Quantitative Economics 6 (2015), 429–461 1759-7331/20150429

Flexible Bayesian analysis of first price auctions using a simulated likelihood

Dong-Hyuk Kim
Department of Economics, Vanderbilt University

I propose a Bayesian method to analyze bid data from first-price auctions under private value paradigms. I use a series representation to specify the valuation density so that bidding monotonicity is always satisfied, and I impose density affiliation by the nonparametric technique of Beresteanu (2007). This flexible method is, therefore, fully compatible with the underlying economic theory. To handle such a rich specification, I use a simulated likelihood, yet obtain a correct posterior by regarding the draws used for simulation as a latent variable to be augmented in the Bayesian framework; see Flury and Shephard (2011). I provide a step-by-step guide of the method, report its performance from various perspectives, and compare the method with the existing one for a range of data generating processes and sample sizes. Finally, I analyze a bid sample for drilling rights in the outer continental shelf that has been widely studied and propose a reserve price that is decision theoretically optimal under parameter uncertainty.

Keywords. First price sealed bid auctions, affiliated private values, revenue maximizing reserve price, Bayesian analysis, method of series, simulated likelihood, shape restriction.

JEL classification. C11, C13, C15, C44, D44, L38.

1. Introduction

Structural data analysis of first-price auctions has, in its early stage, specified the density of bidders' latent values (willingness to pay) and constructed the likelihood using the density of observed bids, which is linked to the valuation density via an equilibrium. When a complex statistical model is employed, the likelihood evaluation was computationally impractical and, therefore, the early literature has employed strong parametric assumptions mostly within the independent private value paradigm (IPVP). See, for example, Donald and Paarsch (1993), Laffont, Ossard, and Vuong (1995).

Dong-Hyuk Kim: dong-hyuk.kim@vanderbilt.edu

This paper is based on Chapter 2 of my PhD dissertation at University of Arizona in 2010. Many thanks to my PhD supervisor, Keisuke Hirano, and committee members Gautam Gowrisankaran and Mo Xiao. I appreciate comments from seminar participants at University of Arizona, University of Sydney, University of Queensland, University of New South Wales, University of Auckland, University of Illinois, Urbana–Champaign, Australian National University, the 7th International Symposium of Econometric Theory and Applications, the Econometric Society Australasian Meeting in 2011, and the Asian Meeting of Econometric Society in 2011. I also thank the handling coeditor and three anonymous referees for useful and insightful suggestions.

DOI: 10.3982/QE257

The seminal article by Guerre, Perrigne, and Vuong (2000) has greatly widened the scope of this line of research. By showing that the inverse bidding function is a functional of the density of the optimal bid, the article established the nonparametric identification of the valuation density and proposed a two step estimation method. This method first estimates the bid density functions to form the inverse bidding function and then uncovers the distribution of values by evaluating the inverse bidding function at the bid data. Since the method starts from the bid density, which is not the model primitive, the authors named it the *indirect* approach. This approach enables nonparametric analysis as well as extensions to more general paradigms such as the conditional independent private value paradigm (CIPVP) and the affiliated private value paradigm (APVP); see Li, Perrigne, and Vuong (2000, 2002).

The indirect approach may, however, be incompatible with the paradigm on which the empirical method is based. In particular, the estimated inverse bidding function may not be increasing and/or the estimated densities of bids and values may not be affiliated. Even if the estimates are consistent, such problems could in practice lower efficiency and invalidate policy recommendations. This is important to researchers and policy makers because policy prescriptions may have to be drawn even when the sample is small. For example, Li, Perrigne, and Vuong (2003) estimated the revenue maximizing reserve price (RMRP) for the outer continental shelf (OCS) wildcat auctions under the APVP using a bid sample of 217 auctions, each with two bidders. Although the RMRP in this article is valid only when the bid density is affiliated and the inverse bidding strategy is strictly increasing, the article estimated the bid density without imposing the shape restrictions.

Motivated by this, I propose an empirical framework that is fully consistent with the underlying paradigm.[1] Following the earlier literature, I directly specify the valuation density so that the bidding monotonicity is automatically satisfied.[2] I use a flexible series representation on which the density affiliation can be imposed via the nonparametric shape restriction method of Beresteanu (2007): interdependence among valuations is allowed to be flexible. To handle such a rich specification, the method employs a simulated likelihood, yet obtains the exact posterior by regarding the draws used for simulation as a latent variable to be augmented in the Bayesian framework; see Andrieu, Doucet, and Holenstein (2010), Flury and Shephard (2011). To make use of this, I finely discretize the sample space to construct a multinomial likelihood, which is then unbiasedly estimated by simulated bid data: the Bayesian method with a simulated likelihood (BSL), therefore, does not suffer simulation error.

An important objective of this paper is also to provide a step-by-step guide of the BSL because this paper introduces these recent techniques to the auction literature. I discuss

[1] Similarly motivated, two recent articles developed the idea to impose shape restrictions independently and simultaneously with the present paper, but adopting a different approach. Henderson, List, Millimet, Parmeter, and Price (2012) imposed bidding monotonicity in the IPVP and Hubbard, Li, and Paarsch (2012) restricted density affiliation using a parametric copula in the APVP. These articles took the indirect approach.

[2] The failure of bidding monotonicity arises only in the indirect approach. Bierens and Song (2012) also proposed a method of simulated moments using a sieve representation of the valuation density for the IPVP.

in detail the theoretical justification and implementation of the method, including the conditions under which the posterior inference is exact and how to carry out the BSL while imposing density affiliation. I also explain the Bayesian model selection to choose the statistical model that fits the data the best, and document the performance from various perspectives such as precision and accuracy of inference, robustness against the prior and discretization, and computing time. In particular, the BSL is more accurate than the previous indirect method for a wide range of data generating processes (DGPs) and sample sizes. The posterior analysis is robust against the prior and discretization for a number of well behaved distributions.

Finally, the Bayesian method naturally provides a decision theoretic framework, which is shown to be useful for auction design problems in Aryal and Kim (2013), Kim (2013). I revisit the bid sample from OCS wildcat sales that was investigated by Li, Perrigne, and Vuong (2000, 2003) under the CIPVP and the APVP, respectively. The distribution of the bid data is highly skewed to the left, having a long tail to the right with outliers. These outliers imply a positive probability of extremely large values, suggesting an unreasonably large RMRP. First, I use a strong prior to control the tail behavior and implement the BSL to obtain the posterior of the valuation density. I then choose a reserve price of \$163 that maximizes the posterior predictive revenue. This proposal is optimal under the subjective expected utility principle of Savage (1954), Anscombe and Aumann (1963) as well as under the average *frequentist* risk principle. I find that the posterior revenue distribution at the reserve price of \$163 first-order stochastically dominates the one at the actual reserve price of \$15. The revenue gain is *economically* significant.

In the paper, I assume that the policy maker commits not to sell the auctioned item when no bid exceeds the reserve price following the convention of the empirical auction literature.[3] However, if the policy maker may sell the unsold item in a future auction and bidders expect this, then the auction today competes with the auction tomorrow. Therefore, the RMRP that accounts for such dynamic effects should be lower, and it depends on many factors, including bidders' discount factor and the number of bidders; e.g., for a fixed number of bidders, the more patient are the bidders, the lower is the RMRP (see McAfee and Vincent (1997)).

In the next section, I provide an overview of the empirical environment and the BSL. In Section 3, I illustrate the BSL using the simplest paradigm, the IPVP, and extend, in Section 4, the BSL to the APVP and analyze the bid sample from the OCS wildcat sales. I then conclude the paper by discussing auctions with many bidders. The Appendix collects computational details.

2. Methodology overview

The goal of bid data analysis in this paper is to make inference on the valuation distribution and propose policy recommendations for future auctions. Having this in mind, consider an environment with a bid sample $\mathbf{z}_T := \{(b_{1,t}, \ldots, b_{n,t})\}_{t=1}^{T}$, where $n \geq 2$ is the

[3]See, for example, Paarsch (1997), Li, Perrigne, and Vuong (2003), Krasnokutskaya (2011), Aryal and Kim (2013), Kim (2013) among many.

bid profile at auction $t \in \{1, \ldots, T\}$. All the T sales are in the format of the first-price sealed-bid auction: each bidder i at auction t bids $b_{i,t}$ after observing her own value $v_{i,t}$, and obtains the auctioned item by paying $b_{i,t}$ if and only if $b_{i,t} = \max(b_{1,t}, \ldots, b_{n,t})$.[4] It is important to note that without any a priori knowledge, the bid sample $\mathbf{z}_T$ is useless for achieving the goal of data analysis: one can learn the valuation density only when he knows how the observed bids are related to the unobserved values, and one can predict the bidding behavior under different auction rules only under some behavioral assumptions. Thus, a theoretical paradigm is assumed to be known.

Let $\mathcal{P}$ be the set of all absolutely continuous distributions on $\mathbb{R}^n_+$ with a differentiable density. Let $\mathcal{F}_{\text{APV}} \subset \mathcal{P}$ collect all exchangeable and affiliated distributions; a distribution $F \in \mathcal{P}$ is said to be affiliated if and only if its density f satisfies

$$\frac{\partial^n}{\partial v_1 \cdots \partial v_n} \log f(v_1, \ldots, v_n) \geq 0, \tag{1}$$

and f is said to be exchangeable if and only if

$$f(v_1, \ldots, v_n) = f(v_{i_1}, \ldots, v_{i_n}) \tag{2}$$

for every permutation $(i_1, \ldots, i_n)$, e.g., $f(v_1, v_2) = f(v_2, v_1)$ for $n = 2$. If $\mathbf{v} := (v_1, \ldots, v_n) \sim F \in \mathcal{F}_{\text{APV}}$, the first-price sealed-bid auction induces a game with incomplete information among n bidders for which Milgrom and Weber (1982) derive a symmetric Bayesian Nash equilibrium with a strictly increasing bidding function. Let $f_{y_1|v_1}(\cdot|\cdot)$ be the conditional density of $y_1 := \max\{v_2, \ldots, v_n\}$ given v_1. The equilibrium bidding function is then given by

$$\beta(v; F, \rho) := v - \int_\rho^v \exp\left[- \int_\alpha^v \frac{f_{y_1|v_1}(u|u)}{\int_0^u f_{y_1|v_1}(t|u)\, dt} du \right] d\alpha \tag{3}$$

for $v \geq \rho$, the reserve price; otherwise, bidding any $b < \rho$ is optimal. Let $\beta(v; F) := \beta(v; F, \rho = 0)$. The APVP includes the CIPVP and the IPVP as special cases. In particular, let $\mathcal{F}_{\text{IPV}} \subset \mathcal{F}_{\text{APV}}$ collect all distributions such that $F(v_1, \ldots, v_n) = \prod_{i=1}^n F_1(v_i)$, where F_1 is the marginal distribution of v_1. If $F \in \mathcal{F}_{\text{IPV}}$, the bidding strategy (3) simplifies to

$$\beta(v; F, \rho) := v - \int_\rho^v \left\{ \frac{F_1(\alpha)}{F_1(v)} \right\}^{n-1} d\alpha. \tag{4}$$

I consider the empirical environment where the correct paradigm is known and $b_{i,t} = \beta(v_{i,t}|F)$ for all (i, t), where $\mathbf{v}_1, \ldots, \mathbf{v}_T \overset{\text{i.i.d.}}{\sim} F$ and the valuation distribution F satisfies all the theoretical shape restrictions under the given paradigm. Such prior knowledge is critical not only because it links bids to values, but also because it allows a prediction on bidding behavior and seller's revenue under a different ρ.

[4] A nonbinding reserve price is assumed. This assumption is consistent with many empirical situations including the bid sample analyzed in Section 4 of this paper.

To analyze the bid data, I use a flexible statistical model $\{F(\cdot|\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ to approximate F, where $\boldsymbol{\Theta}$ is the parameter space. Throughout the paper, the theoretical auction model is referred to as the *paradigm*, but the flexible density family that is used to approximate the underlying valuation density is referred to as the *statistical model* or, simply, the model. Let $\beta(\cdot|\boldsymbol{\theta}) := \beta[\cdot|F(\cdot|\boldsymbol{\theta})]$, let $\bar{b}(\boldsymbol{\theta}) := \lim_{v\to\infty} \beta(v|\boldsymbol{\theta})$, and let $\mathbb{1}(A)$ be the indicator for event A. The likelihood is proportional to

$$\prod_{t=1}^{T} g(b_{1,t}, \ldots, b_{n,t}|\boldsymbol{\theta}), \tag{5}$$

where the joint density of the equilibrium bids is given as

$$g(b_1, \ldots, b_n|\boldsymbol{\theta}) = f\big[\beta^{-1}\big(b_1\big|\boldsymbol{\theta}\big), \ldots, \beta^{-1}\big(b_n\big|\boldsymbol{\theta}\big)|\boldsymbol{\theta}\big] \times \prod_{i=1}^{n} \frac{\mathbb{1}[b_i \le \bar{b}(\boldsymbol{\theta})]}{\beta'[\beta^{-1}(b_i|\boldsymbol{\theta})|\boldsymbol{\theta}]}.$$

Since there is no closed form expression for the likelihood (5) for a richly parametrized model, the inverse bidding function $\beta^{-1}(b_{i,t}|\boldsymbol{\theta})$ must be numerically approximated for every data point $b_{i,t}$ in the sample $\mathbf{z}_T$ for each evaluation of the likelihood; an optimization routine needs to run $n \times T$ times. Notice that for one evaluation of the inverse bidding function, the routine computes (just) the bidding function $\beta(\cdot|\boldsymbol{\theta})$ repeatedly many times until it finds a point $x \in \mathbb{R}_+$ to solve $b_{i,t} = \beta(x|\boldsymbol{\theta})$; moreover, one evaluation of the bidding function $\beta(\cdot|\boldsymbol{\theta})$ can already be time consuming, as it involves many integrals in the bidding strategy (3) under the APVP. For this reason, a direct evaluation of the original likelihood (5) has been regarded as impractical, especially for large T. Therefore, the literature has employed either a tightly parametrized statistical model (see Donald and Paarsch (1993), Laffont, Ossard, and Vuong (1995)) or the indirect approach (see Guerre, Perrigne, and Vuong (2000), Li, Perrigne, and Vuong (2002, 2003), Campo, Perrigne, and Vuong (2003), Krasnokutskaya (2011)). Typically, the latter accommodate a nonparametric specification of the density functions.

In this paper, I propose a computationally feasible approach that directly specifies the valuation density using a flexible statistical model. Instead of evaluating the true likelihood (5), however, I employ a method of simulated likelihood that evaluates $\beta^{-1}(\cdot|\boldsymbol{\theta})$ only a small number of times.[5] I shall discuss in the subsequent sections how to do this.

Once the likelihood is approximated by simulation, one could use the maximum likelihood estimator (MLE), taking the classical framework. But I adopt the Bayesian framework for the following reasons. First, the upper bound $\bar{b}(\boldsymbol{\theta})$ depends on the parameter $\boldsymbol{\theta}$, in which case the statistical model is irregular and the MLE fails to be efficient,

[5]For all the exercises in this paper, I evaluate $\beta(\cdot|\boldsymbol{\theta})$ at equidistant grid points $(x_0, x_1, \ldots, x_{100})$ on the support of the value and evaluate $\beta^{-1}(\cdot|\boldsymbol{\theta})$ by interpolation. I consider the number of grid points as sufficiently large because when I increase the number of grid points to 1001, the approximation quality does not change.

while the Bayes estimator continues to be efficient (see Hirano and Porter (2003)).[6] Second, the Bayesian framework naturally provides a decision theoretic framework that is shown to be useful for auction design under parameter uncertainty (see Aryal and Kim (2013), Kim (2013)). Third, if the simulated likelihood is unbiased for the original likelihood, a Bayesian inference can be exact (see Andrieu, Doucet, and Holenstein (2010), Flury and Shephard (2011)).

To elaborate the last point, let $\mathbf{y}$ be the sample, let $p(\boldsymbol{\theta})$ be the prior, let $p(\mathbf{y}|\boldsymbol{\theta})$ be the likelihood, and let $p(\boldsymbol{\theta}|y)$ be the posterior. The objective is to draw an ergodic sample of random parameters $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^S \sim p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$. Let $\hat{p}_u(\mathbf{y}|\boldsymbol{\theta})$ be the simulated likelihood such that

$$p(\mathbf{y}|\boldsymbol{\theta}) = E_u\big[\hat{p}_u(\mathbf{y}|\boldsymbol{\theta})\big], \tag{6}$$

where the expectation is over the simulation draws, u, which are independent and identically distributed (i.i.d.) uniforms. Since $\hat{p}_u(\mathbf{y}|\boldsymbol{\theta})$ is a joint density of $\mathbf{y}$ based on the uniform u, equality (6) implies that $\hat{p}_u(\mathbf{y}|\boldsymbol{\theta})$ gives $p(\mathbf{y}|\boldsymbol{\theta})$ after being integrated over u. Thus, $\hat{p}_u(\mathbf{y}|\boldsymbol{\theta})$ can be interpreted as a joint density of $(u, \mathbf{y})$, i.e., $\hat{p}_u(\mathbf{y}|\boldsymbol{\theta}) = p(u, \mathbf{y}|\boldsymbol{\theta})$. Thus, if one draws

$$\big(u^1, \boldsymbol{\theta}^1\big), \ldots, \big(u^S, \boldsymbol{\theta}^S\big) \sim p(u, \boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(u, \mathbf{y}|\boldsymbol{\theta}),$$

then $\{\boldsymbol{\theta}^s\}_{s=1}^S$ are draws from the correct posterior. The only requirement here is the unbiasedness (6), for which the number of simulation draws can be moderate and do not need to grow rapidly as T increases. Flury and Shephard (2011) discussed this property in detail and argued that it is a great advantage over the simulated MLE, which requires the simulation size to grow at a rate faster than T^2.

3. Independent private value paradigm

In this section, I propose the BSL for the IPVP; I shall extend it to the APVP in Section 4. Since $F \in \mathcal{F}_{\text{IPV}}$, the marginal density f_1 is the only model primitive of interest. I first demonstrate how to construct the likelihood that is to be unbiasedly estimated by simulated data and provide a detailed implementation guide, documenting various aspects of the method. I then run a series of Monte Carlo experiments to compare the BSL with the nonparametric method of Guerre, Perrigne, and Vuong (2000).

3.1 *Valuation density and simulated likelihood*

Consider $v_1 \in [0, 1]$. Let $\mathcal{H}$ be the Hilbert space of functions from $[0, 1]$ to $\mathbb{R}$ and let $\{\phi_j\} \subset \mathcal{H}$ be a sequence of linearly independent functions whose linear span is dense in $\mathcal{H}$, i.e., for any $h \in \mathcal{H}$ and for any $\varepsilon > 0$, there is an index set I and a sequence of real

[6] For example, consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0, \delta_0]$, where δ_0 is to be estimated. Then the MLE is $\delta_{\text{MLE}} = \max\{X_1, \ldots, X_n\}$, and its asymptotic distribution is a shifted exponential distribution, where δ_0 does not belong to the interior of the support of the sampling distribution of the MLE. Specifically, $n(\delta_{\text{MLE}} - \delta_0) \xrightarrow{d} -Y$, where Y is exponentially distributed with density $p_Y(y) = \delta_0 \exp(-\delta_0 y) \cdot \mathbb{1}(y > 0)$.

numbers $\{\theta_j\}_{j\in I}$ such that $\|h - \sum_{j\in I}\theta_j\phi_j\| < \varepsilon$.[7] Polynomials, splines, or Fourier functions can be used to construct such $\{\phi_j\}$. The log density of v_1 can be approximated by

$$\log f_1(v|\boldsymbol{\theta}) = \sum_{j\in I}\theta_j\phi_j(v) + c(\boldsymbol{\theta}), \tag{7}$$

with $\boldsymbol{\theta} := \{\theta_j\}_{j\in I}$ and the normalizing constant $c(\boldsymbol{\theta})$ for any given accuracy; therefore, the statistical model (7) is said to be *flexible*. The statistical literature develops nonparametric methods using this property. The classical statistics theoretically establishes the asymptotic properties of the estimate and derives the rate at which the number of components, $|I|$, increases as the sample size grows to obtain the desirable asymptotic properties. In practice, the number of components is always chosen informally (as Wasserman (2006) pointed out) by the Bayesian information criterion (BIC), the Akaike information criterion (AIC), or some other data driven method such as cross-validation.

The Bayesian statistics, on the other hand, regards the number of components as one of the parameters and updates its prior over the set of *all* positive integers via Bayes rule; see Ferguson (1973), Escobar and West (1995), Petrone (1999). The Bayesian nonparametric methods are technically complicated, and I do not attempt to develop a fully nonparametric method that adopts a simulated likelihood. In Section 3.3, instead, I illustrate how to choose the number of components using the Bayesian model selection. Note that the BIC and the AIC are rough approximations of this formal model selection, each assuming a different prior.

Since $F \in \mathcal{F}_{\text{IPV}}$ and $(b_1,\ldots,b_n)$ are all equilibrium bids, $g(b_1,\ldots,b_n|F) = \prod_{i=1}^n g_1(b_i|F)$, where $g_1(\cdot|F)$ denotes the marginal bid density of b_1; see Guerre, Perrigne, and Vuong (2000). Therefore, $\mathbf{z}_T$ is considered as a random sample of size $n\times T$ from $g_1(\cdot|F)$ with a one dimensional sample space. Let $B\subset[0,1]$ include all bids in $\mathbf{z}_T$ and let $\{[b^*_{d-1},b^*_d]\}_{d=1}^D$ denote the sequence of bins with $B = \bigcup_{d=1}^D[b^*_{d-1},b^*_d]$. The bin probability under $\boldsymbol{\theta}$ is given by

$$\pi_d(\boldsymbol{\theta}) := \Pr\bigl(b\in\bigl[b^*_{d-1},b^*_d\bigr]|\boldsymbol{\theta}\bigr) = \int_{b^*_{d-1}}^{b^*_d} g_1(b|\boldsymbol{\theta})\,db. \tag{8}$$

To define the likelihood, let $y_d := \sum_{t=1}^T\sum_{i=1}^n \mathbb{1}(b_{i,t}\in[b^*_{d-1},b^*_d])$, the number of bids in $[b^*_{d-1},b^*_d]$ for $d=1,\ldots,D$. The associated sample histogram is then $\mathbf{y} := (y_1,\ldots,y_D)$, which can be viewed as a nonparametric estimate of the bid density up to normalization.[8] Let $\bar{y} := \max\mathbf{y}$. Then the probability mass of $\mathbf{y}$ under $\boldsymbol{\theta}$ (the likelihood of $\boldsymbol{\theta}$ for given $\mathbf{y}$) is

$$p(\mathbf{y}|\boldsymbol{\theta}) := \frac{(n\cdot T)!}{y_1!\times\cdots\times y_D!}\prod_{d=1}^D\bigl[\pi_d(\boldsymbol{\theta})\bigr]^{y_d} \propto \prod_{j=1}^{\bar{y}}\left\{\prod_{d=1}^D\bigl[\pi_d(\boldsymbol{\theta})\bigr]^{\mathbb{1}(y_d>j-1)}\right\}. \tag{9}$$

[7] For example, the index set I can be the set of all nonnegative integers less than some prespecified number. But I do not specify I for the time being because the elements of I depend on the choice of ϕ. The norm is here defined by the usual inner product, i.e., $\|h\| := (\int h^2\,d\nu)^{1/2}$, where ν denotes the Lebesgue measure.

[8] Chamberlain (1987) obtains the asymptotic efficiency bound of the generalized method of moments (GMM) estimator using the fact that a multinomial distribution can be arbitrarily close to the true distribution; see Lemma 3 in the article.

Suppose now that the statistical model $f_1(\cdot|\boldsymbol{\theta})$ is given. One can then draw

$$\tilde{v}_r^j, \ldots, \tilde{v}_R^j | \boldsymbol{\theta} \overset{\text{i.i.d.}}{\sim} f_1(\cdot|\boldsymbol{\theta}) \tag{10}$$

independently also across $j = 1, \ldots, \bar{y}$ by a standard scheme such as the inverse cumulative distribution function (CDF), which first draws $u_1^j, \ldots, u_R^j \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0,1]$ and defines $\tilde{v}_r^j := F_1^{-1}(u_r^j|\boldsymbol{\theta})$ with $F_1(v|\boldsymbol{\theta}) := \int_0^v f_1(\alpha|\boldsymbol{\theta})\, d\alpha$ for all j and r.[9] Then the bin probability (8) is unbiasedly estimated by

$$\hat{\pi}_d^j(\boldsymbol{\theta}) := \frac{1}{R}\sum_{r=1}^{R} \mathbb{1}\big(\tilde{v}_r^j \in \big[v_{d-1}^*, v_d^*\big]\big) \quad \text{for any } j = 1, \ldots, \bar{y},$$

with $v_d^* := \beta^{-1}(b_d^*|\boldsymbol{\theta})$ for $d = 1, \ldots, D$. Notice that $\beta^{-1}(\cdot|\cdot)$ needs to be evaluated only a small number of times, D. Observe also that $\prod_{d=1}^D [\pi_d(\boldsymbol{\theta})]^{\mathbb{1}(y_d > j-1)}$ in (9) can be unbiasedly estimated by $\prod_{d=1}^D [\hat{\pi}_d^j(\boldsymbol{\theta})]^{\mathbb{1}(y_d > j-1)}$ for each $j \in \{1, \ldots, \bar{y}\}$ up to a multiplicative constant. To see this, suppose that the indicators are all 1. Then

$$\prod_{d=1}^{D} \hat{\pi}_d^j(\boldsymbol{\theta}) \propto \sum_{r_1=1}^{R} \cdots \sum_{r_D=1}^{R} \prod_{d=1}^{D} \mathbb{1}\big(\tilde{v}_{r_d}^j \in \big[v_{d-1}^*, v_d^*\big]\big) = \sum_{r_1 \neq \cdots \neq r_D} \prod_{d=1}^{D} \mathbb{1}\big(\tilde{v}_{r_d}^j \in \big[v_{d-1}^*, v_d^*\big]\big),$$

where the equality holds because the event is negligible that a simulated value belongs to more than one bin. Moreover, since the simulated values are all independent, the expectation of (9) with all the indicators being 1 is written as

$$E\left[\prod_{d=1}^{D} \hat{\pi}_d^j(\boldsymbol{\theta}) \middle| \boldsymbol{\theta}\right] \propto \sum_{r_1 \neq \cdots \neq r_D} \prod_{d=1}^{D} E\big[\mathbb{1}\big(\tilde{v}_{r_d}^j \in \big[v_{d-1}^*, v_d^*\big]\big)|\boldsymbol{\theta}\big] \propto \prod_{d=1}^{D} \pi_d(\boldsymbol{\theta}).$$

When some indicators are 0, the same argument holds as the bin probability with $\mathbb{1}(y_d > j-1) = 0$ becomes 1. Notice that the total number of simulation draws is $\bar{y} \times R$ with $\bar{y}$ much smaller than T. Moreover, $\beta^{-1}(\cdot|\cdot)$ needs to be evaluated only at the grid points: $(b_0^*, b_1^*, \ldots, b_D^*)$. For the rest of the paper, let

$$\hat{p}_u(\mathbf{y}|\boldsymbol{\theta}) := \frac{(n \cdot T)!}{y_1! \times \cdots \times y_D!} \times \prod_{j=1}^{\bar{y}} \left\{ \prod_{d=1}^{D} \big[\hat{\pi}_d^j(\boldsymbol{\theta})\big]^{\mathbb{1}(y_d > j-1)} \right\},$$

which is the simulated likelihood, where u denotes the auxiliary uniform variable used for simulation.

3.2 *Illustration*

I explain the implementation of the BSL in the IPVP using an artificial bid sample for which the valuation distribution is characterized by $v = \frac{\tilde{v}-0.055}{2.5-0.055}$, where $\tilde{v} \sim$

[9] One could alternatively use the accept/reject sampler. For one dimensional problem, however, the inverse CDF is simpler and quicker. I use the accept/reject sampler for the multivariate problem in the next section. See the Appendix for more details.

$\log \mathrm{N}(0,1) \cdot \mathbb{1}(\tilde{v} \in [0.055, 2.5])$. Guerre, Perrigne, and Vuong (2000) used $\log \mathrm{N}(0,1) \cdot \mathbb{1}(\tilde{v} \in [0.055, 2.5])$ for the Monte Carlo study in their article. I rescale this distribution so that the support is the unit interval. I consider $(n, T) = (2, 200)$. I generate a sample of $T = 200$ pairs of bids, $\mathbf{z}_T$, for which I find (average, standard deviation, skewness) $= (0.158, 0.095, 0.316)$. In this subsection, I illustrate the BSL only using this artificial bid sample, but the exercise here is only one of many that are soon discussed. For example, for this given data, I run the BSL with 10 different statistical models in Section 3.3. Moreover, I implement the method for a number of different pairs of valuation distributions and sample sizes in Section 3.4.

3.2.1 *Specification of $f_1(\cdot|\boldsymbol{\theta})$ and the prior* To construct the basis functions, I use the Legendre polynomials $\phi_j(v) := \sqrt{2j+1} \cdot \tilde{\phi}_j(2u-1)$, where $\tilde{\phi}_j(x) = \frac{d^j}{dx^j}(x^2-1)^j/(2^j j!)$. I use the prior in the form of $p(\boldsymbol{\theta}) = \prod_{j \in I} p(\theta_j)$ with $\theta_j \sim N(0, (10 \cdot 2^j)^{-1})$ for all $j \in I := \{1, \ldots, k\}$. Since the prior mean of θ_j is zero for all j, the prior predicts the uniform density on $[0,1]$, i.e., the density (7) is a constant function if $\theta_j = 0$ for all j. As shown in Figure 1, ϕ_j has $j-1$ extrema. Since the prior variance of θ_j decreases in j, θ_j would get probabilistically close to 0 as j increases. This suggests that a noisy density is unlikely under the prior. I use $k = 7$. As noted above, I consider alternative models in Section 3.3, and the statistical model with $k = 7$ can be chosen by the formal model selection discussed there.

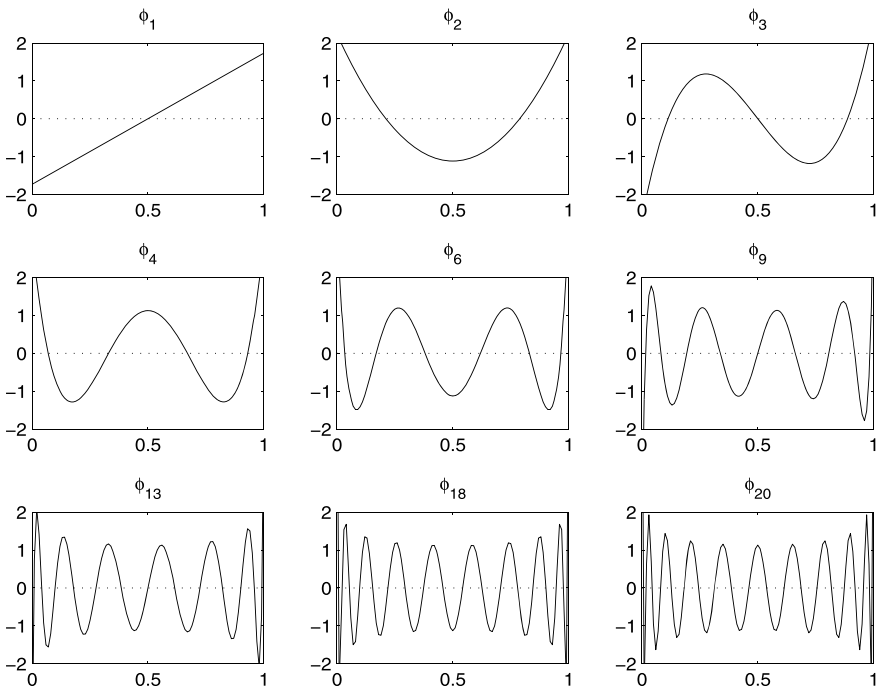

FIGURE 1. Legendre polynomials (recentered at 0.5 with support $[0,1]$). Each panel shows a basis function of the Legendre polynomials ϕ_j for some $j \in \{1, 2, \ldots, 20\}$. The function ϕ_j has $j-1$ extrema with support $[0,1]$, i.e., as j increases, ϕ_j gets noisy.

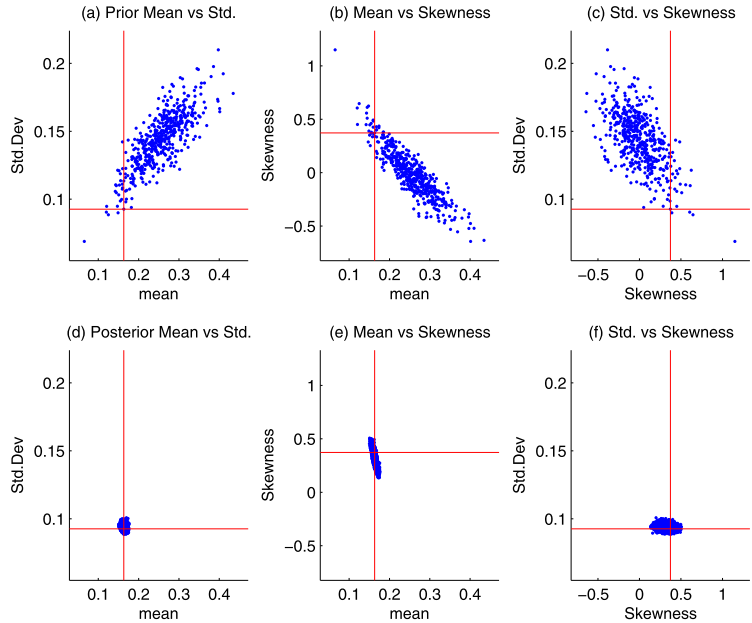

FIGURE 2. Log normal with $(n, T) = (2, 200)$. Panels (a)–(c) [(d)–(f)] demonstrate the distribution of the summary statistics by dots of the bid data under the prior [posterior] along with the summary statistics of the original data in solid lines.

Before computing the posterior, it would be useful to check what the prior and the model say about the data. To do so, I draw $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ and generate a bid sample under $\boldsymbol{\theta}$ of the same size as $\mathbf{z}_T$, and obtain its sample mean, standard deviation, and skewness. After repeating this many times, Figure 2(a)–(c) scatters the distributions of the summary statistics under the prior, marking the summary statistics of the original data $\mathbf{z}_T$ by the plain lines. The data $\mathbf{z}_T$ can be considered as a realization under the prior. This exercise is called a prior predictive analysis; see Geweke (2005).

I discretize the sample space by equidistant grid points $(b_0^*, b_1^*, \ldots, b_D^*)$ with $(b_0^*, b_D^*) = (0, b_{\max})$ with $D = 20$; I find $\bar{y} = 34$. So β^{-1} only needs to be evaluated $D = 20$ times. As will be shown in Section 3.2.4, the posterior inference is almost identical under much more coarse discretizations and, therefore, the loss of information due to the discretization with $D = 20$ seems to be small.

3.2.2 *Posterior computation* To explore the posterior distribution, I employ a Markov chain Monte Carlo (MCMC) algorithm. In particular, I consider the Metropolis–Hastings (MH) algorithm where each component in $\boldsymbol{\theta}$ is updated one-by-one in a prespecified order. To simplify notation, let $\boldsymbol{\theta}_{-j}$ be the (sub)vector of all θ_a with $a < j$ in $\boldsymbol{\theta}$, and let $\boldsymbol{\theta}_{+j}$ collect all θ_a with $a > j$, i.e., $\boldsymbol{\theta} = (\boldsymbol{\theta}_{-j}, \theta_j, \boldsymbol{\theta}_{+j})$. At each MCMC iteration s, let u be a new uniform draw. At each index j, going from 1 to k, the algorithm draws a candidate $\tilde{\theta}_j \sim q(\tilde{\theta}_j | \theta_j^{s-1})$, the proposal density. Let $\tilde{L} := \hat{p}_u(\mathbf{y} | \boldsymbol{\theta}_{-j}^s, \tilde{\theta}_j, \boldsymbol{\theta}_{+j}^{s-1})$. Then the algorithm

sets $(\theta_j^s, L_j^s) := (\tilde{\theta}_j, \tilde{L})$ with probability

$$\min\left\{\frac{\tilde{L}}{L_*} \times \frac{p(\boldsymbol{\theta}^s_{-j}, \tilde{\theta}_j, \boldsymbol{\theta}^{s-1}_{+j})}{p(\boldsymbol{\theta}^s_{-j}, \theta_j^{s-1}, \boldsymbol{\theta}^{s-1}_{+j})} \times \frac{q(\theta_j^{s-1}|\tilde{\theta}_j)}{q(\tilde{\theta}_j|\theta_j^{s-1})}, 1\right\}, \tag{11}$$

where $L_* := L^s_{j-1} \cdot \mathbb{1}(j > 1) + L^{s-1}_k \cdot \mathbb{1}(j = 1)$, or it sets $(\theta_j^s, L_j^s) := (\theta_j^{s-1}, L_*)$, otherwise. After updating θ_k, the algorithm moves on to the $(s+1)$th iteration. Under mild conditions, Tierney (1994) shows that for any measurable function h, $S^{-1}\sum_{s=1}^S h(\boldsymbol{\theta}^s) \xrightarrow{\text{a.s.}} \int_\Theta h(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}$ as S grows, regardless of the starting point $\boldsymbol{\theta}^0 \in \Theta$. A sufficient condition for this is that the posterior is absolutely continuous with respect to the proposal density; see Theorem 4.5.5 in Geweke (2005). Thus, if $q(\cdot|\cdot)$ has full support in $\mathbb{R}$, the algorithm converges. In this paper, I draw a candidate $\tilde{\theta}_j$ from $N(\theta_j^{s-1}, \sigma_j^2)$ with a prespecified σ_j^2 for all $j = 1, \ldots, k$. Then the algorithm converges and $q(\theta_j^{s-1}|\tilde{\theta}_j)/q(\tilde{\theta}_j|\theta_j^{s-1}) = 1$ in (11) due to the symmetry of the Gaussian density. This method is known as the Gaussian MH algorithm.

In practice, the performance of the Gaussian MH algorithm depends on the choice of the scale parameters $\{\sigma_j^2\}_{j\in I}$. If σ_j^2 is too small, $\tilde{\theta}_j$ will be very close to θ_j^{s-1} and the algorithm would not effectively explore Θ. If σ_j^2 is too large, the proposal density very often generates $\tilde{\theta}_j$ that is unlikely under the posterior and thereby mostly rejected: the algorithm will seldom move to another point. Typically, when the model is simple, the algorithm works fine for a wide range of the scale parameters. But tuning up $\{\sigma_j^2\}_{j\in I}$ can be hard when $\boldsymbol{\theta}$ is high dimensional, and it depends on a number of factors including the sample, the prior, and the statistical model. In this paper, I use some preliminary MCMC outcomes to determine $\{\sigma_j^2\}_{j\in I}$. Here is an example. For the artificial sample $\mathbf{z}_T$ generated above, I consider 10 statistical models, as shown in Table 1, each with a different k. For the simplest model ($k = 3$), the algorithm works well for a wide range of scale parameters. For the models with larger k, I set $\{\sigma_j^2\}_{j=1}^{k-1}$ at the posterior standard deviations obtained from the model with $k-1$ components, and set σ_k^2 at a small number such as (0.02^2).[10]

I check the convergence of the MCMC outcomes by the separated partial means test of Geweke (2005). The idea of the test is as follows. Suppose I have $\{\theta_j^s\}_{s=1}^S$ for each component j drawn from a fixed distribution. Then the null hypothesis that the mean of $\{\theta_j^s\}_{s=S/4+1}^{S/2}$ equals the mean of $\{\theta_j^s\}_{s=S\times 3/4+1}^{S}$ must be true.[11] I test the null for each component θ_j, $j \in I$. Then I have $|I|$ p-values. I terminate the MH algorithm if the smallest p-value exceeds 0.01. If not, the MH algorithm collects 100 additional draws and runs the test again until the test fails to reject. Thus, the final S is random. I run the test at $s = 20{,}000$ for the first time. Once the chain terminates, I use the last 75% of the

[10]For the samples of $T = 500$ or 1000 in Section 3.4, I set σ_j^2 to be proportional to the posterior standard deviation of θ_j obtained from the MCMC outcomes for smaller samples of $T = 200$ or 500. All the programming codes used in this paper pass the software validation test of Cook, Gelman, and Rubin (2006).

[11]In other words, the sample is divided into four equally sized blocks, and tests whether the second and the fourth have the same mean. Section 4.7 in Geweke (2005) discusses how to conduct the test when accounting for the autocorrelation.

TABLE 1. Posterior analysis for alternative specifications.

k	log Marg. Likelihood (A)	L_2 Diff. From f_0 (B)	Convergence (min p-Value) (C)	Bayes Action ρ_B (D)	Posterior Pred. Rev. (E)	95% Posterior Credible Set (F)	True Rev. at ρ_B (G)
3	−68.533	0.159	0.532	0.312	0.263	[0.253, 0.273]	0.267
4	−65.151	0.121	0.277	0.329	0.260	[0.249, 0.270]	0.267
5	−64.295	0.114	0.444	0.368	0.261	[0.249, 0.274]	0.266
6	−64.105	0.104	0.351	0.364	0.262	[0.250, 0.274]	0.267
7	−64.093	0.096	0.493	0.359	0.263	[0.251, 0.274]	0.267
8	−64.167	0.105	0.314	0.357	0.262	[0.250, 0.273]	0.267
9	−64.126	0.097	0.480	0.357	0.262	[0.250, 0.274]	0.267
10	−64.141	0.102	0.228	0.358	0.262	[0.250, 0.274]	0.267
11	−64.085	0.096	0.190	0.358	0.262	[0.250, 0.275]	0.267
12	−64.060	0.100	0.047	0.359	0.262	[0.250, 0.275]	0.267

iterations, i.e., $\{\boldsymbol{\theta}\}_{s=S/4+1}^{S}$, for inference and decision making because the convergence test suggests that they are drawn from the posterior. This decision rule is conservative because the component of the worst case has to pass the test. In the exercises of this section, the algorithm mostly passes the convergence test at $s = 20{,}000$ and, thereby, $S = 20{,}000$. For example, for the model with $k = 7$, the smallest p-value among the nine p-values is 0.493 at $s = 20{,}000$ while the average of the p-values is 0.765.

For given S, the computing time increases linearly in k.[12] For example, for the model with $k = 5$ for the given sample of $T = 200$ auctions, the computing time is 3352 seconds (≈ 56 minutes). When $k = 7$ and $k = 10$, it increases to 4522 seconds and 6510 seconds, respectively. Similarly, the computing time increases linearly in T as well. In Section 3.4, I consider larger samples. For $k = 7$, the computing time for $T = 500$ ($T = 1000$) is roughly 120% (340%) longer than the case of $T = 200$.

3.2.3 *Inference and decision making* First of all, to see what the posterior says about the data, I generate a bid sample of the same size as the data $\mathbf{z}_T$ under each $\boldsymbol{\theta}^s$ and compute its summary statistics. Panels (d)–(f) in Figure 2 show the distributions of the summary statistics under the posterior. The posterior predicts the original summary statistics more precisely and accurately than the prior. This exercise is called posterior predictive analysis; see Geweke (2005).

Second, the posterior predictive valuation density is $\int f(v|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}$, which is the most widely used Bayesian density estimate. Panel (b) in Figure 3 shows the predictive valuation density and the pointwise 2.5 and 97.5 percentiles of the posterior distribution of the valuation density (dashed lines) along with the true valuation density (plain line). This 95% posterior credible band is narrow, i.e., the inference is precise, yet it contains the true density over the entire support [0, 1], i.e., the estimate is accurate. The L_2 difference of the posterior predictive density from the true density is 0.096; see Table 1 under column B.

[12] I use an iMac 27, which has 2.9 GHz Intel Core i5 and 8 GB 1600 MHz DDR3 with OS X 10.9.2 (13C64).

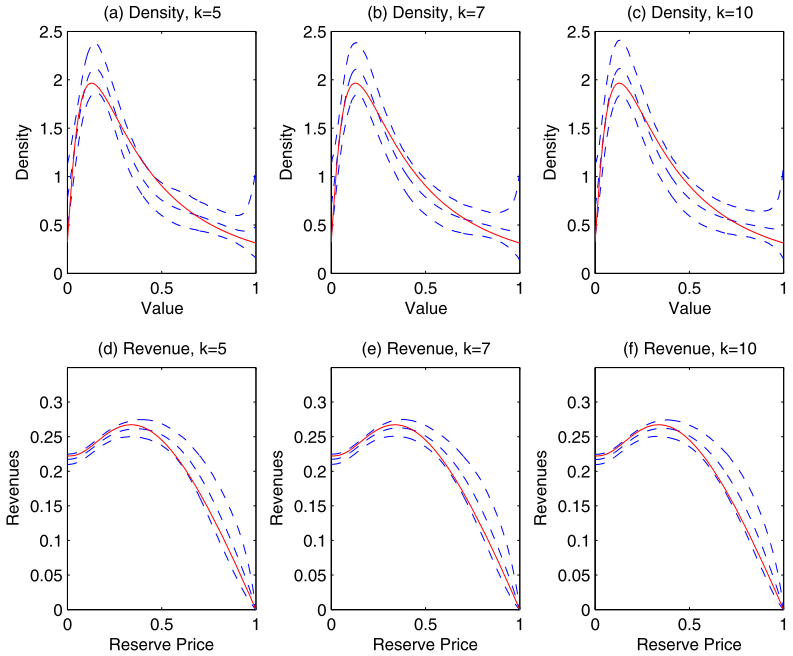

FIGURE 3. Posterior predictive density and revenue. Panels (a)–(c) [(d)–(f)] plot the posterior predictive density [revenue] functions and the 95% credible bands, (dashed lines) along with the true density [revenue] functions (solid lines) for the statistical models with the Legendre polynomials and $k \in \{5, 7, 10\}$.

Third, the posterior predictive revenue of the seller at reserve price ρ is

$$\int \Pi(\boldsymbol{\theta}, \rho)\, p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}, \tag{12}$$

where $\Pi(\boldsymbol{\theta}, \rho)$ is the revenue function under $\boldsymbol{\theta}$ at ρ.[13] Under the preference orderings of Savage (1954), Anscombe and Aumann (1963), it is optimal for the seller to maximize the posterior predictive revenue in (12); the solution to this problem is called the *Bayes action*. Moreover, the decision rule that chooses the Bayes action is called the *Bayes rule*, which is shown to be optimal under the average risk frequentist decision principle; see Berger (1985), Kim (2013). Panel (e) in Figure 3 shows the predictive revenue function and the pointwise 2.5 and 97.5 percentiles of the posterior of the revenue function (dashed lines) along with the true revenue function (plain line). The posterior analysis of the revenue function is precise and accurate. The Bayes action is $\hat{\rho}_B = 0.359$, at which the (maximized) predictive revenue is 0.263 with the 95% posterior credible interval $[0.251, 0.274]$, which includes the true revenue at $\hat{\rho}_B$, 0.267; see Table 1 under columns D–G. The true revenue is maximized at $\rho_0 = 0.340$ with the revenue 0.2672.

[13]The revenue function under the IPVP is given as $\Pi(\boldsymbol{\theta}, \rho) := n\rho[1 - F(\rho|\boldsymbol{\theta})]F(\rho|\boldsymbol{\theta})^{n-1} + n(n-1)\int_{\rho}^{\bar{v}} y[1 - F(y|\boldsymbol{\theta})]F(y|\boldsymbol{\theta})^{n-2} f(y|\boldsymbol{\theta})\, dy$; see Riley and Samuelson (1981).

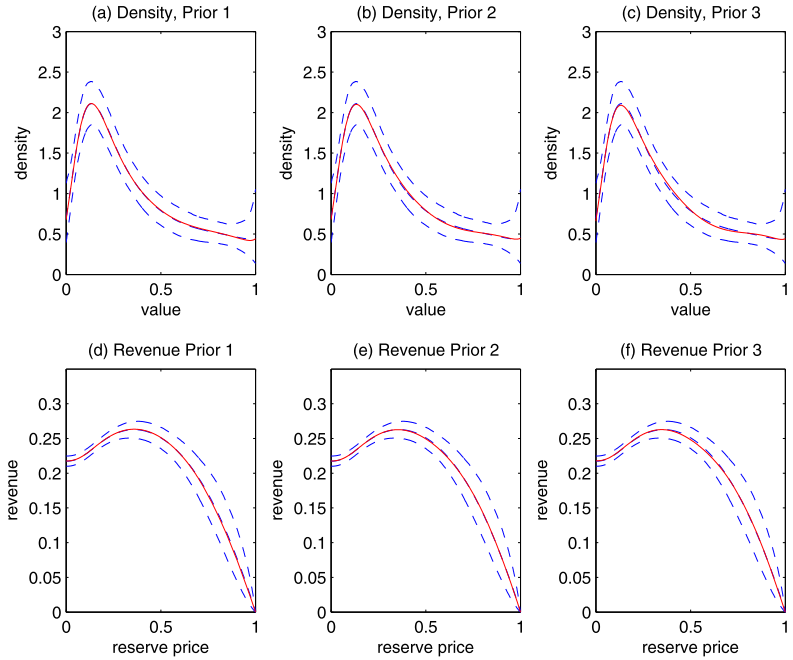

FIGURE 4. Robustness to prior. Panels (a)–(c) [(d)–(f)] reprint the dashed lines from Figure 3(b) [Figure 3(e)], i.e., the posterior predictive density [revenue] and the 95% credible band under the original prior and, additionally, put the posterior predictive density [revenue] under the alternative priors (solid line).

3.2.4 *Robustness: Prior and discretization* The posterior analysis here is robust with respect to the prior and the discretization. Recall that the prior is in the form of $p(\boldsymbol{\theta}) = \prod_{j=1}^{k} p(\theta_j)$, each $\theta_j \sim N(0, c^2 \cdot (10 \cdot 2^j)^{-1})$ with $c = 1$. I additionally consider $c \in \{0.5, 2, 5\}$—priors 1, 2, and 3, respectively. The upper (lower) panels in Figure 4 reprint the dashed lines in Figure 3(b) [Figure 3(e)], i.e., the posterior predictive density (revenue) and the 95% credible band under the original prior and, additionally, put the posterior predictive density (revenue) under the alternative priors (plain lines). Those predictive functions under alternative priors are almost identical to the ones under the original prior.

The original discretization is formed by the equidistant knot points $(b_0^*, b_1^*, \ldots, b_D^*)$ with $(b_0^*, b_D^*) = (0, b_{\max})$ and $D = 20$. I consider three coarser discretizations, i.e., $(\hat{b}_0^*, \hat{b}_1^*, \ldots, \hat{b}_{\hat{D}}^*)$;

Disc. 1. We have $(\hat{b}_0^*, \hat{b}_1^*, \hat{b}_2^*, \ldots, \hat{b}_9^*, \hat{b}_{10}^*) := (b_0^*, b_2^*, b_4^*, \ldots, b_{18}^*, b_{20}^*)$, i.e., two adjacent bins are combined.

Disc. 2. We have $(\hat{b}_0^*, \hat{b}_1^*, \ldots, \hat{b}_9^*, \hat{b}_{10}^*) := (b_0^*, b_1^*, \ldots, b_9^*, b_{10}^*)$ and $(\hat{b}_{11}^*, \hat{b}_{12}^*, \hat{b}_{13}^*, \hat{b}_{14}^*, \hat{b}_{15}^*) := (b_{12}^*, b_{14}^*, b_{16}^*, b_{18}^*, b_{20}^*)$, i.e., the first 10 bins are the same, but the 2 adjacent bins are combined for the rest.

Disc. 3. We have $(\hat{b}_0^*, \hat{b}_1^*, \ldots, \hat{b}_{13}^*, \hat{b}_{14}^*) := (b_0^*, b_1^*, \ldots, b_{13}^*, b_{14}^*)$ and $(\hat{b}_{15}^*, \hat{b}_{16}^*, \hat{b}_{17}^*) := (b_{16}^*, b_{18}^*, b_{20}^*)$, i.e., the first 14 bins are the same, but the 2 adjacent bins are combined for the rest.

Those predictive functions under alternative discretizations are almost identical to the ones under the original discretization. As a result, the graphical representation is essentially the same as Figure 4.

3.3 *Bayesian model selection*

I employ the Bayesian model comparison to choose the statistical model that fits the data the best. The model is indexed by the number of components k for a given specification of $\{\phi_j\}$. Let $\boldsymbol{\theta}_k$ denote the parameter of the model k. The marginal likelihood of model k with the sample $\mathbf{y}$ is then given as $p(\mathbf{y}|k) = \int p(\boldsymbol{\theta}_k)\, p(\mathbf{y}|\boldsymbol{\theta}_k)\, d\boldsymbol{\theta}_k$, which can be estimated using the MCMC outcomes $\{\boldsymbol{\theta}_k^s\}$ by the Bridge sampler of Fruhwirth-Schnatter (2004), Meng and Wong (1996). Let $p(k)$ be the prior probability mass function over the model space. The posterior probability of model k is then given by $p(k|\mathbf{y}) \propto p(k)\, p(\mathbf{y}|k)$. The Bayesian model selection chooses k to maximize the posterior probability. This formal model selection is often approximated by the BIC and AIC, each assuming a different prior over the model space.

Column A in Table 1 documents the log marginal likelihoods for the statistical models under consideration. The marginal likelihoods do not vary much across k. Kass and Raftery (1995) proposed a widely used rule of thumb: model k is strongly preferred to model $\tilde{k}$ if $\log p(\mathbf{y}|k) - \log p(\mathbf{y}|\tilde{k}) > 1/3$. This condition is not met for all the models with $k \geq 5$ when compared against the model with $k = 7$.[14] Moreover, the posterior analysis is robust across the statistical models. The posterior predictive densities and the posterior predictive revenue functions are all similar (see Figure 3), and the Bayes actions for choosing a reserve price ρ and the policy implications are robust (see Table 1, columns D–G).

3.4 *Comparison with previous methods*

In this section, I compare the BSL and the two-step kernel method of Guerre, Perrigne, and Vuong (2000). I fix $n = 2$ but vary $T \in \{200, 500, 1000\}$ with three different valuation distributions: log normal, exponential, and asymmetric uniform. Thus, there are nine Monte Carlo experiments, each with a different pair of sample size and distribution. In each experiment, I employ 1000 replications. In each replication, I draw a new sample of (n, T) from the fixed DGP and run the empirical methods.

Guerre, Perrigne, and Vuong (2000) showed that the valuation density is nonparametrically identified via the inverse bidding function $\xi(b) := \beta^{-1}(b) = b + G(b)[(n-1)g(b)]^{-1}$, which is a functional of the bid density g and the bid distribution function G. The method first nonparametrically estimates $\hat{g}$ and $\hat{G}$ to construct $\hat{\xi}$,

[14] I have considered alternative statistical models where B-splines construct $\{\phi_j\}$. Those models generate similar posterior analysis for the valuation density and the seller's revenue, but they are all dominated by the models with the Legendre polynomials in the marginal likelihood comparison.

and uncovers the valuation density $\hat{f}_1$ using "pseudo" values $\{\hat{v}_{i,t} := \hat{\xi}(b_{i,t})\}$ for $b_{i,t}$ in the sample. Like all other nonparametric methods, this method is sensitive to the choice of smoothing parameters, i.e., bandwidths for kernels. I consider two bandwidth configurations: one is the bandwidth selection in Guerre, Perrigne, and Vuong (2000); the other is the bandwidths that minimize the L_2-difference of the estimate from the true density.[15] The former, which I call GPV, is the most natural (or initial) choice for practitioners, whereas the latter, "Oracle GPV," gives the most accurate estimate, but it is infeasible in practice because the true density is unknown. For the BSL, I consider the statistical models with $k \in \{5, 7, 10\}$ instead of implementing the formal model selection because the posterior analysis is robust to a wide range of statistical models as shown in Section 3.3 and below as well. The experiments below show that the BSL is much more accurate than the GPV, and even better than the Oracle GPV.

3.4.1 *Log normal* The log normal density used here is defined in Section 3.2. Figure 3(a) shows the posterior predictive density estimate (middle dashed line) in the first replication of the Monte Carlo experiment with $T = 200$. This is a typical density estimate under the BSL with $k = 5$. Since I employ 1000 replications, each generating a different estimate, I have 1000 density estimates, which form the sampling distribution of the density estimator. Figure 5(a) summarizes the sampling distribution by its pointwise mean, and 2.5 and 97.5 percentiles (dashed) along with the true valuation density (solid). For the models with $k \in \{7, 10\}$, Figures 3(b),(c) and 5(b),(c) are read in the same way. For all $k \in \{5, 7, 10\}$, the BSL closely approximates the valuation density with narrow 95% *frequency* bands. Here, a distinction needs to be made: the frequency band represents the variation of the posterior predictive density, $\int f(\cdot|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{z}_T)\, d\boldsymbol{\theta}$, over the distribution of the sample $\mathbf{z}_T$, while the posterior credible band measures the variation of $f(\cdot|\boldsymbol{\theta})$ with respect to the posterior $p(\boldsymbol{\theta}|\mathbf{z}_T)$ for a given sample $\mathbf{z}_T$.

Figure 5(d) illustrates a typical GPV estimate—the one obtained from the first replication (dashed)—and panel (e) summarizes the sampling distribution of GPV by its 95% frequency band. The typical GPV estimate does not well approximate the valuation density and the frequency band is far wider than any of the BSL models. As Table 2 (first block for $T = 200$) documents, the mean integrated squared error (MISE) of the GPV estimate is about five times larger than the MISEs for the BSL models; see columns A and B.[16] Although the Oracle GPV approximates more accurately than the GPV, its MISE is still larger than those of the BSLs.

Table 2 also summarizes policy implications on choosing a reserve price for the revenue maximizing seller across the methods. For the BSL, I maximize the seller's posterior predictive revenues (12), and for the GPV, I use the nonparametric RMRP estimate of Li,

[15] Guerre, Perrigne, and Vuong (2000) used the triweight kernel $(35/32)(1-u^2)^3\mathbb{1}(|u| \le 1)$ for both bid density and valuation density. Their article proposed a combination of bandwidths $h_g = 1.06\hat{\omega}_b(nT)^{-1/5}$ for the bid density estimate and $h_f = 1.06\hat{\omega}_v(n\bar{T})^{-1/5}$ for the valuation density estimate where $\hat{\omega}_b$ and $\hat{\omega}_v$ are the sample standard deviations of the observed T bids and the estimated $\bar{T}$ values after trimming out boundaries.

[16] Let $\hat{f}_z$ be an estimate for the true density f_0. Then $\mathrm{MISE}(\hat{f}_z) = \int E_z(\hat{f}_z(x) - f_0(x))^2\, dx = \int V_z(\hat{f}_z(x))\, dx + \int (E_z\hat{f}_z(x) - f_0(x))^2\, dx = \text{variance} + \text{bias}^2$. The MISE is small only when both variance and bias are small.

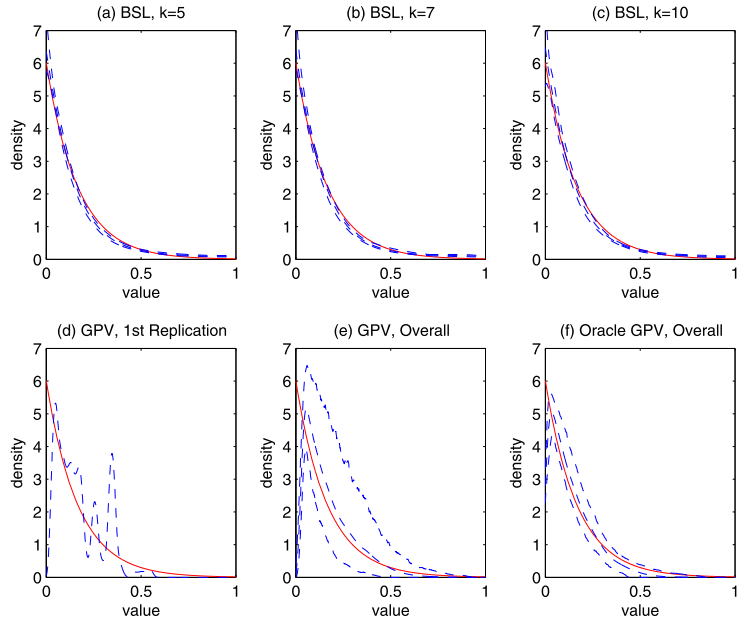

FIGURE 5. Monte Carlo, exponential, $T = 200$. The solid line in each panel is the true valuation density. Panels (a)–(c) show the sampling distributions of the BSLs by the pointwise mean and the 95% frequency band in dashed lines. Panel (d) shows a typical GPV estimate. Panels (e) and (f) are the sampling distribution of the GPV and the Oracle GPV, respectively.

TABLE 2. Summary of Monte Carlo experiments, exponential.

		MISE		Revenue		% Revenue	% of
		BSL	GPV (Oracle)	BSL	GPV	Gain	BSL > GPV
T	k	(A)	(B)	(C)	(D)	(E)	(F)
200	5	0.134	0.975 [0.385]	0.111	0.102	8.77	89.00
	7	0.127		0.111	0.102	8.79	91.70
	10	0.150		0.111	0.102	8.75	87.40
500	5	0.090	0.862 [0.322]	0.111	0.107	3.46	81.90
	7	0.094		0.111	0.107	3.48	82.90
	10	0.103		0.111	0.107	3.45	80.80
1000	5	0.057	0.792 [0.284]	0.111	0.108	2.25	86.70
	7	0.075		0.111	0.108	2.18	77.10
	10	0.075		0.111	0.108	2.18	77.10

Perrigne, and Vuong (2003), which constructs the revenue as a functional of the bid distribution functions, $\hat{G}$ and $\hat{g}$, to choose a point in the bid space, say x, that maximizes this revenue function and proposes $\hat{\rho}_x := \hat{\xi}(x) = x + \hat{G}(x)[(n-1)\hat{g}(x)]^{-1}$ as the RMRP.

In each replication, I choose a reserve price and find the associated *true* revenue under each procedure. Since I employ 1000 replications considering four procedures, I have four sets of 1000 true revenues. Columns C and D in the first block of Table 2

present their averages for $T = 200$.[17] The average revenues of the BSL models are 5.6 percent larger than the GPV (column E), and the BSL produces higher revenues than the GPV $863 \sim 889$ times out of 1000 replications (column F).

The use of the GPV can be justified by its large sample properties, e.g., as the sample size grows, the estimates converge to the true quantities. I repeat the Monte Carlo study with larger sample sizes, $T \in \{500, 1000\}$. The BSL continues to produce smaller MISEs and larger revenues than the GPV; see Table 2, columns A and B. As T grows, the 90% frequency bands get narrower for all the methods under consideration, but typical GPV estimates continue to be noisy and the frequency bands of the GPV are much wider than the frequency bands of the BSL models. I do not present graphical comparisons as they are qualitatively the same as the case of $T = 200$, i.e., Figure 5.

3.4.2 *Exponential* All the settings for the Monte Carlo experiments are the same, but values are drawn from the exponential distribution with mean $1/6$ that is truncated at 1, i.e., $f(v) \propto \exp(-6 \cdot v) \cdot \mathbb{1}(v \in [0, 1])$. The upper block of Figure 5 summarize the sampling distributions of the BSLs for $T = 200$, panel (d) plots a typical GPV estimate, and panels (e) and (f) summarize the sampling distributions of the GPV and the Oracle GPV estimators. The BSL estimates are much closer to the true density than the GPV estimate and even closer than the Oracle GPV estimate, which is confirmed by the MISE comparisons in Table 2. From such results, it is natural to expect that choosing a reserve price using the BSL should be more profitable, which is again confirmed in Table 2. I repeat the exercise for larger sample sizes $T \in \{500, 1000\}$. The BSLs unanimously produce smaller MISEs and larger revenues than the previous methods; see Table 2. I do not graphically present the results for $T \in \{500, 1000\}$ as the results are qualitatively the same as Figure 5, but with narrower frequency bands.

3.4.3 *Nonexchangeable* In practice, the prior knowledge on the shape of the valuation distribution and bidders' bidding behavior can be incorrect. To see how the BSL and GPV perform when the paradigm is misspecified, I run an additional set of Monte Carlo experiments where the valuation density violates exchangeability (2), thereby, leading to an asymmetric bidding game. In particular, I consider the vector of (v_1, v_2) that is distributed as $f(v_1, v_2) \propto \mathbb{1}(v_1 \in [0, \bar{v}_1]) \times \mathbb{1}(v_2 \in [0, \bar{v}_2])$. For the induced game, a Bayesian Nash equilibrium is characterized by the bidding strategies $\beta_i(v_i) = (1 - \sqrt{1 - k_i v_i^2})/(k_i v_i)$, where $k_i := 1/\bar{v}_i^2 - 1/\bar{v}_j^2$ for $i, j \in \{1, 2\}$, $j \neq i$; see Chapter 4 in Krishna (2002). I use $(\bar{v}_1, \bar{v}_2) = (1, 4/5)$. In each Monte Carlo replication, I draw $\{(v_{1,t}, v_{2,t})\}_{t=1}^T$ from the true valuation density and compute $b_{i,t} = \beta_i(v_{i,t})$ for every $i = 1, 2$ and $t = 1, \ldots, T$. Then I implement the empirical methods for the bid sample, $\{(b_{1,t}, b_{2,t})\}_{t=1}^T$, under the false assumption that the true density is exchangeable and the bidding game is symmetric. Figure 6 summarizes the results of the Monte Carlo experiment for $T = 200$. The solid lines on each panel represent the marginal densities of the true valuation distribution. The BSL estimates pass between the two marginals with narrow 95% frequency bands, but the GPV estimate is very different from the true density

[17] The standard deviations of the true revenues are very small, on the order of 1% of the revenue averages for the BSL specifications and 10% for the GPV.

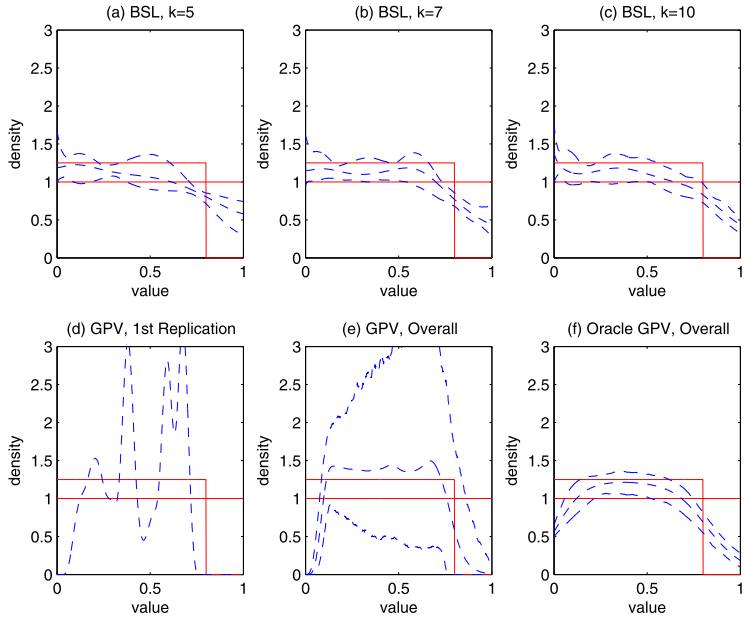

FIGURE 6. Monte Carlo, asymmetric, $T = 200$. The solid lines on each panel are the true marginal valuation density. Panels (a)–(c) show the sampling distributions of the BSLs by the pointwise mean and the 95% frequency band in dashed lines. Panel (d) shows a typical GPV estimate. Panel (e) [(f)] is the sampling distribution of the GPV [Oracle].

with a greater sampling variation. I find that the MISEs of the BSL for $k \in \{5, 7, 10\}$ are less than 50% of the MISEs of the GPV and are slightly smaller than the MISEs of the Oracle GPV.[18]

4. AFFILIATED PRIVATE VALUE PARADIGM: OCS WILDCAT AUCTIONS

In this section, I implement the BSL under the APVP to analyze the data set used in many articles including Li, Perrigne, and Vuong (2000, 2003). Especially, I explain how to flexibly impose the density affiliation (1).

4.1 *Data*

The U.S. Department of the Interior has organized auctions to sell off drilling rights on offshore oil and gas development in areas of the Gulf of Mexico. Li, Perrigne, and Vuong (2003) studied a data set of $T = 217$ auctions, each with $n = 2$ bidders between 1954 and 1969. Let $\mathbf{z}_T := \{(b_{1,t}, b_{2,t})\}_{t=1}^{T}$, the sample of the $T = 217$ bid pairs. The average bid and standard deviation are, respectively, \$145.78 and \$255.72 per acre in 1972 dollars, and the sample extremums are \$19.70 and \$2220.28. Li, Perrigne, and Vuong

[18] I do not evaluate the methods by revenues, not only because the required computation is more involved, but also because the revenue comparisons should be obvious given the MISE comparisons and the estimates shown in Figure 6.

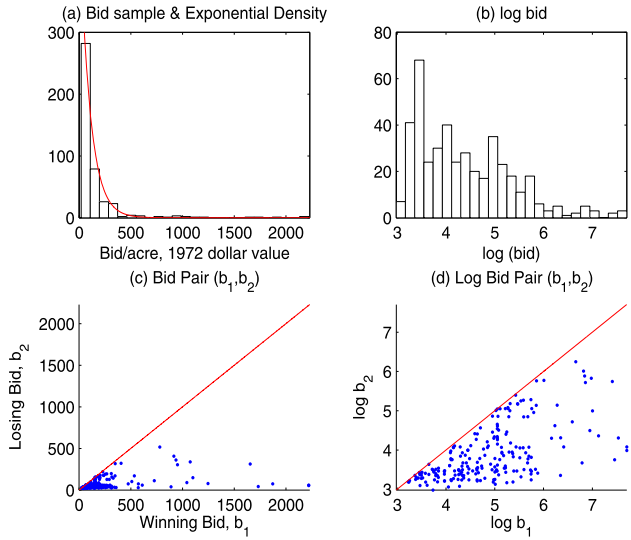

FIGURE 7. OCS wildcat bid data. Panel (a) shows histograms of the bid sample and panel (b) shows histograms of the log of bid data. Panel (c) shows a scatter diagram of the bid data with $b_1 > b_2$, and panel (d) shows a scatter diagram of the log of bid data. The sample average is \$145.78 per acre in 1972 dollar with the standard deviation of \$255.72.

(2003) nonparametrically estimated the RMRP under the APVP. In particular, it is assumed that $\{(v_{1,t}, v_{2,t})\}_{t=1}^{T} \overset{\text{i.i.d.}}{\sim} F(\cdot,\cdot) \in \mathcal{F}_{\text{APV}}$ and $b_{i,t} = \beta(v_{i,t}|F, \rho = 0)$ in (3) for all $(i,t) \in \{1,2\} \times \{1,\ldots,217\}$.[19] Figure 7(a) shows a histogram of the (marginal) bid sample. It is highly skewed to the left (skewness $= 5.09$), having a long tail to the right with some outliers. Let $b_{1,t} > b_{2,t}$ with abuse of notation. Panel (c) shows a scatter diagram of winning bids $\{b_{1,t}\}_{t=1}^{T}$ and losing bids $\{b_{2,t}\}_{t=1}^{T}$.

Panels (b) and (d) demonstrate the log bids. All the log bids are included in $[2.980, 7.233]$. Consider the set of equidistant grid points $\{b_0^*, b_1^*, \ldots, b_D^*\}$ with $(b_0^*, b_D^*) := (2.980, 7.233)$.[20] Then the collection of $B_{d,e} := [b_{d-1}^*, b_d^*] \times [b_{e-1}^*, b_e^*]$ for all $d \in \{1,\ldots,D\}$ and $e \in \{1,\ldots,d\}$ discretizes the joint log bid space (a two dimensional space) under the 45° line by $D(D+1)/2$ equally sized square bins. I use $D = 10$, i.e., 55 bins, for which I find $\bar{y} = 20$. I consider $D \in \{7, 9, 11\}$ later.

4.2 *Specification of valuation density with shape restrictions*

When $F \in \mathcal{F}_{\text{APV}}$, the entire joint density has to be specified, unlike the IPVP case, because the independence is not guaranteed. In a bivariate case, the statistical model (7) extends

[19]The empirical auction literature agrees that the actual reserve price of \$15 per acre is nonbinding as it is too low to screen some bidders with low values; see Li, Perrigne, and Vuong (2003) and refereces therein.

[20]If the original bid sample space was discretized by some equidistant grid points, most bins would have a very small number of data points or be empty because of the large tail area, and there would be a few bins near the origin with a large number of data points. So $\bar{y}$ will be large and, therefore, the number of simulation draws, i.e., $\bar{y} \cdot R$, has to be large. By discretizing the log bid space, not only do I avoid this practical problem, but also I let the prior be more informative on the tail area, which is more coarsely discretized.

to

$$\log f(v_1, v_2|\boldsymbol{\Theta}) = \sum_{i\in I}\sum_{j\in I} \theta_{i,j}\phi_i(v_1)\phi_j(v_2) + c(\boldsymbol{\Theta}), \tag{13}$$

where I is the index set that determines the number of components and $\boldsymbol{\Theta}$ is the matrix of all coefficients $\{\theta_{i,j}\}_{(i,j)\in I^2}$. To exploit the prior knowledge that $F \in \mathcal{F}_{\text{APV}}$, I employ a statistical model on which it is convenient to impose the theoretical shape restrictions, especially density affiliation (1). The statistical model (13) satisfies (1) if and only if

$$\frac{\partial^2}{\partial v_1\,\partial v_1}\sum_{i\in I}\sum_{j\in I} \theta_{i,j}\phi_i(v_1)\phi_i(v_2) \geq 0 \tag{14}$$

for *every* $(v_1, v_2) \in \mathbb{R}^2$. This gives *infinitely* many inequality conditions.

To impose this restriction allowing for a flexible interdependence structure, I employ the nonparametric method of imposing shape restrictions of Beresteanu (2007), who uses B-splines. Thus, I construct the basis functions using B-splines. Let $\phi_j(x) := \tilde{\phi}(\frac{x-j/k}{1/k})$, where $\tilde{\phi}(x) := \sum_{s=0}^{3} \frac{(-1)^s}{2}[\frac{3!}{s!(3-s)!}](x + \frac{3}{2} - s)_+^2 \cdot \mathbb{1}(x \in [-\frac{3}{2}, \frac{3}{2}])$ with $x_+ = x \cdot \mathbb{1}(x > 0)$, with $k \geq 4$. Figure 8 shows $\{\phi_j\}_{j=-1}^{k+1}$ with $k \in \{4, 6, 8\}$. Let $I := \{-1, 0, 1, \ldots, k, k+1\}$ include ϕ_{-1} and ϕ_{k+1} that are centered outside $[0, 1]$. Then almost every point in $[0, 1]$ has the same number of positive kernels, which is necessary for shape restriction. The statistical model (13) is then affiliated if the inequality condition in (14) holds at every point in the *finite* set $\{\frac{0}{k}, \frac{1}{k}, \ldots, \frac{k}{k}\} \times \{\frac{0}{k}, \frac{1}{k}, \ldots, \frac{k}{k}\}$. That is, the infinitely many constraints in (14) are represented only by $(k+1)^2$ inequalities. To simplify the notation, define the matrix of the basis functions evaluated at the points in $\{\frac{0}{k}, \frac{1}{k}, \ldots, \frac{k}{k}\}$

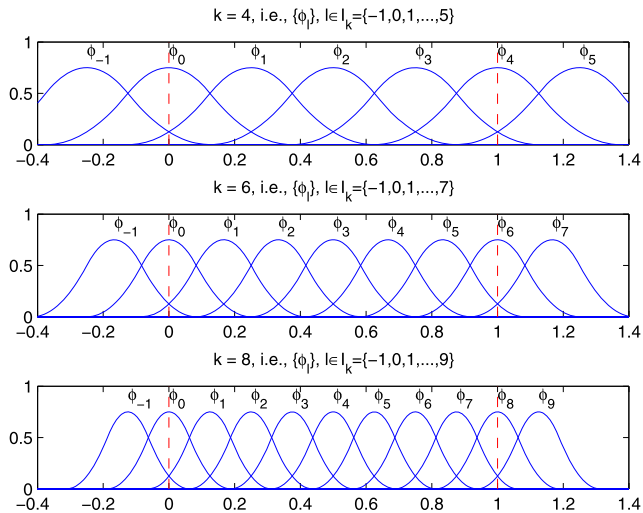

FIGURE 8. Basis functions constructed by B-splines. Each panel shows $\{\phi_\ell\}_{\ell\in I}$ for different k as shown. To impose shape restrictions, ϕ_{-1} and ϕ_{k+1} are required so that every point in $[0, 1]$ has the same number of positive kernels.

as $\boldsymbol{\Phi} := [\phi_i(j/k)]_{(i,j)\in I\times\{0,1,\ldots,k\}}$. This $(k+3)\times(k+1)$ matrix contains the $(k+3)$ basis functions, each evaluated at every point in $\{\frac{0}{k}, \frac{1}{k}, \ldots, \frac{k}{k}\}$. The inequality conditions in (14) with the B-splines is then written as $\boldsymbol{\Phi}'\boldsymbol{\Theta}\boldsymbol{\Phi} \geq \mathbf{0}_{(k+1)\times(k+1)}$ with the *element-by-element* inequality $\geq$. Finally, the statistical model (13) is exchangeable if and only if $\boldsymbol{\Theta}$ is symmetric, i.e., $\theta_{i,j} = \theta_{j,i}$.[21] Thus, the number of effective components in $\boldsymbol{\Theta}$ is $(k+3)(k+4)/2$. Moreover, since only the sample space below 45° line needs to be considered under exchangeability, the number of constraints reduces to $(k+1)(k+2)/2$.

To use the statistical model (13) with the support $[0,1]^2$, one may, in principle, rescale the bid data so that the rescaled pair of values belongs to $[0,1]^2$. Two problems arise. First, the bid distribution with such a heavy tail in Figure 7 suggests that the upper boundary of the valuation may be extremely large, but it is unknown. Second, even if the boundary is known, since it is very large, the true density of the *rescaled value* would be highly condensed on a small neighborhood of the origin $(0,0)$ in the support $[0,1]^2$. However, to approximate such a high peak, the model (13) must have many components, i.e., large k, but most coefficients are close to zero. To see this, look at Figure 8: as k increases, the support of ϕ_0 at zero gets narrower. So, to express a peak highly condensed around zero, k has to be large and only ϕ_0 has a large coefficient, but the other coefficients are all close to zero. Then the marginal likelihood of such a model would be small because the prior over a higher dimensional space is more diffused, penalizing overparametrization.

For this reason, I transform the value in such a way that the distribution of the transformed value spreads out more evenly over the unit square than the case of rescaling. Let $\tilde{F}(\cdot|\mu)$ be a strictly monotone transformation, parametrized by μ, that maps from $\mathbb{R}_+$ to $[0,1]$. I can now employ the model (13) with much smaller k to approximate the joint density of the transformed value $(x_1, x_2) := [\tilde{F}(v_1|\mu), \tilde{F}(v_2|\mu)]$. The statistical model for the original value (v_1, v_2) is then written as

$$\begin{aligned}\log f(v_1, v_2|\mu, \boldsymbol{\Theta}) = \log \tilde{f}(v_1|\mu) + \log \tilde{f}(v_2|\mu) \\ + \sum_{i\in I}\sum_{j\in I} \theta_{i,j}\phi_i(x_1)\phi_j(x_2) + c(\boldsymbol{\Theta}),\end{aligned} \tag{15}$$

where $\tilde{f}(\cdot|\mu)$ denotes a derivative of $\tilde{F}(\cdot|\mu)$. Since $\tilde{F}(\cdot|\mu)$ is a monotone mapping onto $[0,1]$, it is essentially a distribution function. For example, the exponential distribution or the log normal distribution could be used. I interpret (15) to mean that $\tilde{f}(\cdot|\mu)$ first approximates the marginal valuation density and that the additional terms improve the approximation, accounting for a flexible interdependence structure. For a given accuracy, therefore, if $\tilde{f}(\cdot|\mu)$ is close to the true density, a moderate number of components would suffice.

Since the role of $\tilde{F}$ is simply to spread out the probability mass more evenly over the support of the statistical model (13), I choose a simple parametric distribution

[21] Exchangeability has been considered in the auction literature. Especially, Li, Perrigne, and Vuong (2002) estimated the valuation density without first imposing exchangeability and then symmetrized the density estimate by taking an average over all possible permutations of the arguments in the density.

that is skewed to the left. In particular, since the histogram of the bid data (Figure 7(a)) is similar to the exponential density (solid line), believing that the value might be similarly distributed, I use the exponential distribution with density $\tilde{f}(v|\mu) = \mathbb{1}(v \geq 0)\exp(-v/\mu)/\mu$.[22] Such beliefs do not have to be exact because the additional terms explain the gap between the leading terms and the true density function. One could, in principle, choose $\tilde{F}$ formally via the Bayesian model selection discussed in Section 3.3 with a larger set of statistical models, each indexed by a pair of $(\tilde{F}, k)$, but the computing costs would then increase substantially, without much improvement in the data analysis.

4.3 *Prior specification and implications*

The statistical model for the valuation density is parametrized by $(\mu, \boldsymbol{\Theta})$ for a given k. Recall that $\theta_{i,j}$ is the coefficient attached to the kernel centered at the grid point $(i/k, j/k)$ with $(i, j) \in I^2$. Therefore, if $\theta_{i,j}$ is constant for all $(i, j) \in I^2$, the density (13) of the transformed value $(x_1, x_2) = (\tilde{F}(v_1|\mu), \tilde{F}(v_2|\mu))$ is the uniform over $[0, 1]^2$. If $\theta_{i,j}$ for a certain (i, j) is larger than other elements in $\boldsymbol{\Theta}$, the density would have a mode at $(i/k, j/k)$; the larger is the $\theta_{i,j}$, the bigger is the mode. Now consider the prior given as $\mu/1000 \sim N(1, 1) \cdot \mathbb{1}(\mu > 0)$ and

$$\boldsymbol{\Theta}|\mu \sim \mathbb{1}\big(\boldsymbol{\Phi}'\boldsymbol{\Theta}\boldsymbol{\Phi} \geq \mathbf{0}\big) \cdot \prod_{i \geq j} \big[N\big(0, \tau_{i,j}^2\big) \cdot \mathbb{1}(\theta_{i,j} = \theta_{j,i})\big], \tag{16}$$

$$\tau_{i,j} = \eta + (10 - \eta)(1 - i/k)^4 \cdot \mathbb{1}(i \geq 0) \tag{17}$$

with $\eta \in (0, 10)$. The indicators in (16) are associated with the theoretical shape restrictions; affiliation (1) and exchangeability (2). Hence, whenever the proposal function draws a parameter that violates (1), the candidate is automatically rejected because the prior at the candidate is zero; see the acceptance probability (11). All $\{\theta_{i,j}\}$ are distributed around zero. The standard deviation in (17) starts with 10 at $(i, j) = (0, 0)$, decreases quickly, and becomes η at $i = k$, the right boundary of the unit square. Therefore, the density of the transformed value (x_1, x_2) has a greater variation around the origin than the tail area: the smaller is the η, the more strongly the prior controls the tail behavior. I use first $\eta = 0.1$ and check alternative values later.

To see the implication of the prior on the data distribution, I draw $(\mu, \boldsymbol{\Theta})$ from the prior, generate a bid sample of $T = 217$ auctions, each with two bidders, under $(\mu, \boldsymbol{\Theta})$, and obtain summary statistics: the average, the standard deviation, and the skewness of the winning bids. After repeating many times, panels (a)–(c) in Figure 9 scatter the predictive distributions of the summary statistics ($\div 100$). The plain lines indicate the summary statistics of the original data $\mathbf{z}_T$ from the OCS wildcat sales.

[22] Such beliefs can be verified by a simple simulation exercise. I draw a bid sample of size 1000 from the exponential distribution with mean 1, and compute associated values by evaluating the inverse bidding function. The simulated value has a distribution similar to the exponential distribution with mean 2.4, but with a heavier tail. Roughly, its density is decreasing and convex toward the origin, and its three quartiles are (0.618, 1.694, 4.403), whereas the three quartiles of the exponential with mean 2.4 are (0.690, 1.664, 3.327).

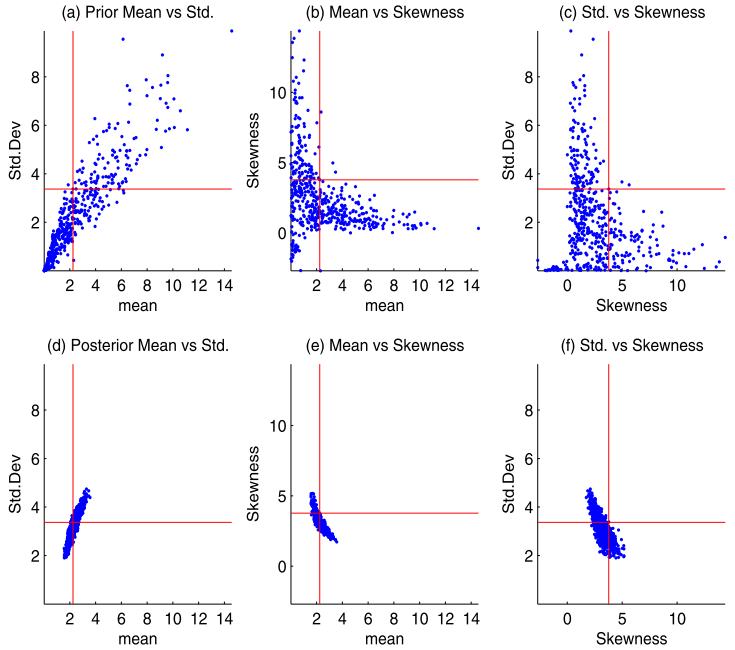

FIGURE 9. Predictive distributions of summary statistics under prior and posterior. Panels (a)–(c) demonstrate the distribution of the summary statistics (dots) of the bid data under the prior along with the summary statistics of the original data (solid line). Panels (d)–(f) do similarly for the posterior.

4.4 *Posterior computation, inference, and decision making*

I draw $\{\mu^s, \boldsymbol{\Theta}^s\}_{s=1}^S$ from the posterior using the MH algorithm; see Section 3.2.2. The parameters here include a matrix, whereas the parameters in the previous section are one dimensional. Although slightly modified, the MH algorithm is essentially the same: it updates, as before, one component at a time with a deterministic order at each MH iteration s unbiasedly estimating the likelihood by simulation.[23] The simulated values in (10) drawn from $f(v_1, v_2|\mu, \boldsymbol{\Theta})$ are now $n = 2$ dimensional; see the Appendix for the accept/reject sampling scheme. Employing the statistical model with $k = 4$, I iterate the MCMC algorithm $S = 20{,}000$ times, recording every 10th outcome, and check the convergence by the separated partial means test.[24] The smallest p-value among the 29 p-values is 0.012 and the average of the p-values is 0.426.

To see the implications of the posterior on the summary statistics, I perform the posterior predictive analysis. I generate a bid sample of size $T = 217$ and compute its

[23]Specifically, I update μ and then update $\theta_{i,j}$ in the order of $j = i, \ldots, k, k+1$ for each i, where $i = -1, 0, 1, \ldots, k, k+1$. I employ $R = 4000$.

[24]One could choose k formally by the Bayesian model section. In the previous version of this paper, I found that the log marginal likelihood is maximized at $k = 4$ and drops sharply as k increases because the number of components increases at a rate of $O(k^2)$ and thereby the statistical model gets quickly overparametrized. Moreover, when k increases, the MH algorithm must iterate more to explore dramatically expanding parameter space and, therefore, the computing time increases.

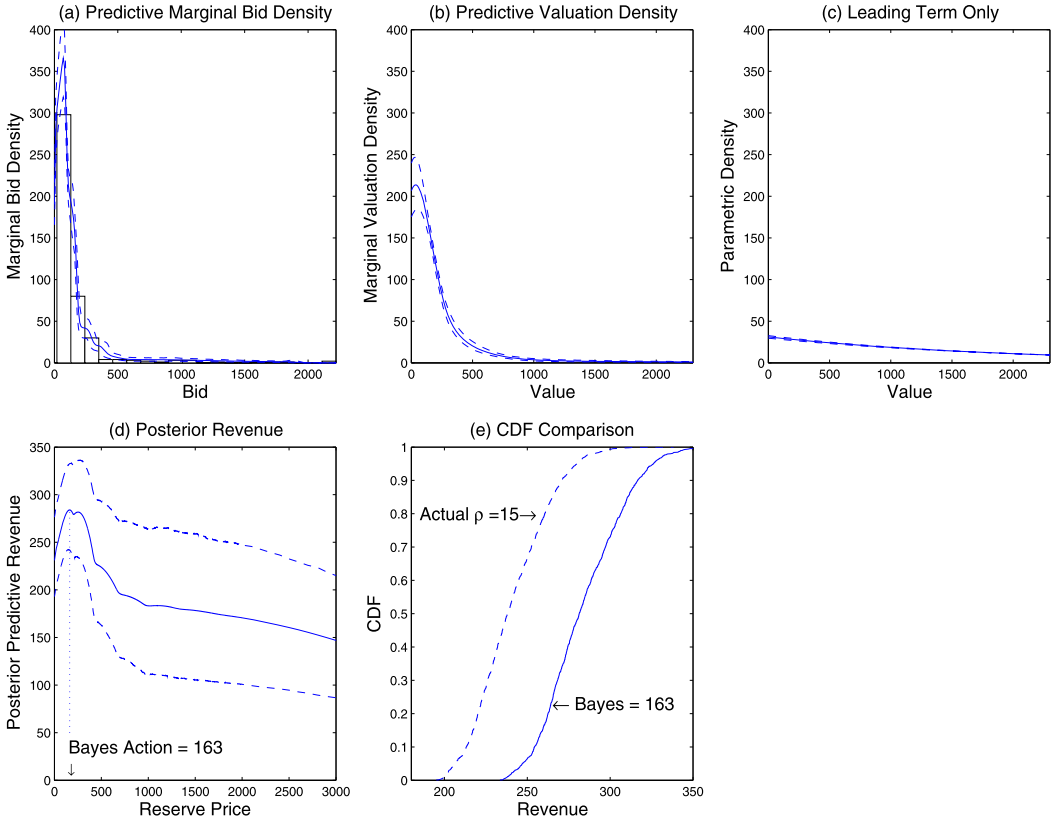

FIGURE 10. Prior predictive distributions. Panel (a) plots the data histogram along with the posterior predictive marginal bid density (solid) and a 95% posterior credible band (dashed); panel (b) plots similarly for the marginal valuation density. Panel (c) shows the valuation density that is predicted only by the leading term $\tilde{f}(\cdot|\mu)$. Panel (d) demonstrates the posterior predictive revenue (plain) along with the 95% posterior credible band (dashed). Panel (e) shows the economic significance of the Bayes action of \$163 per acre relative to the actual reserve price of \$15 per acre.

summary statistics: the average, the standard deviation, and the skewness of winning bids. After repeating this exercise many times, panels (d)–(f) in Figure 9 show the predictive distributions of the summary statistics under the posterior along with the summary statistics of the data $\mathbf{z}_T$. The distributions are more condensed than the prior and the summary statistics of the original data $\mathbf{z}_T$ may be considered as a realization under the posterior.

Panel (a) in Figure 10 plots the data histogram along with the posterior predictive marginal bid density (solid) and a 95% posterior credible band (dashed); panel (b) plots similarly for the marginal valuation density.[25] The predictive bid density well explains the sample overall, but the prior controls the tail behavior. The valuation density is

[25]The posterior distribution of the bid density is obtained via a usual kernel smoothing method over simulated data under each parameter value $(\mu, \boldsymbol{\Theta})$ drawn from the MCMC algorithm.

more diffuse toward the right than the bid densities because values are larger than bids. Panel (c) shows the valuation density that is predicted only by the leading term $\tilde{f}(\cdot|\mu)$, which suggests that a large portion of the valuation density is still explained by the additional terms in the statistical model (15).

The seller may wish to choose a reserve price to extract the largest revenue from the future auction. For simplicity, the seller's value for the auctioned tract is assumed to be zero. Let $\mathcal{A} \subset \mathbb{R}_+$ be the set of all feasible reserve prices. The seller's revenue under $(\mu, \boldsymbol{\Theta})$ with $\rho \in \mathcal{A}$ is given by $\Pi(\mu, \boldsymbol{\Theta}, \rho) := E[\beta(v_h|\rho, \mu, \boldsymbol{\Theta}) \cdot \mathbb{1}(v_h > \rho)|\mu, \boldsymbol{\Theta}]$, where $v_h := \max\{v_1, \ldots, v_n\}$. Panel (d) in Figure 10 demonstrates the posterior predictive revenue (plain) along with the 95% posterior credible band (dashed) The Bayes action (12) for this problem is the reserve price of \$163 per acre conditional on the prior and the likelihood. The posterior predictive revenue at \$163 is \$283.96 per acre with the 95% posterior credible interval of [\$241.71, \$332.61]. On the other hand, the posterior expected revenue at the actual price of \$15 is \$240.56 per acre with the 95% posterior credible interval of [\$202.42, \$285.88], which includes the average revenue (winning bid) in the sample of \$224.32. The predictive revenue distribution at the optimal choice dominates, first-order stochastically, the predictive revenue distribution at the actual reserve price as shown on panel (e). Thus, the revenue gain is *economically* significant.

I have so far employed the discretization with $D = 10$. For each $D \in \{7, 9, 11\}$, I repeat the posterior inference and decision making. The posterior predictive bid density and valuation density are very similar to the base specification with $D = 10$, i.e., Figure 10. The Bayes actions for choosing a reserve price are also similar; they are \$158, \$162, and \$161 per acre for $D = 7, 9$, and 11, respectively.

4.5 *Prior sensitivity and robust decision making*

I check the prior sensitivity on posterior inference and decision making. Unlike the Monte Carlo experiments in Section 3, the data set from the OCS wildcat sales has a long tail with some outliers, which may suggest the existence of very high values. I first consider four priors, say, priors 1–4, that all control the tail behavior but with different degrees; they use $\eta \in \{0.01, 0.2, 0.3, 0.5\}$, respectively, in equation (17). Only prior 1 with $\eta = 0.01$ controls the tail behavior more strongly than the original prior with $\eta = 0.1$. Panels (a)–(c) in Figure 11 illustrate the posterior predictive densities of the bid and the value (plain lines on panels (a) and (b)) along with the 95% posterior credible bands under the original prior (dashed). All the posterior predictive densities are included in the associated credible bands and they are very similar to the predictive densities under the original prior. The posterior predictive densities and revenues are robust as long as the prior is not substantially different. Similarly, panel (c) plots the posterior predictive revenues under the priors 1–4 along with the 95% posterior credible band under the original prior. For the priors under consideration, the posterior analysis is robust.

In addition, I consider substantially different priors with $\eta \in \{1, 2\}$, under which the posteriors predict fairly different densities of bids and valuations (panels (d) and (e)), and completely different revenues (panel (f)); the Bayes actions are 664 and 714 for each $\eta \in \{1, 2\}$, respectively. In this case, it would be useful to employ a decision method that

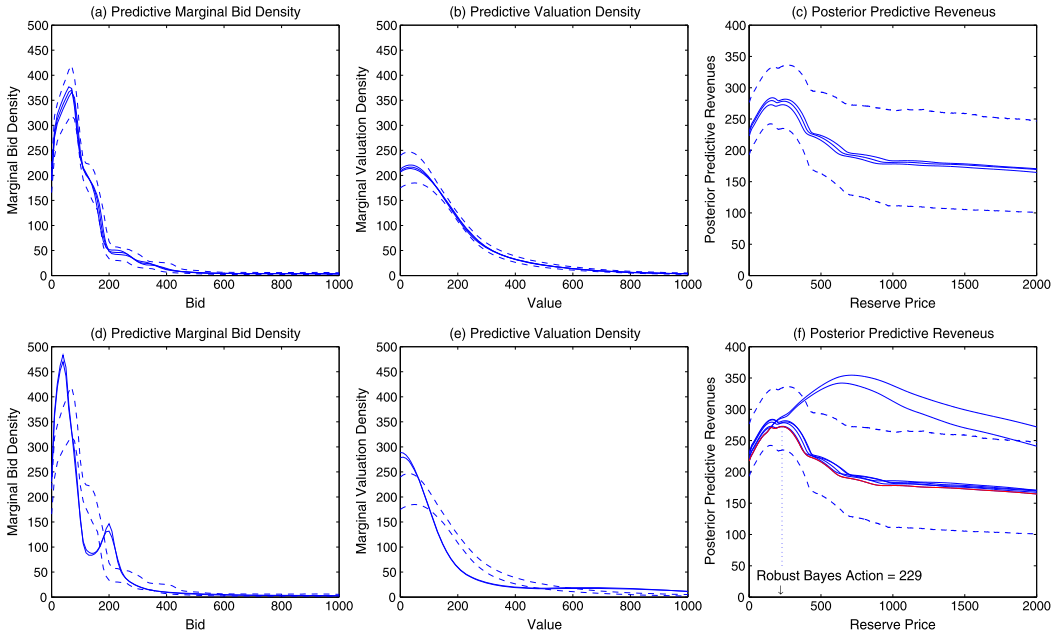

FIGURE 11. Prior sensitivity check. Panels (a)–(c) show the posterior predictive bid density, valuation density, and revenue functions, respectively, under alternative priors 1–4 that control tail behavior in the solid line along with the 95% credible bands under the original priors. Panels (d)–(f) do similarly but under additional priors that do not strongly control tail behavior.

is robust against the prior. Let Γ be the set of all reasonable priors. Under the preference orderings of Gilboa and Schmeidler (1989), it is optimal to maximize

$$\min_{p(\cdot)\in\Gamma} \int \Pi(\boldsymbol{\theta}, \rho)\left[\frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{\int p(\tilde{\boldsymbol{\theta}})p(\mathbf{y}|\tilde{\boldsymbol{\theta}})\,d\tilde{\boldsymbol{\theta}}}\right] d\boldsymbol{\theta},$$

which represents the lower envelop of these posterior predictive revenues.[26] The Γ-*maxmin rule* chooses a reserve price to maximize this lower envelop, securing the decision maker against the most pessimistic revenue. As an illustration, suppose that Γ includes all the priors considered so far. Then the Γ-maxmin rule chooses \$229; see Figure 11(f).

5. CONCLUDING REMARK: LARGE n

I conclude the paper by discussing the empirical environment with a large number of bidders, $n \geq 2$. Since this problem does not affect the BSL for the IPVP, I focus on the

[26] Aryal and Kim (2013) proposed the Γ-maxmin rule for the situations where the reserve price policy is sensitive to the prior of unidentified parameters.

APVP where the BSL would suffer the curse of dimensionality, just like all other flexible statistical methods. For $n \geq 2$, the statistical model (13) extends to

$$\log f(v_1, \ldots, v_n|\boldsymbol{\Theta}) = \sum_{i_1 \in I} \cdots \sum_{i_n \in I} \theta_{i_1,\ldots,i_n} \prod_{j=1}^{n} \phi_{i_j}(v_j) + c(\boldsymbol{\Theta}).$$

For a small n such as 2 or 3, the BSL that I have illustrated so far would be practical. As n grows, however, the number of components increases at a rate of $O(k^n)$: not only does the statistical model get quickly overparametrized, but also the computation becomes impractical. To gain tractability, therefore, it is inevitable either to employ a simple parametric model or to make a stronger assumption on the auction paradigm.

I consider the CIPVP as a reasonable alternative to the APVP because the CIPVP allows for a flexible specification, while values are still affiliated in a restrictive, but intuitive, way: the affiliation arises through an unknown common component, denoted by κ. Note that Li, Perrigne, and Vuong (2000) analyzed the OCS wildcat data under the CIPVP. The joint density of $(v_1, \ldots, v_n, \eta)$ has the form $f_\kappa(\kappa) \prod_{i=1}^{n} f_1(v_i|\kappa)$, and $f(v_1, \ldots, v_n)$ is obtained by integrating κ out. Following Li, Perrigne, and Vuong (2000), if I assume $f_1(v_i|\kappa) = f_\alpha(\alpha_i)$ with $\alpha_i = v_i/\kappa$, then I only need to specify two one dimensional densities $f_\kappa(\cdot)$ and $f_\alpha(\cdot)$ using a flexible statistical model. Let $f_\kappa(\cdot|\boldsymbol{\theta}_\kappa)$ and $f_\alpha(\cdot|\boldsymbol{\theta}_\alpha)$ be such specifications where $\boldsymbol{\theta}_\kappa$ and $\boldsymbol{\theta}_\alpha$ are the parameter vectors. Then the posterior of $(\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\alpha, \kappa)$ can be obtained.[27]

Even if the number of components does not explode, however, there would still be the curse of dimensionality regarding the sample space and its discretization, because the number of bins also increases at a rate of $O(D^n)$ and the simulation size must increase accordingly.[28] To get around this problem, I propose to use a summary statistic $H : \mathbb{R}_+^n \to \mathbb{R}_+^m$ with small m such as 2 or 3, and discretize the space of $\{(h_{1,t}, \ldots, h_{m,t}) := H(b_{1,t}, \ldots, b_{n,t})\}_{t=1}^{T}$. As before, let y_d denote the number of the summary statistics in the dth bin and let $\bar{y} := \max\{y_d\}$. Then one could approximate the associated multinomial likelihood by the simulated bids, $\{(\tilde{h}_{1,r}^j, \ldots, \tilde{h}_{m,r}^j) := H(\tilde{b}_{1,r}^j, \ldots, \tilde{b}_{n,r}^j)\}_{(r,j) \in \{1,\ldots,R\} \times \{1,\ldots,\bar{y}\}}$. This approach is closely related to the rapidly growing area of approximate Bayesian computation (ABC), where high dimensional data are replaced by a low dimensional statistic; see Marin, Pudlo, Robert, and Ryder (2012). The optimal choice of H is one of the main topics of ongoing research in the area of ABC. In general, one should consider several reasonable summary statistics and check the robustness of the analysis. For auction data analysis, there are useful theoretical results that shed a light on the problem of choosing H. For example, Esponda (2008) argues that when each bidder knows the joint density of her own bid and two top bids, she has a correct belief on the joint density of all bids in a symmetric private value paradigm. This suggests that the summary statistic

[27] In Bayesian analysis, both parameters and latent variables are unobserved quantities that are distributed as some prior, which is updated via the Bayes theorem whenever data are observed. There is no distinction between the parameters $(\boldsymbol{\theta}_\eta, \boldsymbol{\theta}_\alpha)$ and the latent variable η in the Bayesian framework.

[28] When there are too many bins, some bins would have no simulated bids unless the simulation size is large. Then the estimated likelihood would often be zero and the MH algorithm does not effectively explore the posterior.

collecting (any) one bid and two top bids would be very informative on the entire joint bid density, which identifies the joint valuation distribution; see Li, Perrigne, and Vuong (2000, 2002). In addition, when the bidders' identities are not observed, but the objective of analysis is to choose a reserve price, it would be sufficient to use only the top two bids; see Athey and Haile (2002).

Appendix: Simulation of values and evaluation of bidding functions

I explain how to draw values from the statistical model and to evaluate the bidding functions. Let $(x_0, x_1, \ldots, x_{100})$ be the equidistant knot points on $[0, 1]$ with $(x_0, x_{100}) = (0, 1)$.

A.1 *Independent private value paradigm*

First, I approximate the associated CDF via the trapezoid rule by evaluating the statistical model in (7) at each point in $(x_0, x_1, \ldots, x_{100})$. Let $\tilde{i}$ be the ith element in I. For the Legendre polynomial basis functions, $I = (1, \ldots, k)$ and $\tilde{i} = i$, but for the B-splines basis functions, $I = (0, 1, \ldots, k)$ and $\tilde{i} = i - 1$. Let Φ_x be the matrix whose (i, j) element is $\phi_{\tilde{i}}(x_j)$ and let $\boldsymbol{\theta}$ be the column vector whose ith element is $\theta_{\tilde{i}}$. Also let a_j^1 be the jth element in $\exp[\Phi_x \boldsymbol{\theta}]$, i.e., $a_j^1 = \exp[\sum_{j \in I} \theta_j \phi_j(x_j)]$ for $j = 0, 1, \ldots, 100$. Then let $a_0^2 := 0$ and $a_{\tilde{j}}^2 := a_{\tilde{j}-1}^2 + (x_j - x_{j-1})(a_j^1 + a_{j-1}^1)/2$ for $j > 0$, i.e., $a_{\tilde{j}}^2 \approx \int_0^{x_j} \exp[\sum_{j \in I} \theta_j \phi_j(t)]\, dt$. Then let $p_j := a_{\tilde{j}}^2 / a_{100}^2$, i.e., $p_j \approx F_1(x_j|\boldsymbol{\theta})$ for $j = 0, 1, \ldots, 1$. Note that $0 = p_0 < p_1 < \cdots < p_{100} = 1$. To draw $\tilde{v} \sim F_1(\cdot|\boldsymbol{\theta})$, I draw $\tilde{u} \sim \text{Uniform}[0, 1]$ and let $\tilde{j}$ be the index such that $\tilde{u} \in (p_{\tilde{j}}, p_{\tilde{j}+1})$. Finally, let

$$\tilde{v} := \left(\tilde{u} - \left\{ p_{\tilde{j}} - \left[\frac{p_{\tilde{j}+1} - p_{\tilde{j}}}{x_{\tilde{j}+1} - x_{\tilde{j}}} \right] \cdot x_{\tilde{j}} \right\} \right) \div \left[\frac{p_{\tilde{j}+1} - p_{\tilde{j}}}{x_{\tilde{j}+1} - x_{\tilde{j}}} \right],$$

i.e., $\tilde{v} \approx F_1^{-1}(\tilde{u}|\boldsymbol{\theta})$. Second, observe that for $n = 2$,

$$\begin{aligned} \hat{b}_j := \beta(x_j|\boldsymbol{\theta}) &= x_j - \int_0^{x_j} \frac{F_1(\alpha|\theta)}{F_1(x_j|\theta)}\, d\alpha \\ &\approx x_j - (2p_j)^{-1} \sum_{s=1}^{j} (x_s - x_{s-1}) \cdot (p_s - p_{s-1}). \end{aligned}$$

Redefine $\tilde{j}$ as the index such that $b_d^* \in (\hat{b}_{\tilde{j}}, \hat{b}_{\tilde{j}+1})$. Then

$$v_d^* := \left(b_d^* - \left\{ p_{\tilde{j}} - \left[\frac{\hat{b}_{\tilde{j}+1} - \hat{b}_{\tilde{j}}}{x_{\tilde{j}+1} - x_{\tilde{j}}} \right] \cdot x_{\tilde{j}} \right\} \right) \div \left[\frac{\hat{b}_{\tilde{j}+1} - \hat{b}_{\tilde{j}}}{x_{\tilde{j}+1} - x_{\tilde{j}}} \right]. \tag{18}$$

Thus, I obtain the knot points $(v_0^*, v_1^*, \ldots, v_D^*)$ in the valuation space that are associated with $(b_0^*, b_1^*, \ldots, b_D^*)$.

A.2 *Affiliated private value paradigm*

First, I employ the accept/reject sampler to draw (transformed) values from the statistical model in (13). Consider the kernel of (13) given by $\eta(v_1, v_2) := \exp\{\sum_{i\in I}\sum_{j\in I}\theta_{i,j}\phi_i(v_1)\phi_j(v_2)\}$, where $(v_1, v_2) \in [0,1]\times[0,1]$. Let $\bar{\eta}(\cdot,\cdot)$ be the piecewise linear density to approximate $\eta(\cdot,\cdot)$ as follows: for any given $(v_1, v_2) \in [0,1]\times[0,1]$,

$$\bar{\eta}(v_1, v_2) \propto \eta\left(\frac{i}{10}, \frac{j}{10}\right) + \eta\left(\frac{i+1}{10}, \frac{j}{10}\right) + \eta\left(\frac{i}{10}, \frac{j+1}{10}\right) + \eta\left(\frac{i+1}{10}, \frac{j+1}{10}\right),$$

where (i, j) is the pair of integers such that $(v_1, v_2) \in (\frac{i}{10}, \frac{i+1}{10}) \times (\frac{j}{10}, \frac{j+1}{10})$. Then I draw the (transformed) value $(\tilde{v}_1, \tilde{v}_2)$ using the accept/reject algorithm as follows:

(i) Draw a proposal $(\tilde{v}_1', \tilde{v}_2') \sim \bar{\eta}(\cdot,\cdot)$.

(ii) Let $(\tilde{v}_1, \tilde{v}_2) = (\tilde{v}_1', \tilde{v}_2')$ with probability $\frac{\eta(\tilde{v}_1', \tilde{v}_2')}{Q\bar{\eta}(\tilde{v}_1', \tilde{v}_2')}$, where $Q \geq \sup_{(v_1,v_2)\in[0,1]^2}\frac{\eta(v_1,v_2)}{\bar{\eta}(v_1,v_2)}$.

(iii) If the proposal is not accepted, go back to step 1.

Note that the original value can be drawn by evaluating $\tilde{F}^{-1}(\cdot|\mu)$ at the transformed value.

Second, I evaluate the bidding function as follows: Let $\boldsymbol{\Theta}$ be the $k+3$ symmetric square matrix whose (i, j) element is $\theta_{\tilde{i},\tilde{j}}$, where $\tilde{i}$ and $\tilde{j}$ are the ith and jth indexes in $I = \{-1, 0, 1, \ldots, k, k+1\}$, i.e., $\tilde{i} = i - 2$. Also let Φ_x be the $(k+3)\times 101$ matrix whose (i, j) element is $\phi_i(x_j)$ with the 101 equidistant knot points $\{x_j\}$ in $[0, 1]$, defined above. Let $\mathbf{f}_x := \exp(\Phi_x^\top\boldsymbol{\Theta}\Phi_x - \text{constant})$, i.e., the (i, j) element, say $f_{i,j}^x$, in $\mathbf{f}_x$ is proportional to $f(x_i, x_j|\boldsymbol{\Theta})$; see the statistical model (13).[29] Let $\boldsymbol{\delta}_x$ be the matrix whose (i, j) element is $\delta_{i,j}^x := \delta_{i-1,j}^x + 0.005\cdot(f_{i,j}^x + f_{i-1,j}^x)$ with $\delta_{i,1} \approx 0$. Let $\mathbf{a}_1$ be the diagonal element of the element-by-element product of $\mathbf{f}_x$ and $\boldsymbol{\delta}_x$, i.e., its jth element is

$$a_j^1 := \frac{f_{j,j}^x}{\delta_{i,j}^x} \approx \frac{f(x_j, x_j|\boldsymbol{\Theta})}{\displaystyle\int_0^{x_j} f(x_j, t|\boldsymbol{\Theta})\,dt} = \frac{f(x_j|x_j, \boldsymbol{\Theta})}{\displaystyle\int_0^{x_j} f(t|x_j, \boldsymbol{\Theta})\,dt}.$$

(If there is an `inf` in $\mathbf{a}_1$, such an element must be replaced by some large number.) Let $\mathbf{a}_2$ collect $a_j^2 := a_{j-1}^2 + 0.005\cdot(a_j^1 + a_{j-1}^1)$ with a small a_1^2, i.e.,

$$a_j^2 \approx \int_0^{x_j}\left[\frac{f(u|u, \boldsymbol{\Theta})}{\displaystyle\int_0^u f(t|u, \boldsymbol{\Theta})\,dt}\right]du.$$

Let $\mathbf{A}_2$ be the 101×101 matrix whose columns are all $\mathbf{a}_2$ and let $\mathbf{A}_3 := \exp\{-\max[\mathbf{A}_2^\top - \mathbf{A}_2, \mathbf{0}_{101\times 101}]\}$ so that its (i, j) element with $j > i$ is

$$a_{i,j}^3 \approx \exp\left\{-\int_{x_i}^{x_j}\left[\frac{f(u|u, \boldsymbol{\Theta})}{\displaystyle\int_0^u f(t|u, \boldsymbol{\Theta})\,dt}\right]du\right\}$$

[29] The constant is often necessary because, otherwise, some element in h will be recognized as `inf` in the machine. I use the average of all elements in $\Phi_x^\top\boldsymbol{\Theta}\Phi_x$. This constant is canceled out.

and all other elements are zero. Now let $\bar{\mathbf{v}} := \{\bar{v}_j := \tilde{F}^{-1}(x_j|\mu)\}_{j=0}^{100}$ be the knot points in the original value space. Let $\mathbf{A}_4$ be the matrix whose (i, j) element $a^4_{i,j} := a^4_{i-1,j} + (\bar{v}_i - \bar{v}_{i-1}) \cdot (a^3_{i-1,j} + a^3_{i,j})/2$ for all j. Then the jth element of the diagonal of $\mathbf{A}_4$ approximates

$$\int_0^{\bar{v}_j} \exp\left\{ -\int_\alpha^{\tilde{F}(\bar{v}_j|\mu)} \left[\frac{f(u|u, \boldsymbol{\Theta})}{\int_0^u f(t|u, \boldsymbol{\Theta})\, dt} \right] du \right\} d\alpha.$$

Let $\mathbf{a}_5$ be the diagonal of $\mathbf{A}_4$. Then the jth element of the diagonal of $\hat{\boldsymbol{b}} := \hat{\boldsymbol{v}} - \mathbf{A}_4$ approximates the equilibrium bidding function $\hat{b}_j := \beta(\bar{v}_j|\mu, \boldsymbol{\Theta})$. Then one may find the knots in the valuation space that are associated with the knots in the bid space similarly to (18).

References

Andrieu, C., A. Doucet, and R. Holenstein (2010), "Particle Markov chain Monte Carlo." *Journal of the Royal Statistical Society, Ser. B*, 72 (3), 269–342. [430, 434]

Anscombe, F. J. and R. J. Aumann (1963), "A definition of subjective probability." *Annals of Mathematical Statistics*, 34 (1), 199–205. [431, 441]

Aryal, G. and D.-H. Kim (2013), "A point decision for partially identified auction models." *Journal of Business & Economic Statistics*, 31 (4), 384–397. [431, 434, 455]

Athey, S. and P. A. Haile (2002), "Identification of standard auction models." *Econometrica*, 70 (6), 2107–2140. [457]

Beresteanu, A. (2007), "Nonparametric estimation of regression functions under restrictions on partial derivatives." Working paper. [429, 430, 449]

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer, New York. [441]

Bierens, H. J. and H. Song (2012), "Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method." *Journal of Econometrics*, 168 (1), 108–119. [430]

Campo, S., I. Perrigne, and Q. Vuong (2003), "Asymmetry in first-price auctions with affiliated private values." *Journal of Applied Econometrics*, 18 (2), 179–207. [433]

Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34 (3), 305–334. [435]

Cook, S. R., A. Gelman, and D. B. Rubin (2006), "Validation of software for Bayesian models using posterior quantiles." *Journal of Computational and Graphical Statistics*, 15 (3), 675–692. [439]

Donald, S. and H. Paarsch (1993), "Piecewise pseudo-maximum likelihood estimation in empirical models of auctions." *International Economic Review*, 34, 121–148. [429, 433]

Escobar, M. D. and M. West (1995), "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90 (430), 577–588. [435]

Esponda, I. (2008), "Information feedback in first price auctions." *RAND Journal of Economics*, 39 (2), 491–508. [456]

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1 (2), 209–230. [435]

Flury, T. and N. Shephard (2011), "Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models." *Econometric Theory*, 27 (5), 933–956. [429, 430, 434]

Fruhwirth-Schnatter, S. (2004), "Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques." *Econometrics Journal*, 7 (1), 143–167. [443]

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, New Jersey. [438, 439, 440]

Gilboa, I. and D. Schmeidler (1989), "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18, 141–153. [455]

Guerre, E., I. Perrigne, and Q. Vuong (2000), "Optimal nonparametric estimation of first-price auctions." *Econometrica*, 68 (3), 525–574. [430, 433, 434, 435, 437, 443, 444]

Henderson, D., J. A. List, D. L. Millimet, C. F. Parmeter, and M. K. Price (2012), "Empirical implementation of nonparametric first-price auction models." *Journal of Econometrics*, 168 (1), 17–28. [430]

Hirano, K. and J. Porter (2003), "Asymptotic efficiency in parametric structural models with parameter-dependent support." *Econometrica*, 71, 1307–1338. [434]

Hubbard, T. P., T. Li, and H. J. Paarsch (2012), "Semiparametric estimation in models of first-price, sealed-bid auctions with affiliation." *Journal of Econometrics*, 168 (1), 4–16. [430]

Kass, R. E. and A. E. Raftery (1995), "Bayes factors." *Journal of the American Statistical Association*, 90 (430), 773–795. [443]

Kim, D.-H. (2013), "Optimal choice of a reserve price under uncertainty." *International Journal of Industrial Organization*, 31 (5), 587–602. [431, 434, 441]

Krasnokutskaya, E. (2011), "Identification and estimation of auction models with unobserved heterogeneity." *Review of Economic Studies*, 78 (1), 293–327. [431, 433]

Krishna, V. (2002), *Auction Theory*. Academic Press, San Diego, California. [446]

Laffont, J., H. Ossard, and Q. Vuong (1995), "Econometrics of first-price auctions." *Econometrica*, 63, 953–980. [429, 433]

Li, T., I. Perrigne, and Q. Vuong (2000), "Conditionally independent private information in OCS wildcat auctions." *Journal of Econometrics*, 98, 129–161. [430, 431, 433, 447, 456, 457]

Li, T., I. Perrigne, and Q. Vuong (2002), "Structural estimation of the affiliated private value auction model." *RAND Journal of Economics*, 33, 171–193. [430, 433, 450, 457]

Li, T., I. Perrigne, and Q. Vuong (2003), "Semiparametric estimation of the optimal reserve price in first-price auctions." *Journal of Business & Economic Statistics*, 21 (1), 53–64. [430, 431, 433, 445, 447, 448]

Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012), "Approximate Bayesian computational methods." *Statistics and Computing*, 22 (6), 1167–1180. [456]

McAfee, P. R. and D. Vincent (1997), "Sequentially optimal auctions." *Games and Economic Behavior*, 18, 246–276. [431]

Meng, X.-L. and W. H. Wong (1996), "Simulating ratios of normalizing constants via a simple identity: A theoretical exploration." *Statistica Sinica*, 6, 831–860. [443]

Milgrom, P. and R. Weber (1982), "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122. [432]

Paarsch, H. (1997), "Deriving an estimate of the optimal reserve price: An application to British Columbian timber sales." *Journal of Econometrics*, 78, 333–357. [431]

Petrone, S. (1999), "Bayesian density estimation using Bernstein polynomials." *Canadian Journal of Statistics*, 27, 105–126. [435]

Riley, J. and W. Samuelson (1981), "Optimal auctions." *American Economic Review*, 71, 381–392. [441]

Savage, L. (1954), *Foundation of Statistics*. Wiley, New York. Reissued in 1972 by Dover, New York. [431, 441]

Tierney, L. (1994), "Markov chains for exploring posterior distributions." *The Annals of Statistics*, 22 (4), 1701–1728. [439]

Wasserman, L. (2006), *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer-Verlag, New York. [435]

Submitted February, 2012. Final version accepted June, 2014.