}_{1,1}, \tilde{v}_{2,1}), \ldots, (\tilde{v}_{1,R}, \tilde{v}_{2,R}) \sim f(\cdot | \theta)\)

We employ an accept/reject method. In particular, we use a piecewise uniform density mimicking (2.9) for the source function. That is, we use the kernel of (2.9) given by 
\[ k_h(x_1, x_2 | \psi) := \exp \left\{ \sum_{i \in I} \sum_{j \in I} \psi_{i,j} \phi_i(x_1) \phi_j(x_2) \right\}. \]
Then, construct the kernel for the piecewise uniform as follows
\[ k_{pu}(x_1, x_2 | \psi) = k_h \left( \frac{i}{10}, \frac{j}{10} | \psi \right) + k_h \left( \frac{i}{10}, \frac{j+1}{10} | \psi \right) + k_h \left( \frac{i+1}{10}, \frac{j}{10} | \psi \right) + k_h \left( \frac{i+1}{10}, \frac{j+1}{10} | \psi \right) \]
for \((x_1, x_2) \in [\frac{i}{10}, \frac{i+1}{10}] \times [\frac{j}{10}, \frac{j+1}{10}]\) for each \(i, j \in \{0, 1, 2, \ldots, 9\}\). Then, the accept/reject algorithm is as follows. For \(r = 1, \ldots, R\).

1. Draw a proposal \((\tilde{u}^*_1, \tilde{u}^*_2)\) from the density proportional to \(k_{pu}\).
2. Let \((\tilde{u}_{1,r}, \tilde{u}_{2,r}) = (\tilde{u}^*_1, \tilde{u}^*_2)\) with probability \(\frac{k_h(\tilde{u}^*_1, \tilde{u}^*_2 | \psi)}{Q k_{pu}(\tilde{u}_{1,r}, \tilde{u}_{2,r} | \psi)}\) with \(Q \geq \sup_{(u_1, u_2) \in [0,1]^2} \frac{k_h(u_1, u_2 | \psi)}{k_{pu}(u_1, u_2 | \psi)}\).
3. If the proposal is not accepted, go back to step 1.

Once \((\tilde{u}_{1,1}, \tilde{u}_{2,1}), \ldots, (\tilde{u}_{1,R}, \tilde{u}_{2,R})\) are drawn, then \((\tilde{v}_{1,1}, \tilde{v}_{2,1}), \ldots, (\tilde{v}_{1,R}, \tilde{v}_{2,R})\) are obtained by \(\tilde{v}_{i,r} = \tilde{F}^{-1}(\tilde{u}_{i,r} | \mu)\) for each \(i = 1, 2\) and \(r = 1, \ldots, R\).

B.1.3 Evaluation of Equilibrium Bidding Function for APV

We compute the equilibrium bids \(\{(\tilde{b}_{1,r}, \tilde{b}_{2,r})\}_{r=1}^R\) using (2.1). It may be quite time consuming to evaluate (2.1), since it requires us to estimate many triple integrals. However, we find that the following recursive algorithm reduces the computing time, significantly, which has three steps.

1. We compute \(a_1, \ldots, a_{2R}\) with
\[ a_j := \frac{f(\tilde{v}_{(j)}, \tilde{v}_{(j)} | \theta)}{\int_{0}^{t(j)} f(\tilde{v}_{(j)}, t | \theta) dt} \]

\(^2\)The inverse CDF not only runs slowly in a multivariate case but also gives no help to computing (2.1) as in the IPV.
for \( j = 1, \ldots, 2R \) where \( \tilde{v}(j) \) is the \( j \)-th smallest out of the \( R \times 2 \) simulated random values. Note that the integral on the denominator is quickly estimated by the Gaussian quadrature.  

2. Construct

\[
A_{i,j} := \int_{\tilde{v}(i)}^{\tilde{v}(j)} f(u, \theta) \, du \approx A_{i,j-1} + \frac{1}{2} (a_j + a_{j-1}) \left( \tilde{v}(j) - \tilde{v}(j-1) \right)
\]
for \( i = 1, \ldots, 2R \) and \( j = i + 1, \ldots, 2R \) with \( A_{i,i+1} \approx \frac{1}{2} (a_i + a_{i+1}) \left( \tilde{v}(i+1) - \tilde{v}(i) \right) \).

3. \( L(\alpha|v) := \exp \left\{ -\int_0^v f_{\sigma_1|v_1}(u|u) \, du \right\} \) is approximated by

\[
\int_0^{\tilde{v}(j)} L(\alpha|\tilde{v}(j); \theta) \, d\alpha \approx \frac{1}{2} \sum_{i=1}^{j-1} \left( \exp(-A_{i,j}) + \exp(-A_{i+1,j}) \right) (\tilde{v}_{i+1} - \tilde{v}_i)
\]

The last two steps amount to a trapezoidal rule approximation with reference points \( \tilde{v}(1), \ldots, \tilde{v}(2R) \).

B.1.4 Alternative Method to Estimate Likelihoods

This bidding function evaluation is still computationally expensive. We use an alternative to estimate \( \pi_d(\theta) \) on the implied valuation space discretization. First, we approximate \( \beta^{-1}(.|\theta) \) using a monotonicity preserving interpolation: evaluate (2.1) at every percentile of \( \{\tilde{v}(1), \ldots, \tilde{v}(2R)\} \) and find a strictly increasing function connecting \( \{(b_{p\text{-th}} \%, \tilde{v}_{p\text{-th}} \%)\}_{p=1}^{100} \). Second, we construct the valuation space discretization by evaluating this function at each corner point of the bins in the sample space. Then, we estimate \( \pi_d(\theta) \) by counting the simulated valuations in the \( d \)-th bin in the

---

3 The upper limit of the integral can be a very large number and, in that case, the Gaussian quadrature approximation with finite points may be poor. However, the integral can be expressed as \( \int_0^{\tilde{F}(\tilde{v}(j)|\mu)} h(\tilde{F}(\tilde{v}(j)|\mu), s, \psi) \, ds \) for which the upper limit is always less than one. Then, for this integral, we find that the Gaussian quadrature even with three points provides fairly accurate estimates. Moreover, \( \tilde{F} \) should not be evaluated, because we already have \( \tilde{u}_{r,i} = \tilde{F}(\tilde{u}_{r,i}|\mu) \) for \( i = 1, 2, r = 1, \ldots, R \).
valuation space. Since this procedure evaluates (2.1) only 100 times regardless of $R$, we may use a large $R$.

**B.2 Basis Functions and Prior Specification**

**B.2.1 Legendre Polynomials and Smoothness Control**

We employ the Legendre polynomials, i.e., $\phi_j(u) := \sqrt{2j + 1}\tilde{\phi}_j(2u - 1)$ with $\tilde{\phi}_j(x) = \frac{d^j}{dx^j}(x^2 - 1)^j/(2^jj!)$. Note that the order of $\phi_j$ gets higher as $j$ increases. Hence, to control the smoothness, we use the prior such that $\psi_j \sim N(0, \tau/2^j)$ for $\tau > 0$. Then, $\psi_j$ is more condensed around zero for higher $j$ and the posterior picks a smooth density more likely. We use Unif[0,1] for $\tilde{F}$ just for simplicity.\(^4\)

**B.2.2 Normalized B splines**

We construct the basis functions as follows:

\[
\left\{ \phi_i(x) := K\left(\frac{x - i/k}{1/k}\right) \right\}_{i \in I} \tag{B.1}
\]

where $K(x) := \sum_{j=0}^{l} (-1)^j \binom{n}{j(l-j)!} \left(x + \frac{l}{2} - j\right)^{l-1}_+ 1 \{ |x| < l/2 \}$ with an integer $l > 1$ and $x^a_+ = x^a \{ x > 0 \}$. $K$ is a kernel symmetric about zero and $l - 1$ times differentiable with support $[-l/2, l/2]$. In addition, $I$ is the set of all integers in $[-m, k + m]$, with an integer $k > l$ and $m := \lfloor (l/2) - 1 \{ l \text{ is even} \} \rfloor$. (|I| = 2m + k + 1, cardinality of $I$.) Then, $\{ \phi_i \}$ are centered on equidistant grid points $\{-m/k, -m+1/k, \ldots, k+m/k\}$.

Note that (B.1) are $l - 2$ times differentiable and the $l - 1$-th derivative does not exist at the grid point. Some are located outside [0, 1], since every point in [0, 1] must have equal number of nonzero bases for the affiliation condition below.\(^5\)

---

\(^4\) Hence, the properly selected $\tilde{F}$ may result in smaller MISE’s than presented below.

\(^5\) For example, $\phi_{-m}$ is the most left located with a positive tail at zero. That is, $\phi_{-m-1}(0) = 0$, if $-m - 1 \in I$. Similarly, $\phi_{k+m}(1) > 0$, while $\phi_{k+m+1}(1) = 0$. 
B.2.3 Affiliation Restrictions with Normalized B splines

Affiliation is equivalent to supermodularity of the log-density. \( \kappa_{\psi}(x_1, x_2) := \sum_{i \in I} \sum_{j \in I} \psi_{i,j} \phi_i(x_1) \phi_j(x_2) \) for simplicity. Then, (2.8) is log-supermodular if and only if \( \frac{\partial^2}{\partial x_1 \partial x_2} \kappa_{\psi}(x_1, x_2) \geq 0 \) for all \((x_1, x_2) \in [0, 1] \times [0, 1]\), which requires infinitely many restrictions. However, Beresteanu (2007), using (B.1), characterizes supermodularity of the log-density by

\[
\kappa_{\psi} \left( \frac{i}{k}, \frac{j}{k} \right) + \kappa_{\psi} \left( \frac{i+1}{k}, \frac{j+1}{k} \right) \geq \kappa_{\psi} \left( \frac{i+1}{k}, \frac{j}{k} \right) + \kappa_{\psi} \left( \frac{i}{k}, \frac{j+1}{k} \right)
\]

for all \(i, j \in \{0, 1, \ldots, k\}\), which is now \(k^2\) linear inequalities. Furthermore, the exchangeability reduces this to \(\frac{k(k+1)}{2}\), because only the ones associated with grid points below (or above) the 45 degree line are relevant due to \(\kappa_{\psi}(x_1, x_2) = \kappa_{\psi}(x_2, x_1)\). Recall that \(\psi\) denotes the vector of \(\psi_{i,j}\) with \(i \geq j\). Hence, it is \(|I|^2\) dimensional. Therefore, \(A\) in (2.11) is \(\frac{k(k+1)}{2} \times \frac{|I||I|+1}{2}\).

B.2.4 Tail Behavior

The discontinuity of optimal reserve price with respect to the tail behavior leads us to construct a prior to control the tail behavior. First of all, we assume \(p(\mu, \psi) = p(\mu)p(\psi)\) for simplicity. Let \(N(a, b)\) denote a normal density with mean vector \(a\) and covariance \(b\). For the AM algorithm to converge, the posterior must have a bounded support. Hence, we use a prior whose supports are given by \(\mu \in (0, 30]\) and \(\max(|\psi|) < 30\). In addition, the prior should be zero for \(\psi\) violating (2.11). Then, we employ the prior in the form of

\[
p(\mu) \propto N(25, 1) \cdot 1 \{\mu \in (0, 30]\} \\
p(\psi) \propto N(0, \Sigma_{\psi}) \cdot 1 \{A\psi \geq 0 \text{ and } \max(|\psi|) < 30\}
\]

Now, recall that \(\phi_j(\cdot)\) given in (B.1) is centered around \(\frac{j}{k}\). Hence, \(\psi_{i,j}\) is related with the shape of (2.9) around \(\left(\frac{i}{k}, \frac{j}{k}\right)\). Let \(d_{i,j} := 10 \cdot \max \left\{ \left(\frac{i}{k}\right)^2 + \left(\frac{j}{k}\right)^2 \right\}^{\frac{3}{2}}, 0.02 \right\}, \) which is increasing in the distance of \(\left(\frac{i}{k}, \frac{j}{k}\right)\) from the origin. We employ a diagonal
matrix $\Sigma_\psi$ whose element associated with $\psi_{i,j}$ equals $d_{i,j}^{-1}$. Then, this prior puts a smaller variance on $\psi_{i,j}$ located further from the origin so that the posterior selects a valuation density whose tail resembles the exponential $\tilde{f}$.

B.3 The Adaptive Metropolis Algorithm

For a given prior $p(\theta)$, we can simulate the posterior using a Metropolis-Hastings algorithm. However, since choosing a good proposal density is hard for a high dimensional $\theta$, we use an adaptive MCMC algorithm. In particular, we employ the Adaptive Metropolis (AM) algorithm of Haario, Saksman, and Tamminen (2001). Suppose $\theta$ is $d$ dimensional. Let $I_d$ be the $d \times d$ identity matrix and $s_d := (2.38)^2/d$. At each $s$-th iteration, we draw a proposal $\tilde{\theta}$ from $N(\theta_{s-1}, \Omega_{s-1})$ where, for a small number $\varepsilon > 0$ and a prespecified initial period $s_0$,

$$
\Omega_{s-1} := \begin{cases} 
\Omega_0 & \text{if } s \leq s_0 \\
 s_d \widehat{\text{Cov}}(\theta_0, \theta_1, \ldots, \theta_{s-1}) + s_d \varepsilon I_d & \text{otherwise}
\end{cases}
$$

with $\Omega_0$ and $\widehat{\text{Cov}}$ denoting some initial covariance and the sample covariance, respectively. Then, $\theta_s := \tilde{\theta} \cdot 1\{u < \alpha\} + \theta_{s-1} \cdot 1\{u > \alpha\}$ with $u \sim \text{Unif}[0, 1]$ and

$$
\alpha := \min \left\{ \frac{p(\tilde{\theta}) L(\tilde{\theta}|y)}{p(\theta_{s-1}) L(\theta_{s-1}|y)}, 1 \right\} \tag{B.2}
$$

Haario, Saksman, and Tamminen (2001) show that the AM algorithm converges to the correct posterior for any $\theta_0$ with $p(\theta_0) > 0$ provided that the posterior is bounded from above and has a bounded support. They also note that we can update $\Omega_s$ for any increasing subset of $\{\theta_s\}$.

B.3.1 Sampling $\psi$ under $A\psi > 0$

Following Geweke (1995) draws a random vector under a set of linear inequality restrictions using the GHK sampler. As noted earlier, we draw a candidate under the affiliation restrictions (2.11) in order to obtain a proposal function that generates
an acceptable candidate more frequently. For this purpose, we use Geweke (1995) as follows. Let $\bar{A}$ denote a $\frac{n(n+1)}{2}$ dimensional invertable matrix that contains $A$ in (2.11) as its first $\frac{k(k+1)}{2}$ rows. Then, (2.11) can be written as

$$\bar{A}\psi \geq \bar{a}$$

for which first $\frac{k(k+1)}{2}$ elements of $\bar{a}$ are zeros and the others are all $-\infty$. Therefore, the additional restrictions brought by $\bar{A}$ are not relevant.

Now, let $\psi$ be the current parameter of the MCMC sequence. Then, we draw a candidate $\tilde{\psi}$ for the next iteration under the linear inequality constraints:

$$\tilde{\psi} \sim N(\psi, \Omega)1(\bar{A}\psi \geq \bar{a})$$

Let $u := \tilde{\psi} - \psi$. Then,

$$\bar{A}u = \bar{A}\tilde{\psi} - \bar{A}\psi \sim N(0, \bar{A}\Omega\bar{A}^t)1(\bar{A}u \geq \bar{a}^* := \bar{a} - \bar{A}\psi)$$

Let $C$ be the cholesky decomposition of $\bar{A}\Omega\bar{A}^t$. Then, it can be shown that $\bar{A}u = C\varepsilon$ with $\varepsilon \sim N(0, I)1(\varepsilon \geq \varepsilon)$. The lower bounds are given by

$$\varepsilon_j = c_{j,j}^{-1}\left(a_j^* - \sum_{i=1}^{j-1} c_{j,i}\varepsilon_i\right)$$

for $j > 1$ and $\varepsilon_1 = c_{1,1}^{-1}a_1^*$ where $c_{i,j}$ is the $(i, j)$ element of $C$. We draw $\varepsilon$'s recursively from the truncated standard normal distribution. Then, $\tilde{\psi} = \psi + \bar{A}^{-1}C\varepsilon$.

B.4 More Discussion on Monte Carlo Studies

B.4.1 Implementation of BSL

For $\tilde{f}$ we use the uniform $[0,1]$ and employ Legendre polynomials to construct $\{\phi\}$. Especially, we use 20 components. Then, the AM algorithm iterates 500,000 times and we choose every 20 parameters from the last 200,000 iterations to compute the
predictive density $\hat{f}_{BSL}(v|y) := E_\theta[f(v|\theta)|y]$ and choose the reserve price maximizing the future revenue. Figures B.1 to B.4 plots some AM outputs for each Monte Carlos studies. (We record every 20 iterations.) For each iteration, $R = 10,000$ bids are simulated. The bid space is discretized into twenty bins with equal size.

For each data generating process, we use 1,000 Monte Carlo replications. If we run the BSL for each replication separately, it would be very time consuming. In this reason, we implement our method only for the first replication. Then, we estimate the predictive valuation density using an importance sampling method.

Specifically, the predictive valuation density estimate for $j$-th replication is estimated by $\hat{f}_{BSL}(v|y^j) := \frac{\sum_{s=1}^{S} w_j(\theta_s)f(v(\theta_s))}{\sum_{s=1}^{S} w_j(\theta_s)}$ where $S$ is the number of random parame-
ters drawn from the posterior of the first Monte Carlo replication and $w_j(\theta)$ is the ratio of target density (the $j$-th posterior) and the source density (the 1st posterior). That is, $w_j(\theta) = \frac{p(u, \theta|y^j)}{p(u, \theta|y^1)} = \prod_{d=1}^D \left\{ \hat{p}_d(\theta) \right\} y_d^j - y_d^1$ and $y_d^j$ is the histogram implied by the discretization and the $j$-th sample for $j = 2, \ldots, 1,000$. Similarly, we conduct the Bayesian decision method to compute the reserve price for last 999 replications.

B.4.2 Specification for GPV

We employ the estimation procedure that Guerre, Perrigne, and Vuong (2000) use for their Monte Carlo study. They take the triweight kernel $K(u) := \frac{35}{32} (1 - u^2)^3$. 

Figure B.2: The AM outputs for the Exponential DGP
Figure B.3: The AM outputs for the Nonsmooth DGP

Then, the bid density is given by \( \hat{g}(b|h_g) := \frac{1}{h_g TN} K\left(\frac{b-b_{i,t}}{h_g}\right) \) using the bandwidth \( h_g = 1.06 \cdot \text{stdv}(z) \cdot (TN)^{-1/5} \). The pseudo values are computed by \( \hat{v}_{i,t} = \beta^{-1}(b_{i,t}) = b_{i,t} + \frac{\hat{G}(b_{i,t})}{(N-1)g(b_{i,t})} \) only for \( b_{i,t} \in [\min(z) + h_g, \max z - h_g] \). Finally, they estimate the valuation density using \( \hat{f}(v|h_g, h_f) := \frac{1}{h_f |\{\hat{v}\}|} K\left(\frac{v-\hat{v}_{i,t}}{h_f}\right) \) with \( h_f = 1.06 \cdot \text{stdv}(\{\hat{v}\}) \cdot |\{\hat{v}\}|^{-1/5} \).

B.4.3 Specification for Oracle GPV

The bandwidths for Oracle GPV are given by \( (h_g^*, h_f^*) := \arg\min_{(h_g, h_f)} \int_0^1 \left\{ \hat{f}(x|h_g, h_f) - f(x) \right\}^2 dx \). We run a grid search to solve this optimization problem. \( (h_g^*, h_f^*) \) does not imply a monotone inverse bidding function.
Note that the smallest bandwidth for strictly monotone $\hat{\beta}^{-1}$ is typically too large. Therefore, it leads to too flat bid density estimate which does not fit the data well. Instead, Oracle GPV chooses a very small $h_g^*$ and a large $h_f^*$: this $h_g^*$ implies a very noisy $\hat{\beta}^{-1}$ which covers the true inverse bidding function so that the pseudo values from this $\hat{\beta}^{-1}$ look roughly distributed as the true DGP. Then, the large $h_f^*$ provides a smooth density estimate over these pseudo values.

B.4.4 Information Loss of GPV

We summarize the loss of information from GPV in Table B.1. The first column is the fraction of the sample points trimmed out the boundary problem of the kernel.
method. The second column is the fraction of data points whose order is distorted due to the nonincreasing inverse bidding function. The third column is simply sum of the first two columns.

B.5 Appendix to Wildcat Auction Analysis

B.5.1 Specification for $f(\cdot|\theta)$

We employ $(l,k) = (4, 10)$. Then, $\psi$ has 91 components and (2.9) is twice differentiable.\footnote{$|\psi| = \frac{|I|(|I|+1)}{2} = 91$ for which $|I| = 2m + k + 1 = 2(1) + 10 + 1 = 13$. The $l$-th derivative does not exist at each grid point $\frac{i}{k}$ for $i \in I$.} This is differentiable enough for the affiliation condition to be written as (2.10). We judge this specification to be sufficiently flexible and computationally practical. We iterate the AM algorithm 100,000,000 times and throw away first 60,000,000 of them. Then, we take every 4,000 iterations to make the posterior inference. Hence, $S = 10,000$. Figure B.5 plots the AM algorithm for some parameters. (We record every 1,000 iteration.)

B.5.2 Computation

Now, we discuss the computation of (2.13). We consistently estimate (2.14) using

$$\hat{\Pi}(\theta, \rho) := \frac{1}{R} \sum_{r=1}^{R} \{\beta(\hat{v}(2), r|\rho, \theta) \cdot 1(\hat{v}(2), r > \rho)\} \quad (B.4)$$

<table>
<thead>
<tr>
<th>DGP’s</th>
<th>Trimming Rate (boundaries)</th>
<th>Distortion Rate (Monotonicity Violation)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Normal</td>
<td>10.918</td>
<td>13.993</td>
<td>24.911</td>
</tr>
<tr>
<td>Exponential</td>
<td>18.420</td>
<td>15.503</td>
<td>33.923</td>
</tr>
<tr>
<td>Nonsmooth</td>
<td>7.567</td>
<td>14.772</td>
<td>22.339</td>
</tr>
<tr>
<td>Wildcat-like</td>
<td>18.883</td>
<td>14.404</td>
<td>33.288</td>
</tr>
</tbody>
</table>
Figure B.5: The AM outputs for the OCS wildcat auction data analysis

where \(\{(\tilde{u}_{1,r}, \tilde{v}_{2,r})\}_{r=1}^{\tilde{R}} \sim f(\cdot, \cdot | \theta)\). Hence, we approximate the Bayes action (2.13) with

\[
\hat{\rho}_B(y) := \arg \max_{\rho \in A} \frac{1}{S} \sum_{s=1}^{S} \hat{\Pi}(\theta_s, \rho)
\]  

(B.5)

where \(\{\theta_s\}_{s=1}^{S}\) are the MCMC output.

Note that for the AM algorithm we have used new simulation draws at each iteration to obtain the unbiased likelihood estimates. However, we use a fixed uniform draw \(\{(\tilde{u}_{1,r}, \tilde{u}_{2,r})\}_{r=1}^{\tilde{R}}\) to compute (B.4) for each \(\rho\) and \(\theta\), because, otherwise, the sample mean in (B.5) would not be smooth. Now, to evaluate (B.5), we employ...
the algorithm as follows. For a fixed $\theta$

1. Compute $\{\tilde{v}_{1,r}, \tilde{v}_{2,r}\}_{r=1}^{\tilde{R}}$ using an inverse CDF method:
   
   $\tilde{v}_{1,r} = F_1^{-1}(\tilde{u}_{1,r} | \theta)$ and $\tilde{v}_{2,r} = F_2^{-1}(\tilde{u}_{2,r} | \tilde{v}_{1,r}, \theta)$ with the marginal CDF, $F(v_1 | \theta) := \int_0^{v_1} \int_0^\infty f(s,t | \theta) dt ds$, and the conditional CDF, $F(v_2 | v_1, \theta) := \int_0^{v_2} \int_0^{\infty} f(v_1, s | \theta) ds$.

2. Compute $\{\beta(\tilde{v}_{(2),r} | \theta)\}_{r=1}^{\tilde{R}}$ using the recursive method we have employed for the AM algorithm.

   Note that $\tilde{v}_{(2),r} > \tilde{v}_{(1),r}$ for each $r = 1, \ldots, \tilde{R}$ and $\beta(v | \theta) := \beta(v | 0, \theta)$

3. For each $\rho \in \mathcal{A}$, compute $\{\beta(\tilde{v}_{(2),r} | \rho, \theta)\}_{r=1}^{\tilde{R}}$ using $\beta(v | \rho, \theta) = \{\beta(v | \theta) + L(\rho | v, \theta)(\rho - \beta(v | \theta))\} \cdot 1(v > \rho)$ following Li, Perrigne, and Vuong (2003) and construct (B.4).

We do this for each $\theta_s$, $s = 1, \ldots, S$. Then, solving (B.5) is easy.

Only the larger valuations $\{\tilde{v}_{(2),r}\}_{r=1}^{\tilde{R}}$ can be used for step 2 above. However, we compute all the bids $\{(\tilde{b}_{1,r}(\theta), \tilde{b}_{2,r}(\theta))\}_{r=1}^{\tilde{R}}$ with $\tilde{b}_{i,r}(\theta) = \beta(\tilde{v}_{i,r} | \theta)$ for $i = 1, 2$ and $r = 1, \ldots, \tilde{R}$. This is useful to estimate the predictive bid density. That is, we may estimate the marginal of bid density using

$$\tilde{g}(b | f(\cdot | \theta)) := \frac{1}{2\tilde{R}h_m} \sum_{r=1}^{\tilde{R}} \sum_{i=1}^{2} K\left(\frac{b - \tilde{b}_{i,r}(\theta)}{h_m}\right)$$ (B.6)

Then, the marginal predictive bid density is estimated by

$$\tilde{g}(b | y) := \frac{1}{S} \sum_{s=1}^{S} \tilde{g}(b | f(\cdot | \theta_s))$$ (B.7)

In addition, to estimate the joint predictive bid density, we first estimate

$$\tilde{g}(b_1, b_2 | y) := \frac{1}{S} \sum_{s=1}^{S} \left\{ \frac{1}{\tilde{R}h_j} \sum_{r=1}^{\tilde{R}} K\left(\frac{b_{1,s} - \tilde{b}_{1,r}(\theta_s)}{h_j}\right) K\left(\frac{b_{2,s} - \tilde{b}_{2,r}(\theta_s)}{h_j}\right) \right\}$$
with a kernel $K$ and some bandwidths $h_m$ and $h_j$ and symmetrize this using

$$\tilde{g}(b_1, b_2 | y) = \frac{1}{2} \{\tilde{g}(b_1, b_2 | y) + \tilde{g}(b_2, b_1 | y)\} \quad (B.8)$$

We employ $\tilde{R} = 1,500$. 
REFERENCES


