

# Identification and estimation of a bidding model for electronic auctions

BRENT R. HICKMAN

Department of Economics, University of Chicago

TIMOTHY P. HUBBARD

Department of Economics, Colby College

HARRY J. PAARSCH

Department of Economics, University of Central Florida

Because of discrete bid increments, bidders at electronic auctions engage in shading instead of revealing their valuations, which would occur under the commonly assumed second-price rule. We demonstrate that misspecifying the pricing rule can lead to biased estimates of the latent valuation distribution, and then explore identification and estimation of a model with a correctly specified pricing rule. A further challenge to econometricians is that only a lower bound on the number of participants at each auction is observed. From this bound, however, we establish nonparametric identification of the arrival process of bidders—the process that matches potential buyers to auction listings—which then allows us to identify the latent valuation distribution without imposing functional-form assumptions. We propose a computationally tractable, sieve-type estimator of the latent valuation distribution based on B-splines, and then compare two parametric models of bidder participation, finding that a generalized Poisson model cannot be rejected

---

Brent R. Hickman: [hickmanbr@gmail.com](mailto:hickmanbr@gmail.com)

Timothy P. Hubbard: [timothy.hubbard@colby.edu](mailto:timothy.hubbard@colby.edu)

Harry J. Paarsch: [hjpaarsch@gmail.com](mailto:hjpaarsch@gmail.com)

Previous drafts of this paper circulated under the title, “Investigating the Economic Importance of Pricing-Rule Misspecification in Empirical Models of Electronic Auctions.” Additional material can be found in the Supplemental Appendix provided on the journal’s website. We thank Dan Quint, Ken Judd, and Stéphane Bonhomme as well as the anonymous referees for feedback that led to significant improvements in both the substance and the presentation of the paper. We also thank Steven Mohr, to whom we are deeply indebted for the in-depth technical support that made the complex computations in this project possible. For useful feedback, we thank Jorge Balat, Michael Dinerstein, Joachim Freyberger, Srihari Govindan, Tong Li, Stephen Ryan, and Matthew Allen, seminar participants at the College of William and Mary, the University of Chicago, the University of Oklahoma, Wake Forest University, and Vanderbilt University, and conference participants at the eleventh annual International Industrial Organization Conference and the second annual Conference on Auctions, Competition, Regulation, and Public Policy. Hickman thanks the Department of Economics at the University of Wisconsin–Madison for hospitality during the period when much of this project was in development. During our research, we benefited immensely from the advice and insights of our friend and colleague Che-Lin Su, who passed away too soon on 31 July 2015. We dedicate this paper to his memory.

by the empirical distribution of observables. Our structural estimates enable us to explore information rents and optimal reserve prices on eBay.

KEYWORDS. eBay, electronic auctions, bid increments, pricing rule.

JEL CLASSIFICATION. D44, C72.

## 1. INTRODUCTION

During the past two decades, electronic auctions (EAs) have become important market mechanisms in the world economy, reducing frictions by bringing together large numbers of buyers and sellers of an incredibly diverse range of goods. Between the first quarter of 2011 and the fourth quarter of 2013, the leading host of EAs, eBay, averaged \$4.95 billion in sales per quarter from auctions alone.<sup>1</sup> EAs are of interest to economists not just because of these sales volumes, but also because the sites are data-rich environments within which fundamental economic phenomena can be studied, for example, market design, market frictions, private information rents, and price discovery. Typically modeled along the lines of traditional auction formats, EAs also display other features that derive from their online implementations and present unique challenges for empirical researchers. An especially important feature is eBay's proxy bidding procedure: a bidder reports a number to the server—kept private until the bidder is outbid—that represents the maximal amount she authorizes the server to bid, and, in turn, the server acts on the bidder's behalf to increase her standing offer up to her reported maximum. For this reason, researchers have typically modeled EAs as some variant of a second-price auction (SPA) because the winner (usually) pays a price linked to the second-highest proxy bid for the object on sale. Important qualifications apply to this procedure: when the EA software overtakes one bidder's maximum bid on behalf of another, it forces the new lead bidder, whenever possible, to surpass the former lead bidder by a commonly known, discrete amount  $\Delta$ , referred to as the *bid increment*.<sup>2</sup> Because the EA software forces the lead bidder to surpass the second-highest bid by  $\Delta$ , a necessary exception occurs when the top two proxy bids are within  $\Delta$  of one another, because a jump in the full amount  $\Delta$  would then surpass the high bidder's maximum authorized bid. In this eventuality, the price is set at the value of the high bid, that is, the winner pays her bid as in the case of a first-price auction (FPA). Thus, transacted prices on eBay (and many other electronic bidding sites) are actually determined by a hybrid second-price/first-price rule. In other words, a positive probability always exists that the highest bid is the sale price; specifically, the winner pays her bid.

Because EAs constitute one of the largest applications of auctions in history, understanding the impact of this nonstandard pricing rule is an important and useful endeavor. In fact, several other real-world situations exist where minimum bid increments

---

<sup>1</sup>These data were downloaded on 25 August 2015 from <http://www.statista.com/statistics/242267/ebays-quarterly-gross-merchandise-volume-by-sales-format>.

<sup>2</sup>Originally, bid increments may have been implemented as a security measure against cyber attacks. For example, setting an increment of  $\Delta = \$2.50$  would increase considerably the cost of automated, high-frequency bid submission relative to the case where the price is allowed to adjust by as little as a penny. Bid increments are fixed by online auction houses and are openly advertised to market participants prior to bidding; in other words, they are common knowledge.

have been noted, but ignored, when the pricing rule is modeled. For example, [McAfee and McMillan \(1996\)](#) noted that the Federal Communications Commission (FCC) spectrum auctions have minimum bid increments that are typically 5–10% of the current price of a given license. We use the simpler, single-unit EA as an opportunity to account formally for the complications minimum bid increments introduce and to highlight the consequences of ignoring this feature of the pricing rule.

[Hickman \(2010\)](#) derived the theoretical implications of the EA pricing rule and found that dominant strategies cease to exist in private-valued auction games, unlike at SPAs. In equilibrium, participants engage in bid shading, similar to that at FPAs, because with positive probability the winner's bid determines the transaction price. Hickman also presented empirical evidence that even seemingly small bid increments can play a non-trivial role in affecting bidder behavior. Using data from a sample of laptop auctions held on eBay, where the average sale price was roughly \$300 and  $\Delta$  was \$5, Hickman found that the final sale price was determined by the winner's bid nearly one-quarter of the time.

Because the structural econometric analysis of data from auctions hinges critically on the form of the theoretical equilibrium, bid shading at EAs implies that the traditional SPA approach leads to an incorrectly specified bidding model, which can induce significant bias in the estimated latent valuation distribution, which in turn can bias inference concerning various other questions relating to bidders' private values, including information rents, optimal reservation price, or predicted revenue changes from additional bidders at an auction.<sup>3</sup>

In this paper, based on the theory developed by [Hickman \(2010\)](#), we develop a complete structural model of EA bidding with a correctly specified pricing rule. The first contribution of our paper is to demonstrate the empirical relevance of the pricing-rule misspecification using a simple Monte Carlo exercise calibrated to resemble a host of realistic scenarios at EAs.

Our second contribution is to develop a fully structural econometric model designed to be implemented using actual eBay data with possible bid shading as part of the equilibrium. We demonstrate that the model is nonparametrically identified by commonly available observables at EAs. Along the way, we also overcome another, independent challenge inherent in data from EAs: in many real-world settings, it is notoriously difficult to measure precisely participation at an auction, that is, to get an accurate measure of how many bidders competed for the object on sale; eBay auctions (and EAs in general) are no different. Within the structural econometrics of auctions literature, considerable

---

<sup>3</sup>One might also wonder about the possible revenue implications of the bid increment  $\Delta$  itself. If, however, bidder valuations are independent and identically distributed, bidders are risk neutral, and the expected payment to a bidder with a value of zero is zero, then in a symmetric, increasing equilibrium of an auction game under a variety of auction formats and pricing rules, the seller will earn the same expected revenue—the well known revenue equivalence proposition, first demonstrated by [Vickrey \(1961\)](#) for uniformly distributed valuations and later proven in general by [Myerson \(1981\)](#) as well as [Riley and Samuelson \(1981\)](#). Thus, even though the EA pricing rule changes bidder behavior, it will not change expected revenue when the above assumptions hold. Whether these assumptions are empirically relevant is beyond the scope of this paper. Instead, we focus on developing and implementing an empirical bidding model that takes into account the pressure to shade bids induced by positive increments at EAs.

research exists addressing this problem, but we take a novel approach to identification by modeling explicitly the process through which some bidder identities are revealed to the econometrician and others are not.<sup>4</sup> Our model also takes into account the fact that on electronic platforms, bidders and econometricians share the same informational disadvantage: neither observes the total number of potential bidders watching at home an object for sale with intent to bid on it.

In short, we model the number of competitors  $N$  as a random variable from the bidders' perspectives and demonstrate that its distribution is nonparametrically identified by the observable lower bounds at each auction. Even if the EA pricing rule were simply the second-price rule, this problem of imperfectly observed participation would still represent a significant roadblock to empirical work. Previous researchers, such as Song (2004a, 2004b), developed methods to sidestep the need to observe  $N$  when identifying the latent valuation distribution. However, without knowing  $N$  (or at least its distribution), many counterfactual simulations based on private-values estimates remain impossible. Ours is the first paper to solve this problem: we establish nonparametric identification of a bidding model with stochastic  $N$  under the available observables, which in principle makes counterfactual simulations possible. For tractability, our estimator relies on a parametric, generalized Poisson model of the bidder arrival process. We find in our eBay data strong evidence that there is little additional gain from a fully nonparametric estimator.

Our third contribution is to propose a novel sieve-type estimator of the latent valuation distribution, which incorporates B-splines for flexibility and Galerkin methods for tractability. Our choice of B-splines permits us to match a crucial prediction of the theory: at points where bid increments change discretely from one value to another, the equilibrium bid function contains an abrupt jog. Other popular estimation methods, such as kernel smoothing or orthogonal polynomials, either cannot accommodate the needs of the identification strategy or require an inordinate amount of numerical complexity to do so. B-splines provide a remarkable degree of functional flexibility and numerical stability, which we demonstrate below.

Our estimator also incorporates Galerkin methods—commonly used in physical sciences applications—to solve the differential equation defined by a representative bidder's first-order condition. In ours and many other contexts, structural estimation involves reverse-engineering the distribution of a latent random variable (e.g., private values) from an observable distribution (e.g., bids), with an equilibrium mapping linking the two together (e.g., Bayes–Nash bidding strategies). This situation often necessitates a *nested fixed-point* approach to estimation, where the equilibrium equations are reevaluated each time model parameters are updated by the solver, that optimizes the empirical criterion function. Our use of Galerkin methods reduces substantially the computational costs of implementing the estimator in a way that is similar in spirit to the mathematical programming with equilibrium constraints (MPEC) approach pi-

---

<sup>4</sup>See Hickman, Hubbard, and Sağlam (2015) for a complete survey of previous methods for coping with an imperfectly observed number of bidders.

oneered by [Su and Judd \(2012\)](#). Rather than sequentially solving a new differential equation every time the empirical objective function is evaluated—which may occur thousands of times—our proposed method essentially requires solving a bidder’s first-order conditions *only once* by simultaneously optimizing the parameters of both the equilibrium mapping and the underlying structural primitives. Although we tailor our estimator to the particular differential equations governing equilibrium conditions for bidding in electronic auctions, our general method combining B-splines and Galerkin methods may be applicable in broader structural contexts outside of auctions as well.

We use our structural estimates to explore counterfactual simulations of the model that inform us about two aspects of online market design: we quantify the distribution of information rents to winners and explore optimal reserve prices for sellers on eBay. First, we find that despite the large numbers of buyers participating in the market, winners on eBay garner significant information rents. Second, we resolve the puzzle of why the majority of sellers on eBay set reserve prices at zero: although the expected revenue curve (as a function of reserve price) has an interior global optimum, the difference between expected revenue at a reserve price of zero and the optimal expected revenue involves mere pennies.

The remainder of the paper has the following structure: so as to place our research in context, in the following subsection we briefly review the related literature, while in Section 2, we present the basic theory of bidding, identification, and estimation for a simplified version of the model—one with a fixed, known  $N$ . This simplification permits a parsimonious Monte Carlo exercise to explore the impact of pricing-rule misspecification. The results of this misspecification analysis are presented in Section 3, while in Section 4, we extend the basic bidding model, identification results, and estimation method to handle more realistic aspects of real-world EAs, such as randomness in  $N$ , limited data availability, and nonconstant  $\Delta$ . In Section 5, we implement our estimator using a sample of data from eBay laptop auctions and discuss our empirical results, while in Section 6, we present a summary of and the conclusion from our research. In the Supplemental Appendix, available in a file on the journal website, <http://qeconomics.org/supp/233/supplement.pdf>, we provide Monte Carlo results, additional figures, results from robustness checks, and a brief primer on B-splines. We also provide our data and code at [http://qeconomics.org/supp/233/code\\_and\\_data.zip](http://qeconomics.org/supp/233/code_and_data.zip).

### 1.1 *Related literature*

[Lucking-Reiley \(2000\)](#) provided a guide to EAs for economists. In surveying Internet auctions in 1998, he found that 121 of 142 sites used the ascending-price format. Lucking-Reiley recognized the use of proxy bidding, but did not distinguish between SPAs and EAs. Similarly, [Bajari and Hortaçsu \(2004\)](#) discussed the prevalence of the ascending-price format at online auctions. Although they noted that bid increments existed and discussed the eBay case explicitly, Bajari and Hortaçsu implicitly assumed that EAs were SPAs. Other empirical researchers (such as [Roth and Ockenfels \(2002\)](#) and [Adams \(2007\)](#) as well as [Zeithammer and Adams \(2010\)](#)) recognized that bid increments exist, but nevertheless assumed that EAs are SPAs. Given that the ascending-price format is predomi-

nant at EAs, a natural first step in analyzing these auctions is to examine how such auctions are typically modeled.

Haile and Tamer (2003) presented what remains to be the most flexible open-auction format method available. Two important points should be kept in mind when comparing their research to ours: First, in our case, the EA rules are clear enough that we can use the structure to identify the model fully; Haile and Tamer focused on partial identification, in part because their assumptions are quite flexible and tailored to fitting a host of settings that may deviate from the canonical “clock” model of an ascending-price auction. Second, Haile and Tamer considered an environment within which bidders are only permitted to tender bids from a discrete grid, for example, when the auctioneer raises the price in discrete jumps and asks for audience volunteers to pay the proposed price, or when bidders cry out discrete bid jumps. At EAs, however, bidders may choose their own bids from a continuum (or at least a very fine grid), and the fixed increment instead governs the probability that they will pay their own bid on winning.

We investigate the economic and statistical importance of misspecifying the EA pricing rule as a second-price rule. In the next section, we present a simple game-theoretic model of bidding at EAs where the number of competitors is known and constant across different auctions. We begin with the most austere model possible so as to focus attention on the effects of neglecting a minimum bid increment when considering the pricing rule at EAs. Hortaçsu and Nielsen (2010) emphasized how crucial using the correct mapping between valuations and bids is to identification. We demonstrate that within the EA model this mapping is nonparametrically identified, and we present a sensitivity analysis, describing the effects of misspecification. We find that it can lead to significantly biased point estimates and policy prescriptions, for example, concerning optimal auction design. We then consider model identification and estimation under more empirically realistic assumptions when the true number of competitors is random and unknown to the econometrician.

## 2. BASELINE MODEL: FIXED PARTICIPATION

Consider a seller who seeks to divest a single object at the highest price. There are  $N$  potential buyers, each of whom has a privately known value for the object for sale. Each potential buyer views the private value of each of her competitors as a random variable  $V$ , which represents an independent draw from the cumulative distribution function (CDF)  $F_V(v)$  that is twice continuously differentiable and has a strictly positive probability density function (PDF)  $f_V(v)$  on a compact support  $[\underline{v}, \bar{v}]$ , where  $\underline{v}$  weakly exceeds zero.<sup>5</sup> This information is common knowledge to all bidders. For now, we also assume that the number of potential buyers  $N$  is fixed and known. In Section 4, we present a more realistic model of  $N$  when we establish identification and estimation using actual eBay data. For the current purposes, however, it will help to isolate the specific implications of pricing-rule misspecification if we begin with this simple model of auction participation. This

<sup>5</sup>We use uppercase Italic letters to denote random variables and lowercase Italic letters to denote realizations of these random variables.

environment is often referred to as the symmetric *independent private-values paradigm* (IPVP).

To reduce clutter, it will be useful later to let  $Z$  denote  $V_{(1:N-1)}$ , the highest valuation of  $(N - 1)$  draws from the urn  $F_V(v)$ ; in symbols,  $Z = \max\{V_1, V_2, \dots, V_{N-1}\}$ . The random variable  $Z$  represents the highest private value of a bidder's  $(N - 1)$  opponents at the auction. Given that valuations are distributed independently and identically, the CDF and PDF of  $Z$  are  $F_Z(z) = F_V(z)^{N-1}$  and  $f_Z(z) = (N - 1)F_V(z)^{N-2}f_V(z)$ , respectively.

The highest bidder is declared winner, but the price is determined by a special hybrid pricing rule, where the winner pays the smaller of either her bid or the second-highest bid plus a commonly known bid increment  $\Delta$ . Below, we characterize the symmetric equilibrium bid function and investigate the importance of the error that obtains from assuming the EA is a SPA.

### 2.1 Deriving the EA equilibrium bid function

Hickman (2010) showed that a unique, monotone, pure-strategy, Bayes–Nash equilibrium exists within the IPVP. Denote by  $\beta(v)$  the symmetric bid function at an EA with bid increment  $\Delta$ . At an EA two scenarios can determine the sale price: in the first, the highest losing bid is within  $\Delta$  of the winner's bid and she therefore pays her own submitted bid (the first-price rule is used); in the second, the two top bids are further apart and the winner pays the highest losing bid plus  $\Delta$  (a second-price rule is used). Note that the actual bid increment only enters the pricing equation directly in the event that a second-price rule is triggered, in which case a bidder's own bid does not affect the sale price. Therefore, when optimizing bids on the margin, players only consider how the bid increment controls the threshold (below their own bid) at which a first-price rule is triggered, where their own bid determines sale price. With this in mind, we define a *threshold function*,  $\tau(b)$ , as

$$\tau(b) = \begin{cases} \underline{v}, & b \leq \underline{v} + \Delta, \\ b - \Delta, & \underline{v} + \Delta \leq b. \end{cases} \tag{1}$$

Suppose that a bid of  $b$  is the highest and, therefore, the winner. Intuitively, whenever  $b$  is less than  $\underline{v} + \Delta$  the first-price rule is triggered with certainty; otherwise, this only occurs when the maximum competing bid is within  $\Delta$  of  $b$ .

To derive the optimal bidding strategy, let  $\mathbf{B} = (B_1, B_2, \dots, B_N)$  denote the vector of all bids, and let  $B_{-n}^{\max}$  denote the maximum bid among player  $n$ 's opponents. Bidder  $n$ 's optimization problem can be expressed as

$$\begin{aligned} & \max_{b \in \mathbb{R}} \{ (v_n - b) \Pr[\tau(b) < B_{-n}^{\max} \leq b] \\ & + [v_n - \mathbb{E}(B_{-n}^{\max} | B_{-n}^{\max} \leq \tau(b)) - \Delta] \Pr[B_{-n}^{\max} \leq \tau(b)] \}. \end{aligned} \tag{2}$$

The first term in the sum corresponds to winning the EA under a first-price rule, that is, when bidder  $n$ 's bid and the second-highest bid are within  $\Delta$  of each other. The second term corresponds to winning the EA under a second-price rule, that is,  $b_n$  exceeds



the second-highest bid by at least  $\Delta$ , and bidder  $n$  pays  $\Delta$  more than the second-highest bid. We can express the probability of winning under the first-price rule given equilibrium bid function  $\beta(v)$  with inverse  $\beta^{-1}(b)$  as  $\Pr[\tau(b) < B_{-n}^{\max} \leq b] = \Pr[\beta^{-1}[\tau(b)] < Z \leq \beta^{-1}(b)] = F_Z[\beta^{-1}(b)] - F_Z[\beta^{-1}[\tau(b)]]$ , and under the second-price rule as  $\Pr[B_{-n}^{\max} \leq \tau(b)] = \Pr[Z \leq \beta^{-1}[\tau(b)]] = F_Z[\beta^{-1}[\tau(b)]]$ . The conditional expectation in the second term of expression (2) is

$$\mathbb{E}[B_{-n}^{\max} | B_{-n}^{\max} \leq \tau(b)] = \frac{\int_{\underline{v}}^{\beta^{-1}[\tau(b)]} \beta(u) f_Z(u) du}{F_Z[\beta^{-1}[\tau(b)]]}.$$

Maximizing a bidder's expected surplus yields a boundary value problem that defines the function  $\beta(v)$ :

$$\beta'(v) = \frac{[v - \beta(v)]f_Z(v)}{F_Z(v) - F_Z[\beta^{-1}(\tau[\beta(v)])]} \quad \text{and} \quad \beta(\underline{v}) = \underline{v}. \quad (3)$$

Note that for bids  $\beta(v) \leq \underline{v} + \Delta$ , (3) reduces to the standard first-price equilibrium differential equation, since  $F_Z(\underline{v}) = 0$ . This is for good reason since anyone planning to bid below that point will realize that for them the first-price rule will be triggered with certainty in the event that they win. Note also that the derivative in (3) exists everywhere since  $\tau(\cdot)$  is a continuous function. This formulation, which was developed by [Hickman \(2010\)](#), is not particularly helpful to empirical researchers because the differential equation does not admit a closed-form solution. Fortunately though, after applying the inverse function theorem this differential equation can be recast solely in terms of the equilibrium inverse bid function, so equation (3) can be rewritten as

$$\frac{d\beta^{-1}(b)}{db} = \frac{F_Z[\beta^{-1}(b)] - F_Z[\beta^{-1}(b - \underline{v} - \Delta)]}{[\beta^{-1}(b) - b]f_Z[\beta^{-1}(b)]}, \quad (4)$$

where we have used the fact that  $\beta(v) = b$  and  $\beta^{-1}(b) = v$ . Thus, the EA equilibrium is characterized by a piecewise differential equation for optimal bidding on  $[0, v_\Delta)$  and  $[v_\Delta, \bar{v}]$  where  $v_\Delta = \beta^{-1}(\Delta)$ .<sup>6</sup>

As an example of how equilibrium behavior at an EA differs from the standard SPA and FPA, we depict in [Figure 1](#) the bid functions under the three pricing rules.<sup>7</sup> In this figure and below, we refer to the equilibrium bid function at an EA as  $\beta_{EA}(v)$ , the equilibrium bid function at a SPA as  $\beta_{SP}(v) = v$ , and the equilibrium bid function at an FPA as  $\beta_{FP}(v) = \mathbb{E}[Z | Z < v]$ . The important point to note is that the EA equilibrium bid function lies weakly between the SPA and the FPA bid functions. In fact, [Hickman \(2010, Theorem 3.6\)](#) showed that the EA model nests the SPA and FPA as special cases in the sense that, as  $\Delta \rightarrow 0$ , we get  $\beta_{EA}(v) \rightarrow \beta_{SP}(v)$  uniformly, and as  $\Delta \rightarrow \mathbb{E}[Z]$ , we get  $\beta_{EA}(v) \rightarrow \beta_{FP}(v)$  uniformly.

<sup>6</sup>Despite the piecewise definition,  $\beta(v)$  turns out to be continuously differentiable at the point  $v_\Delta$ , since the derivative in equation (4) approaches the first-price derivative as  $v \rightarrow v_\Delta$  from above; see [Hickman \(2010\)](#).

<sup>7</sup>This example was constructed using  $N = 3$ ,  $V \sim \text{Rayleigh}(1)$ , truncated to the interval  $[0, 1]$ , and  $\Delta = 0.1$ .



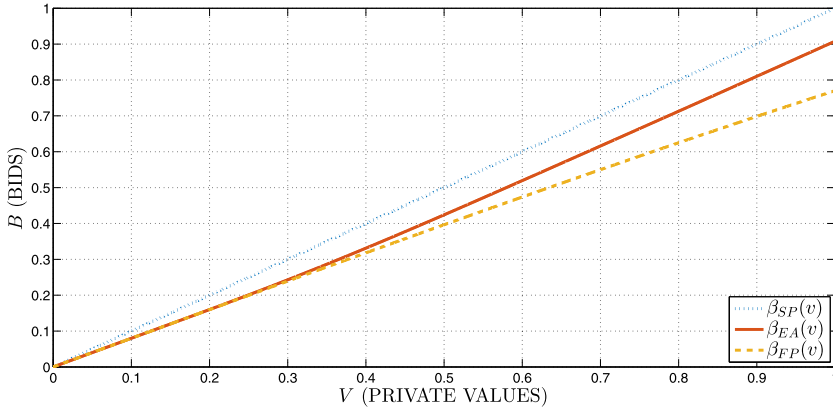


FIGURE 1. SPA, EA, and FPA equilibria.

### 2.2 Nonparametric identification and estimation with fixed $N$

The EA model is identified nonparametrically and can be estimated using a simple modification to the approach proposed by [Guerre, Perrigne, and Vuong \(2000, GPV\)](#) for models of first-price, sealed-bid auctions. To see why, let  $G_B$  denote the cumulative distribution function of equilibrium bids, and note that  $G_B(b) = F_V(v) = F_V[\beta^{-1}(b)]$  and, therefore,  $g_B(b) = \frac{f_V(v)}{\beta'(v)} = f_V[\beta^{-1}(b)] \frac{d\beta^{-1}(b)}{db}$  is the corresponding probability density function. Substituting these terms into equation (4) yields

$$v = b + \frac{G_B(b)^{N-1} - G_B[\tau(b)]^{N-1}}{(N-1)g_B(b)G_B(b)^{N-2}}. \tag{5}$$

This formulation of the equilibrium shows that each bidder’s private value is point identified from a sample of bids. This is particularly useful because, when misspecifying the pricing rule as SPA, the researcher implicitly assumes that bids *are* private valuations, whereas, given an estimate of the bid distribution and density, equation (5) allows for a simple error correction to adjust for demand shading or the difference between private values  $v$  and bids  $b$ .

In the following section, using simulated data, we present a sensitivity analysis to compare empirical results under the (incorrect) SPA assumption, which merely takes bids as private values and estimates their PDF nonparametrically, and the (correct) EA assumption. For the latter case, we construct a two-stage nonparametric estimator in the spirit of GPV: in the first stage, we construct an empirical analog of (5) using a kernel density estimator of  $g_B$  to get a sample of private-value estimates,  $\{\hat{v}\}$ ; in a second stage we then kernel-smooth the density of  $\hat{v}$  to obtain an estimate of  $f_V$ . To avoid problems of sample trimming, we use the boundary-corrected GPV (BCGPV) estimator proposed by [Hickman and Hubbard \(2015\)](#).<sup>8</sup> So as to provide a basis of comparison between the

<sup>8</sup>Traditional kernel density estimators are known to be inconsistent and biased at the boundary of the support. GPV proposed a solution to this problem that involved discarding data within a neighborhood of the sample extremes so as to preserve consistency within the interior of the support. In finite samples this

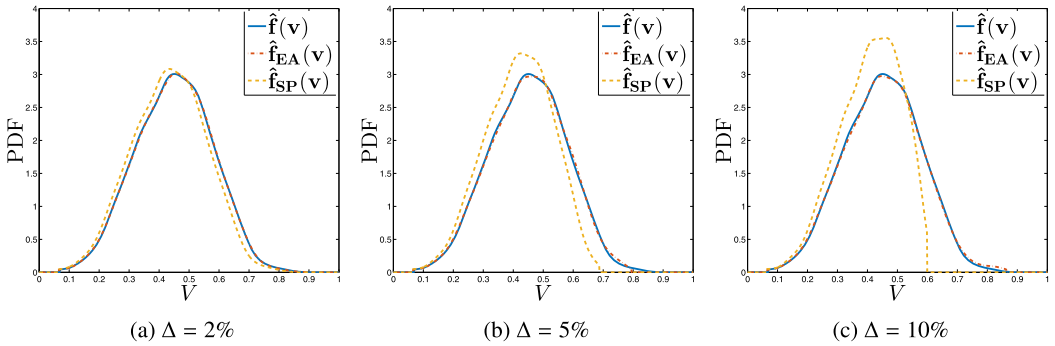


FIGURE 2. Comparison of probability density function estimates.

EA and SPA estimates, we also directly estimate the density  $f_V$  by kernel-smoothing the sample of simulated private values as well. Although such a strategy would never be available to practitioners working with real data, it provides a baseline estimate by which to judge the closeness of the other two estimators to the true distribution, given a finite sample of data with which to work.

As a precursor to our sensitivity analysis, in Figure 2, we plot three estimated PDFs. We considered a simple case in which a researcher observes 1000 auctions each having five bidders who draw valuations from a Weibull(0.5, 4.0) defined on  $[0, 1]$ . In panels (a), (b), and (c) of the figure, we display representative results under a bid increment of 0.02, 0.05, and 0.1, respectively. In each panel the solid line is the kernel-smoothed nonparametric estimate of  $f_V(v)$  based on the random valuations we generated; this is the best one could hope for from a nonparametric estimator; we denote it  $\hat{f}(V)$ . The other two estimators take as input the equilibrium bids from an EA that correspond with the randomly drawn valuations. The dash-dotted line represents  $\hat{f}_{EA}(V)$ —the nonparametric estimate under the correctly specified EA model. The dashed line represents  $\hat{f}_{SP}(V)$ —the nonparametric estimate under the misspecified SPA model. Note that  $\hat{f}_{EA}(V)$  is not visually distinct in the figure because it coincides almost exactly with  $\hat{f}(V)$ . The misspecified nonparametric estimate attains a higher peak and is shifted to the left of the optimal estimate (and the EA nonparametric estimate), with the effect becoming more pronounced the larger is  $\Delta$ . This occurs because, for a given valuation, the SPA model predicts a higher bid (bidders' weakly dominant strategy is to bid their valuation) than the EA model, which involves shading bids in the hope that the item at auction is awarded under a first-price rule. As such, the valuation implied by a given bid is lower under a SPA-assumed pricing rule. It is also worth noting that  $\hat{f}_{EA}$  is naturally handicapped relative to  $\hat{f}_{SP}$  because the former is a two-step nonparametric estimator and has a slower convergence rate than the latter, a one-step nonparametric estimator. Our sensitivity analysis will test varying sample sizes, but the figure suggests that the statistical chal-

creates several problems for inference. Hickman and Hubbard (2015) proposed an alternative approach based on boundary-corrected kernel density estimators that are uniformly consistent on the closure of the support. They described a number of attractive features of the boundary-corrected GPV estimation strategy, but for this paper the most important benefit is perhaps preserving the entire sample of data, making the one-step SPA estimator and the two-step EA estimators more comparable.

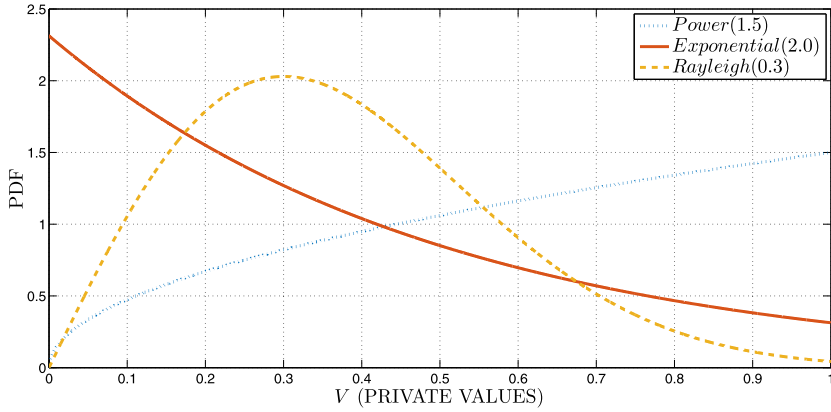


FIGURE 3. Truncated probability density functions.

lenges a two-step nonparametric estimator faces can be less than the cost of pricing-rule misspecification, something we formally pursue in the next section.

### 3. MISSPECIFICATION ANALYSIS

In this section we adopt three alternative specifications for the private value distribution—exponential, with CDF  $F(v) = 1 - \exp(-v/\theta)$ , power, with CDF  $F(v) = v^\theta$ , and Rayleigh, with CDF  $F(v) = 1 - \exp(-v^2/2\theta^2)$ —so as to explore the implications of the nonstandard pricing rule based on bid increments. For each one we assumed that  $[\underline{v}, \bar{v}]$  equals  $[0, 1]$ , and all distributions are truncated and renormalized so that the corresponding densities integrate to 1. For simplicity, in what follows we simply use the names of the untruncated distributions when referring to the truncated ones. The three PDFs are depicted in Figure 3. We chose these three in particular for our Monte Carlo exercise so as to evaluate the effect of the PDF having a mode at  $\underline{v}$ , an interior mode, and a mode at  $\bar{v}$ . Unless explicitly stated, we assumed the bid increment  $\Delta$  is 2% of the highest valuation which is consistent with a major portion of eBay bid increments.<sup>9</sup> Before presenting our Monte Carlo experiments, we perform two exploratory analyses to probe the realism of our three test distributions.

#### 3.1 Error in the equilibrium bid function

To quantify the effect of misspecifying the EA pricing rule, we computed a relative measure of error in the implied bidding function. Define the relative error from modeling and solving for the equilibrium at an EA by assuming a SPA by  $\varepsilon(v) \equiv \frac{\beta_{SP}(v) - \beta_{EA}(v)}{v} = \frac{v - \beta_{EA}(v)}{v} = 1 - \frac{\beta_{EA}(v)}{v}$ , which can be interpreted as the percentage error in the predicted bid for a given valuation as  $\beta_{SP}(v) = v$ .

<sup>9</sup>Bid increments at eBay auctions are discussed at <http://pages.ebay.com/help/buy/bid-increments.html>, which we accessed on 8/25/2015. Bid increments ranged from 5% for low-valued items (under \$5.00) to 1% for items that are selling for between \$2500.00 and \$4999.99. As an example, items with prices between \$5.00 and \$24.99 have a bid increment of \$0.50, which is 10% of \$5 and 2% of \$25.

TABLE 1. Expected relative error in bid function under SPA assumption.

Distribution	$N = 3$	$N = 5$	$N = 10$
Exponential(2.0)	0.11584	0.08945	0.06245
Rayleigh(0.3)	0.06685	0.05607	0.04162
Power(1.5)	0.04540	0.03950	0.03119

TABLE 2. Frequency of first-price rule at EA auctions.

	$\Delta$	Distribution		
		Exponential(2.0)	Rayleigh(0.3)	Power(1.5)
$N = 3$	0.01	2.90%	4.04%	3.80%
	0.02	5.80%	8.04%	7.47%
	0.05	14.65%	20.23%	18.04%
$N = 5$	0.01	3.53%	5.21%	6.72%
	0.02	7.00%	10.27%	13.07%
	0.05	17.04%	25.35%	30.07%
$N = 10$	0.01	5.04%	6.54%	13.52%
	0.02	9.90%	12.85%	25.37%
	0.05	23.71%	31.48%	52.88%

In Table 1, we summarize the expected relative error involved in assuming a SPA—which we computed as  $\mathbb{E}[\varepsilon(V)] = \int_{\underline{v}}^{\bar{v}} \varepsilon(v) f_V(v) dv$ —for each of the three distributions. The expected relative errors reported in Table 1 are all greater than 3% and can be as high as 11%. For each distribution we see that as the number of bidders increases, the expected error decreases. For an EA, the sign of this effect is not obvious: with more competition bidders behave more aggressively, but with more participants at an auction, the probability that the top two bids are within  $\Delta$  of each other also increases. The numbers in the table suggest that the former competitive effect dominates the latter probabilistic effect.

### 3.2 Bid increments and the frequency of a first-price rule

Hickman (2010) found, within a sample of 1128 eBay auctions for laptop computers, that 23.05% of final sale prices were generated by a first-price rule being triggered (because the top two bids were close together), as opposed to the often assumed second-price rule. For our test distributions, we simulated EA auctions involving 3, 5, and 10 bidders, with bid increments of 0.01, 0.02, and 0.05.<sup>10</sup> In each case we simulated 1 million auctions and computed the fraction where the first-price rule determined the transaction price.

In Table 2, we present these frequencies for the simulated scenarios. The results illustrate that, for a given distribution and a fixed number of players at auction, increasing

<sup>10</sup>By calibrating  $\bar{v} = 1$ , the bid increments correspond to a percentage of the highest possible valuation.

the bid increment increases the share of transaction prices determined by the first-price rule. Likewise, for a given distribution and a fixed bid increment, increasing the number of players at auction increases the share of transaction prices determined by the first-price rule. The power distribution dominates the exponential one, but there is no clear ranking between these two and the Rayleigh distribution. For a given  $(N, \Delta)$  pair, the power distribution involves a higher share of bid profiles triggering the first-price rule than the exponential. The distributions we consider, with realistic bid increments, are capable of generating phenomena consistent with what is observed in actual data. Moreover, assuming an EA is a SPA involves potentially serious consequences.

To evaluate the effects of misspecification, we employed a nonparametric empirical model under the correct EA assumption as well as the misspecified SPA assumption. Next, we report the results of simulation experiments in which we used these distributions to demonstrate that misspecification can lead to significantly different estimates of the latent valuation distribution. In Section 3.4, we consider the implications that model misspecification can have, estimating optimal auctions and quantifying the economic importance of the biased policy prescriptions and predictions deriving from the SPA assumption.

### 3.3 *Simulation experiments*

We conducted a series of simulation experiments in which we varied model components, including  $F_V \in \{\text{Exponential}(2.0), \text{Rayleigh}(0.3), \text{Power}(1.5)\}$ ,  $N \in \{3, 5, 10\}$ , and  $\Delta \in \{0.02, 0.05, 0.10\}$ . We also varied the sample size of auctions  $T \in \{100, 300\}$ . We simulated each instance  $S = 1000$  times and allowed the econometrician to observe the bids from all potential bidders. For each simulation  $s$ , we performed the following steps:

Step 1. We generated  $T$   $N$ -tuples of valuations from a given distribution.

Step 2. We used the true EA bid function to map these valuations into bids that we assume the researcher actually observes.

Step 3. We assumed the bids came from a SPA and estimated the model via the one-step nonparametric estimator described in the previous section.

Step 4. We assumed the bids came from an EA auction and estimated the EA model via the two-step nonparametric estimator described in the previous section.

To evaluate the statistical performance of the estimators, we constructed tests of the null hypothesis that the sample of estimated pseudo values recovered under the SPA and EA assumptions, respectively, came from the same distribution as the actual sample of simulated valuations. Specifically, we used a two-sample Kolmogorov–Smirnov test as well as an Anderson–Darling test based on [Scholz and Stephens \(1987\)](#). Results are presented in Tables A.1 and A.2, respectively, in the Supplemental Appendix. Specifically, we present results from this simulation exercise by reporting the number of null hypotheses rejected (that the two distributions are the same, at the 5% level) as well as the median  $p$ -value of the relevant test statistic for the instances involving the various distributions, bid increments, number of bidders at auction, and sample sizes.

The most robust result is that the two-step EA estimator always outperforms the one-step, albeit misspecified, SPA estimator. For the exponential and Rayleigh distribution cases, a null hypothesis is never rejected under the EA estimation. For a given bid increment, as either  $T$  or  $N$  increases, the number of times the SPA-based null hypothesis is rejected increases. Moreover, once  $\Delta$  reaches 5%, nearly every simulation involving the SPA assumption allows the null hypothesis to be rejected. Regardless of the distribution, we never reject the null hypothesis for the 2% bid increment cases involving the EA-based estimates. The power distribution case, however, is notably difficult for both estimators if  $\Delta$  or  $L \equiv NT$  is sufficiently large, though the correctly specified EA model is superior. The EA model gets rejected because the Kolmogorov–Smirnov test statistic was designed to be used on direct observations from a given CDF, whereas in the EA case, the pseudo values are estimated. Moreover, the Kolmogorov–Smirnov test statistic converges at rate  $1/\sqrt{L}$ , which is faster than the optimal nonparametric convergence rate for a two-step estimator  $(\log L/L)^{1/5}$  derived by GPV. The power distribution has a right-hand mode, where kernel-based estimators are known to have difficulty. For the power distribution, in addition to the full sample, we present a 95% sample (we drop the top 5% of each data set) and show that this issue is related to the upper boundary; the EA model is never rejected in these subsamples while the SPA is almost always for large enough  $L$  or  $\Delta$ .

Under the SPA assumption, the private value distribution is the same as the bid distribution. Thus, structural estimation methods that employ the second-price rule will uncover the population bid distribution  $G_B(b)$  as the sample size gets large. That is, if we denote the estimated valuation distribution under a SPA assumption given a sample of  $T$  auctions by  $\hat{F}_{SP}^T(V)$ , then as the number of auctions in the sample increases, we have  $\text{plim}_{T \rightarrow \infty} \hat{F}_{SP}^T(V) = G_B(b) = F_V[\beta^{-1}(b)]$ . Since  $\beta(v)$  does not equal  $v$  when  $\Delta$  is positive, it is clear that  $\hat{F}_{SP}^T(V)$  will fail to converge in probability to  $F_V(v)$ . As such, the estimated demand functions under the SPA assumption will always lie to the left of the estimated demand function under the EA assumption.

### 3.4 Economic importance of misspecification

To investigate the economic importance of the bias that obtains when a researcher estimates an EA under the SPA assumption, we considered two exercises that an econometrician might be asked to conduct: recommending a reserve price and predicting anticipated revenues if another bidder were to enter the auction. First, we computed the implied optimal auctions (involving optimally chosen reserve prices) corresponding to each estimated distribution (based off the EA and the SPA assumptions), for each simulation  $s$  for each instance of our experiment. Denote by  $\Omega$  the set of all auctions at which (i) any bidder can submit a bid as long as it is greater than some value  $r^*$ , (ii) the buyer submitting the highest bid above  $r^*$  is awarded the object, (iii) auction rules are anonymous in that each bidder is treated in the same way, and (iv) there exists a monotone, symmetric, pure-strategy, Bayes–Nash equilibrium. At any auction satisfying these four

conditions, the optimal reserve price  $r^*$  must satisfy

$$r^* = v_0 + \frac{[1 - F_V(r^*)]}{f_V(r^*)},$$

where  $v_0$  is the seller's valuation for the item at auction; see, [Riley and Samuelson \(1981\)](#) as well as [Myerson \(1981\)](#). In our simulation experiments, we assumed  $v_0$  is zero and computed the optimal reserve price implied by the EA and SPA estimates for each simulation experiment.

In Table A.3 in the Supplemental Appendix, we present the mean reservation price  $\hat{r}_{EA}$  and  $\hat{r}_{SP}$  implied by the estimates of the latent valuation distribution and density under each assumption, EA and SPA, respectively, along with their standard errors  $\hat{\sigma}_{r_{EA}}$  and  $\hat{\sigma}_{r_{SP}}$ . The true optimal reserve price for the Exponential(2.0), Rayleigh(0.3), and Power(1.5) cases are 0.36077, 0.29905, and 0.54288, respectively.<sup>11</sup> The mean reservation price under the EA auction is always within a standard deviation of the true optimal reservation price. In contrast, under the SPA assumption, the mean reservation price is regularly at least 2 standard deviations away from the truth for large enough sample size and especially for the larger bid increments, suggesting that, from a policy perspective, misspecification has important effects.

In our second exercise, we used the estimated latent value distributions to estimate what revenues would be were another bidder to participate at auction. This consideration is motivated by researchers who have pointed out that adding another bidder to the auction is often far more valuable than getting the reservation price exactly right. For example, [Bulow and Klemperer \(1996\)](#) showed that, in a world with a fixed number of participants, the optimal auction with a specific number of bidders provides less revenue than an auction with no reserve price, but one additional bidder. We consider an econometrician who might be asked to predict the expected revenues were another bidder to show up at auction. To do this, we appeal to the revenue equivalence theorem and note that we need only draw with replacement from each estimated valuation distributions and compute the average values of the second-highest valuation a sufficiently high number of times. In this way, the root cause of any discrepancy is entirely attributed to the error deriving from the original misspecification in recovering the valuation distribution from the observed bids.

In Table A.4 in the Supplemental Appendix, we present the expected revenue an econometrician would predict were another bidder to enter the auction. For the reasons documented earlier, the SPA-estimated specification always underpredicts expected revenue. Note that since our simultaneous EA model satisfies the assumptions of the revenue equivalence theorem, a practitioner should predict the same expected revenue regardless of the minimum bid increment  $\Delta$  corresponding to the data from which the underlying distribution was estimated. The results in the table illustrate the substantial variation in predictions from the SPA-based estimations for fixed values of  $N$  and

<sup>11</sup>Recall that the optimal reserve price does not depend on the number of bidders at auction, although we present estimates in Table A.3 for which the number of bidders at auction varied according to the cases we considered in our simulation experiments.



$T$  across the sections of the table in which  $\Delta$  is changing. Specifically, as  $\Delta$  increases, a practitioner assuming a SPA pricing rule would underpredict the expected revenue by larger amounts.

Our simulation results suggest that misspecifying the pricing rule can result in significantly different estimates of the latent distribution of valuations. These differences carry over to policy recommendations because the SPA-estimated valuation distributions suggest significantly different policy estimates—optimal reserve prices. There are also consequences in applying estimates from the misspecified model to predict expected revenues; for example, if another bidder were to participate at auction. In contrast, concerns about the two-step estimation process required of a correctly specified EA model do not appear to cause issues in practice as the estimator performs quite well in these dimensions.

#### 4. IDENTIFICATION AND ESTIMATION UNDER STOCHASTIC PARTICIPATION

We now extend our model to make it more compatible with real-world data from eBay. The principal challenges in empirical applications are threefold: first, the econometrician does not observe all bids, but rather a selected sample of bids that may or may not be consistent with equilibrium play. Second, the total number of bidders participating in an auction is unobserved to the econometrician, and instead, only a lower bound on total participation can be gleaned from data. Third, real-world EAs do not use constant bid increments, but bid increment schedules that are piecewise constant (that is, they discretely jump at specific, predetermined points in bid space). In this section, we provide tractable solutions to these three problems and demonstrate that a model with a correctly specified pricing rule is still nonparametrically identified under these more empirically realistic conditions. We then propose a sieve-type estimation strategy, based on B-splines, that we implement in the next section using data from eBay laptop auctions.

##### 4.1 *A bidding model with stochastic $N$*

One common characteristic of EAs is that the web-based interface makes it impossible to observe precisely the number of potential competitors within a given auction; that is, the number of users who are following an item with intent to bid on it. Difficulty in measuring  $N$  has long been a principal challenge within the empirical auctions literature, particularly when the bidders are able to observe  $N$  and adjust their bidding strategies with information unavailable to the econometrician. At an EA, however, the researcher and the bidders are on the same footing in that *neither* observes  $N$ . Therefore, we shall model participation from the perspective of a bidder as a stable stochastic process that exogenously allocates bidders to a given auction.

Specifically, let bidders view  $N$  as a random variable with probability mass function  $\rho_N(n; \boldsymbol{\lambda}) \equiv \Pr(N = n)$  indexed by a parameter vector  $\boldsymbol{\lambda}$  and assume that they do

not know the realization of  $N$  ex ante when they compute their strategic bids.<sup>12</sup> In formulating the exogenous participation process this way, we allow for  $\lambda$  to be infinite dimensional so that the distribution of  $N$  may be fully nonparametric if, for example,  $\lambda = \{\lambda_0, \lambda_1, \lambda_2, \dots\}$ , where  $\lambda_n = \Pr(N = n)$ .

Once again, consider the auction from bidder 1's perspective, and let  $M \equiv N - 1$  denote the number of opponents she faces. We also define

$$\begin{aligned} \rho_M(m; \lambda) &\equiv \Pr(M = m | N \geq 2) = \Pr(N = m + 1 | N \geq 2) \\ &= \frac{\rho_N(m + 1)}{1 - \rho_N(0) - \rho_N(1)}, \quad m \in \{1, 2, 3, \dots\}, \end{aligned}$$

as the probability that bidder 1 faces exactly  $m \geq 1$  opponents.<sup>13</sup> Just as before when  $N$  was known ex ante, a bidder's strategic decision problem within an auction is how to respond optimally to her highest rival bid. We denote the highest rival valuation and bid as random variables  $V_M$  and  $B_M$ , respectively, and we denote their respective distributions as

$$\begin{aligned} F_M(V_M) &= \sum_{m=2}^{\infty} \rho_M(m; \lambda) F_V(V_M)^m \quad \text{and} \\ G_M(B_M) &= \sum_{m=2}^{\infty} \rho_M(m; \lambda) G_B(B_M)^m. \end{aligned} \tag{6}$$

As such,  $F_M$  and  $G_M$  are weighted sums of powers of their respective parent distributions, where the weights represent the probability of a given realization for the number of potential bidders.

With these adjustments to notation, the bidding model based on a hybrid pricing rule can be easily extended to handle stochastic exogenous participation. By inserting  $F_M$  into equation (3) in place of  $F_Z$ , we get a new differential equation to redefine the equilibrium  $\beta$  for the stochastic participation case (with the same boundary condition as before); namely,

$$\beta'(v) = \frac{[v - \beta(v)]f_M(v)}{F_M(v) - F_M[\beta^{-1}(\tau[\beta(v)])]} \quad \text{and} \quad \beta(\underline{v}) = \underline{v}. \tag{7}$$

Intuitively, whether  $N$  is known or stochastic, the highest rival bid is still a random variable to which each player best responds. The only difference now is that randomness

<sup>12</sup>Song (2004b) was the first to propose a bidding model of first-price auctions where bidders view unknown  $N$  as a random variable. She then showed that assuming  $N$  is Poisson distributed makes it possible to identify the distribution of private valuations without knowing the exogenous arrival rate of bidders. However, knowledge of  $\lambda$  is needed for counterfactuals and revenue projections using structural estimates. We take a different approach here that aims to identify both the private value distribution *and* the distribution of  $N$  from observable data. The advantages of our approach are that parametric assumptions are unnecessary and structural counterfactual simulations are well defined.

<sup>13</sup>Conditioning on the event that  $N \geq 2$  comes from two facts. First, if bidder 1 exists, then she knows  $N$  is at least 1. Second, if  $N = 1$  so that bidder 1 faces no opponent, then she wins the object at a price of  $\Delta$ . But since her own bid has no bearing on the likelihood of this possibility, the  $N = 1$  scenario does not enter her decision making.

comes from two sources: a given opponent's valuation is unknown and the quantity of opponents is also unknown.

Going forward we denote the inverse bidding function by  $\xi(b) \equiv \beta^{-1}(b) : [\underline{b}, \bar{b}] \rightarrow [\underline{v}, \bar{v}]$ . As before, equation (7) can be transformed to express the inverse bidding relationship as

$$\xi(b) = v = b + \frac{G_M(b) - G_M[\tau(b)]}{g_M(b)}, \quad (8)$$

where  $g_M(b)$  is the density corresponding to the distribution of the highest rival bid. These two equations establish that the form of the inverse bid function can be inferred, as long as  $\lambda$  and the parent distribution of bids  $G_B$  can be identified. If  $\xi$  and  $G_B$  are known, then  $F_V$  is also known since  $F_V(v) = G_B[\xi(b)]$ . Thus, structural identification now hinges crucially on recovering the distribution of  $N$  from data. This we address in the next subsection.

## 4.2 Identifying exogenous participation rates

The set of observables available from an electronic platform like eBay, however, presents several challenges. We assume the econometrician does not observe all bids, from which the parent distribution  $G_B$  could be easily estimated. Rather, she observes a selected subset, being only the second-order statistic from a sample of stochastic size. Note that on eBay, the maximal bid is only observed at auctions where the first-price rule was triggered, roughly one in every five in our data. There may also be reasons to doubt whether the third-highest and other observed bid submissions are generated by equation (8), as we discuss below in Section 5.1. If the researcher has access to a broader subset of bids than what we describe here, then the estimator resulting from our identification strategy will have more statistical power. But so as to be conservative on the capabilities of our proposed method, we use only the highest losing bid (the second-order statistic).

Furthermore, the econometrician does not observe  $N$ , the total number of bidders, directly, but rather, she sees only the number of bidders who submit tenders to the server, call it  $\tilde{N}$ . This number we argue is merely a lower bound: some bidders who watch an item with intent to bid may find that their planned bid was surpassed before they get around to submitting it. Thus, the list of actual participants is passed through a natural "filter process" that withholds some of them from view *before* the econometrician is allowed to see the list of observed participants.

**4.2.1 The filter process** Underlying this idea is an assumption of simple intra-auction dynamics in the sense that ordering of bidders' submission times is random.<sup>14</sup> We assume that, prior to the auction, Nature generates a list of bidders, indexed  $\{1, 2, \dots, n\}$ , where  $n$  follows known distribution  $\rho_N(n; \lambda)$ , but each bidder is confined to an enclosed

<sup>14</sup>Note that our assumption allows us to be agnostic concerning how agents individually decide on bidding early or late within the auction. We do rule out, however, the possibility of coordination on some observable aspect of the auction so that the *relative ordering* of agents' bid submission times is systematic, rather than random. See further discussion on our assumptions that simplify intra-auction dynamics in Section 5.2 below.

cubicle so that she cannot observe the realization of  $n$ . For each  $i$ , Nature generates an independent and identically distributed (iid) private valuation  $v_i$  from  $F_V$ . Each bidder then formulates her strategic sealed bid,  $\beta(v_i)$ , and waits for Nature to come collect it from her. Nature visits each bidder in order of her index within the list, but if the highest two bids from previous tenders both exceed  $\beta(v_i)$ , then Nature skips bidder  $i$ 's submission, discarding it as if it never happened. At the conclusion, Nature reports to the econometrician the number of *recorded* bidders.

Within this simple environment, for each bidder  $i \geq 3$ , whenever the second-highest bid from among  $\{\beta(v_1), \dots, \beta(v_{i-1})\}$  exceeds  $\beta(v_i)$ , then  $i$  will not appear to have participated, even though she may have intended, *ex ante*, to submit a bid. Observing only a subset of potential bidders presents a challenge to the econometrician, but by explicitly modeling the filter process we can overcome it and still identify  $\lambda$  from observed lower bounds  $\tilde{n}$ . Moreover, if one is interested solely in modeling auction participation, then the filter process can be further simplified. Since equilibrium bidding is monotone and bidder visibility depends on the relation between rank ordering of bids and bid timing, we can recast the filter process in equivalent terms where Nature endows each bidder  $i$  with an iid quantile rank  $Q_i \sim \text{Uniform}(0, 1)$ , rather than a private valuation. Nature then walks through the (unordered) list  $\mathbf{q} = \{q_1, q_2, \dots, q_n\}$  (for a given value of  $n$ ) and reports the number to the econometrician,

$$\tilde{n} = \begin{cases} n, & \text{if } n \leq 2, \\ 2 + \sum_{i=3}^n \mathbf{1}(q_i^* < q_i), & \text{if } n \geq 3, \end{cases} \tag{9}$$

where  $q_i^*$  is the second-highest from among  $\{q_1, q_2, \dots, q_{i-1}\}$ , and  $\mathbf{1}(\cdot)$  is an indicator function. This observation facilitates simulation of the filter process *without* knowing  $F_V$  *ex ante*, which in turn makes it possible to separate identification/estimation of  $\lambda$  and  $F_V$ .

From the above description, it is easy to see that the distribution of  $\tilde{N}$  for a given value of  $N$  is invariant to changes in  $\lambda$ . Therefore, the filter process can be repeatedly simulated for arbitrary hypothetical values of  $n$ , and we can treat the conditional probabilities  $\Pr(\tilde{n}|n)$  as known quantities for arbitrary  $(\tilde{n}, n)$  pairs. Since  $\tilde{N}$  is observable, we can in turn treat its probability mass function, denoted  $\tilde{\rho}_N(\tilde{N})$ , as an observable since it can be directly estimated from data. Moreover, by the law of total probability, we have the following relationship, which establishes identification of the exogenous participation process:

$$\tilde{\rho}_N(\tilde{n}) = \sum_{n=0}^{\infty} \Pr(\tilde{n}|n) \times \rho_N(n; \lambda). \tag{10}$$

Since the above argument does not rely on an assumption that  $\lambda$  is finite-dimensional, our identification result is, in fact, *nonparametric*. In other words, observed participation together with our model of the filter process is enough to identify the distribution of  $N$  on its own, without appealing to specific functional-form assumptions on  $\rho_N$ . In practice, additional parametric assumptions (for example, specifying  $N$

as Poisson) may provide benefits such as statistical efficiency or numerical tractability, but they are not fundamentally necessary from an identification standpoint. Below in our empirical application, we shall see that the parametric generalized Poisson model (Consul and Jain (1973)), a two-parameter distribution, provides a remarkably tight fit to the data, leaving very little room (given our finite sample) for additional improvements to fit through functional-form relaxations.

### 4.3 Identifying $F_V$

With the above result in hand, nonparametric identification of the remainder of the structural model is straightforward. Let  $H(b)$  denote the ex ante distribution of the highest losing bid within an auction (the second-order statistic from a sample of stochastic size  $N$ ). Once again, since the highest losing bid is observable, we can treat  $H$  as observable since it can be directly estimated from data. It relates to the parent distribution of bids via the mapping

$$H(b) = \sum_{n=2}^{\infty} \frac{\rho_N(n; \boldsymbol{\lambda})}{1 - \rho_N(0; \boldsymbol{\lambda}) - \rho_N(1; \boldsymbol{\lambda})} (G_B(b)^n + nG_B(b)^{n-1}[1 - G_B(b)]). \quad (11)$$

For fixed  $\boldsymbol{\lambda}$  this mapping is a bijection for each  $b$  in the bid support. Therefore knowing  $\boldsymbol{\lambda}$  and  $H$  implies that the parent distribution of bids is identified.

With the above arguments in place, we can state formally our identification result. So as to fix notation, we define a model as a set of (potentially nonparametric) arrival probabilities  $\rho_N(n; \boldsymbol{\lambda})$ ,  $n = 0, 1, 2, \dots$ , and a private value distribution  $F_V$ . Moreover, we assume that the observables available to the econometrician include  $H(b)$ , the distribution of the highest losing bid, and  $\tilde{\rho}_N(\tilde{n})$ ,  $\tilde{n} = 0, 1, 2, \dots$ , the probability mass function for observed participation  $\tilde{N}$  (which is a lower bound on actual participation).

**PROPOSITION 4.1.** *Under the assumptions of Section 4.2.1, the bidding model  $(\{\rho_N(n; \boldsymbol{\lambda})\}_{n=0}^{\infty}, F_V)$  is nonparametrically identified from the observables  $(H(b), \{\tilde{\rho}_N(\tilde{n})\}_{\tilde{n}=0}^{\infty})$ .*

**PROOF.** Equation (10) establishes identification of the nonparametric bidder arrival probabilities  $\{\rho_N(n; \boldsymbol{\lambda})\}_{n=0}^{\infty}$  from the distribution of observed lower bounds under the model of the filter process described in Section 4.2.1. Given known arrival probabilities, the bijectivity of the mapping (11) establishes that the parent distribution of bids  $G_B$  is nonparametrically identified from the observables.

This in turn means that we can now construct  $G_M$  from equation (6) using the parent bid distribution and the bidder arrival probabilities.<sup>15</sup> Moreover, if  $G_M$  is known, then

<sup>15</sup>Note that  $H(b)$ , the distribution of the highest of  $(N - 1)$  draws, and  $G_M(b)$ , the distribution of the second highest of  $N$  draws, are *not* the same distribution. To see why, consider the task of repeatedly simulating order statistics based on fixed  $N$ . For each simulation,  $N$  iid realizations are generated from some distribution and stored in an unordered list. To compute the value of the second highest from the list of  $N$ , we find the maximum, discard it, and take the maximum of the remaining  $(N - 1)$  draws; that is, in this case, the highest draw from the original list of  $N$  is discarded with certainty. To compute the highest draw

we can reconstruct the inverse bidding function  $\xi(b) = \beta^{-1}(v)$  from equation (8). Finally, this also implies that the private-value distribution  $F_V$  is nonparametrically identified through the relationship  $F_V(v) = G_B[\beta(v)]$ .  $\square$

#### 4.4 Model extension: Nonconstant $\Delta$

Having presented the basic identification strategy, we now extend our model to handle a final challenge. Above, we assumed that the bid increment  $\Delta$  is constant, but at most real-world EAs  $\Delta$  changes at predetermined transition points. For simplicity, consider a single transition point, as the extension generalizes straightforwardly for two or more transition points.

Suppose we have a transition point, denoted by  $b^*$ , and two bid increments, denoted  $\Delta_1 < \Delta_2$ :  $\Delta_1$  applies when the second-highest bid is strictly less than  $b^*$ , and  $\Delta_2$  applies when it is weakly above.<sup>16</sup> Recall that when optimizing bids on the margin, players only consider how the bid increment controls the threshold (below their own bid) at which a first-price rule is triggered, so that their own bid determines sale price. Thus, the important detail to keep in mind is that when one's own bid  $b$  passes the transition point  $b^*$ , the threshold at which the first-price rule is triggered changes and moves farther away from  $b$ , since  $\Delta_1 < \Delta_2$ . With this in mind, we redefine an adjusted threshold function as

$$\tau(b) = \begin{cases} \underline{v}, & b \leq \underline{v} + \Delta_1, \\ b - \Delta_1, & \underline{v} + \Delta_1 \leq b < b^* + \Delta_1, \\ b^*, & b^* + \Delta_1 \leq b < b^* + \Delta_2, \\ b - \Delta_2, & b^* + \Delta_2 \leq b. \end{cases} \tag{12}$$

Intuitively, whenever one's own bid is less than  $b^* + \Delta_1$ , then the first-price rule will only be triggered by  $B_M$  within  $\Delta_1$  of  $b$ ; likewise, whenever one's bid is weakly above  $b^* + \Delta_2$ , the first-price rule will only be triggered by  $B_M$  within  $\Delta_2$  of  $b$ . Between these two points, the first-price threshold remains constant at  $b^*$ : if one's own  $b$  is in the interval  $[b^* + \Delta_1, b^* + \Delta_2)$ , then  $B_M$  within  $\Delta_1$  of  $b$  implies  $B_M > b^*$ , but at the same time,  $\Delta_2$  cannot be involved in a second-price outcome until  $b \geq b^* + \Delta_2$ . Note, however, that the difference  $b - \tau(b)$  steadily increases from  $\Delta_1$  to a value of  $\Delta_2$ , from which it follows that  $\tau(b)$  is a continuous function.

Given this fact, existing results by [Hickman \(2010\)](#) establish that the equilibrium bid function with a transition point is still continuous, with right- and left-hand derivatives that are the same everywhere.<sup>17</sup> Therefore, inserting the expanded version of the threshold function above into equations (7) and (8) suffices to characterize the equilibrium and establish nonparametric identification of the bidding model with transition points as well. The principal challenge that transition points will pose is on the implementation of an estimator, which we discuss in the following section.

---

from a sample of size  $(N - 1)$ , we merely discard the first observation of the unordered list and then find the maximum of the remaining  $(N - 1)$  draws. But this procedure only discards the maximum of the original  $N$  draws with probability  $(1/N)$ . Therefore, the two random variables cannot have the same distribution.

<sup>16</sup>We shall also assume for simplicity that  $\underline{v} + \Delta_1 < b^*$ .

<sup>17</sup>See Lemma 3.3.1 and Proposition 3.3 of [Hickman \(2010\)](#).

### 4.5 Two-stage estimator

In this section, we propose a simple estimator for the bidder arrival process, as well as a sieve-type estimator of the private value distribution  $F_V$  based on B-splines. Our choice of B-splines is motivated partly by their ability to accommodate elements of the empirical model flexibly, such as the abrupt change in the bid function at  $(b^* + \Delta_1)$ , which alternative methods such as kernel smoothing or global polynomials cannot easily do.

To fix notation, let  $\{y_t, \tilde{n}_t\}_{t=1}^T$  denote a sample of auctions; for each we observe the highest losing bid,  $y_t$ , and the number of observed participants,  $\tilde{n}_t$ . Although we have demonstrated nonparametric identification of the bidder arrival process, we shall focus our discussion on estimation of a model in which  $\boldsymbol{\lambda}$  is finite-dimensional. In our empirical application, we shall specify the distribution of  $N$  as generalized Poisson, or

$$\rho_n(n; \boldsymbol{\lambda}) = \frac{\lambda_1(\lambda_1 + n\lambda_2)^{n-1}}{n!} e^{-\lambda_1 - n\lambda_2}, \quad 0 < \lambda_1, |\lambda_2| < 1.$$

Later, we show that this two-parameter model leaves little room for further improvements to data fitting through more flexible functional forms: the generalized Poisson model is able to generate a distribution over the observables  $\tilde{\rho}_N(\tilde{n}; \boldsymbol{\lambda})$  that lay within the nonparametric 95% confidence bounds of the empirical distribution of  $\tilde{N}_t$ . For the present discussion, however, it suffices to consider any known parametric family  $\rho_N(n; \boldsymbol{\lambda})$  that is indexed by a finite-dimensional parameter vector,  $\boldsymbol{\lambda}$ . Where appropriate, we shall discuss further concerns and complications that would arise if the parametric assumptions are relaxed.

**4.5.1 First stage: Estimating  $\boldsymbol{\lambda}$**  We begin by constructing a simulation routine that mimics the filter process and allows us to estimate  $\Pr(\tilde{n}|n)$ . Fix a finite upper bound  $\bar{N}$  and for each  $n \in \{0, 1, 2, \dots, \bar{N}\}$  simulate  $s = 1, 2, \dots, S$  auctions wherein a list of  $n_s$  independent (unordered) quantile ranks  $\mathbf{q}_{n_s} = \{q_{1s}, \dots, q_{n_s}\}$  are drawn from the Uniform(0, 1) distribution. For each such a list, we then compute  $\tilde{n}_s$  according to the definition in equation (9). For each  $n$ , the simulated conditional frequencies are then computed as

$$\hat{\Pr}(\tilde{n}|n) = \frac{1}{S} \sum_{s=1}^S \mathbf{1}(\tilde{n}_s = \tilde{n}).$$

Note that the simulation frequencies are zero whenever  $n < \tilde{n}$ .

In a slight change of notation, we now redefine the model-generated frequencies of  $\tilde{N}$  as

$$\tilde{\rho}_N(\tilde{n}; \boldsymbol{\lambda}) = \sum_{n=0}^{\bar{N}} \hat{\Pr}(\tilde{n}|n) \times \rho_N(n; \boldsymbol{\lambda}), \quad (13)$$

and define the empirical frequencies as  $\hat{\rho}_N(k) \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{1}(\tilde{n}_t = k)$ . Finally, letting  $\tilde{\mathbf{n}} = \{k_1, k_2, \dots, k_L\}$  denote the complete set of unique observed values of  $\tilde{n}$  in the data, we



define a nonlinear least squares (NLS) estimator as the optimizer of the objective function

$$\hat{\lambda} = \operatorname{argmin}_{\lambda} \left\{ \sum_{l=1}^L [\tilde{\rho}_N(k_l; \lambda) - \hat{\rho}_N(k_l)]^2 \right\}. \tag{14}$$

In words, the estimate  $\hat{\lambda}$  is chosen to make the model-generated frequencies of observed bidders as close to the empirical frequencies as possible.<sup>18</sup>

As a practical matter, specifying  $\bar{N}$  and  $S$  involves a trade-off between computational cost and numerical accuracy. For the former, we judged  $\bar{N} = 100$  to be a sensible choice for several reasons. First, note that the Poisson probability  $\rho_N(100; 40) \approx 7.315 \times 10^{-16}$ , so the terms truncated out of the infinite sum in (13) will be on or below the order of machine precision whenever  $N$  is roughly Poisson with a parameter that is weakly less than 40. Second, while eBay auctions are known for high participation rates, 40 is still quite a large number. For example, in our empirical application with laptop computer data we observed a maximum of 11 participants in any given auction. Thus, our choice of  $\bar{N} = 100$  ensures that the finite truncation that we must impose on equation (10) will have no discernible effect for a wide array of eBay data sets. In turn, a relatively low truncation point allowed us to simulate a large number of auctions, or  $S = 10^{10}$ , which delivers at least 5 (and up to 6) digits of accuracy in each cell of the matrix  $\hat{P}r(\tilde{n}|n)$ . In other words, if the conditional probabilities in equation (13) above are expressed as percentages, then our simulation estimates are accurate to within one one-thousandth of a percentage point.<sup>19</sup> One advantage of our approach is that these need only be simu-

<sup>18</sup>An analogous nonparametric estimator could be similar, but with additional complications. Reverting back to the case where  $\lambda = \{\lambda_0, \lambda_1, \lambda_2, \dots\}$ ,  $\lambda_n = \Pr(N = n)$ , is infinite-dimensional. The main challenge now is that only finitely many elements of  $\lambda$  can be estimated with finitely much data. Therefore, for finite sample size  $T$ , we begin by choosing an upper bound,  $\bar{N}_T < \infty$ , after which we restrict  $\lambda_n = 0$  whenever  $(n - 1) > \bar{N}$  and define

$$\begin{aligned} \{\hat{\rho}_N(n; \lambda)\}_{n=0}^{\bar{N}_T} &= \operatorname{argmin}_{\lambda} \left\{ \sum_{l=1}^L [\tilde{\rho}_N(k_l; \lambda) - \hat{\rho}_N(k_l)]^2 \right\} \\ &\text{subject to } \sum_{n=0}^{\bar{N}} \hat{\rho}_N(n; \lambda) = 1. \end{aligned} \tag{15}$$

The choice of  $\bar{N}_T$  involves the usual variance–bias tradeoff. For larger  $\bar{N}_T$ , less bias arises from setting high-order elements of  $\lambda$  to zero, but as  $(\bar{N}_T/T)$  gets large the variance of the estimator will increase as well. A second challenge involves specifying the rate at which  $\bar{N}_T$  should optimally grow as  $T \rightarrow \infty$ . However, our empirical application suggests strongly that solving these problems would produce little benefit above available finite-dimensional parametric methods, so we do not address them here.

<sup>19</sup>The probabilities  $\hat{P}r(\tilde{n}|n)$  are computed as the sample mean of a Bernoulli random variable  $\mathbf{1}(\tilde{N} = \tilde{n}|N = n)$ . Since the sample mean is known to converge at rate  $\sqrt{S}$ , our simulation error is on the order of  $1/\sqrt{10^{10}} = 10^{-5}$ , but may be even less. Simulation was performed in 100 blocks of  $10^8$  simulations. As a check on accuracy, these can be used to compute 100 different estimates of the conditional probability matrix. Taking standard deviations across all 100 estimates for each  $(\tilde{n}, n)$  pair (and excluding pairs that trivially render a conditional probability of zero), we get mean and maximum standard deviations of  $1.47 \times 10^{-5}$  and  $5.55 \times 10^{-5}$ , respectively. Of course, averaging across these 100 estimates (as our final conditional probability matrix does) should further improve the precision for each  $(\tilde{n}, n)$  pair, reducing the mean and maximum standard deviations further to roughly  $1.47 \times 10^{-6}$  and  $5.55 \times 10^{-6}$ , respectively.

lated once and, given our choice of  $\bar{N}$  and  $S$ , can be reused with any data set for which the average participation rate is less than 40. A copy of the matrix of simulated conditional probabilities  $\hat{\Pr}(\tilde{n}|n)$ ,  $n = 1, \dots, 100$ , as well as MATLAB code implementation of the estimator, is available in the Supplemental Code and Data.<sup>20</sup>

Before moving on, one caveat is worth discussion. In the above proposal we have implicitly maintained a *scale invariance assumption* (SIA)—that the filter process may be simulated using quantile ranks without respect to the underlying equilibrium bid values—so as to make our estimator tractable by simplifying computation of the conditional probability matrix  $\Pr(\tilde{n}|n)$ . While the resulting simulations are quite involved (see footnotes 19 and 20), the SIA allows for  $\Pr(\tilde{n}|n)$  to be computed only once and then reused in many different settings, and, more importantly, it can be reused for each separate evaluation of the objective function in (14). Having to recompute the conditional probability matrix repeatedly during runtime would render implementation infeasible. While the SIA is approximately true, small deviations from it exist due to the bid increment,  $\Delta$ : the EA price evolution in practice may filter out an additional small number of bidders from being observed, even though their planned bids would exceed the second-highest bid from previous bid submissions. Specifically, at a given point in time, with positive probability the next bidder to arrive may wish to submit a bid that is less than  $\Delta$  above the second-highest previous bid, and in that case the posted price will have already updated to a level slightly above her planned bid.<sup>21</sup> Because of this deviation from the SIA induced by the bid increment, our data will tend to somewhat undercount the number of observed bidders, relative to what would be the case if the SIA were never violated.

To address this concern, we propose a simple, data-driven correction that allows us to still use our conditional probability matrix  $\Pr(\tilde{n}|n)$ , whose computation relied on the SIA. Note that under the complication introduced by  $\Delta$ , there is a positive fraction of the time that  $k$  bidders are observed in the data, but if the SIA were perfectly true there would actually have been at least  $(k + 1)$  observed bidders. Fortunately, our data contain some additional information that provides clues as to how this process played out. Within each eBay auction, one can observe the complete price path during the life of the auction, from which one can also deduce whether the final sale price involved the triggering of a first-price rule—this happened whenever the value of the final price adjustment was strictly less than  $\Delta$ . Thus, we add to the set of observables an additional variable  $\mathcal{F}_t$ , being 1 if a first-price rule was triggered within auction  $t$ , and 0 otherwise. Note that, by definition, this variable informs us on the equilibrium frequency with which the top two order statistics were within  $\Delta$  of one another, which is also connected to deviations from the SIA. We use this information to adjust our estimator as follows. First, for

<sup>20</sup> In all, we simulated the filter process for  $10^{10} \times 98$  separate auctions ( $10^{10}$  simulations for each  $N \in \{3, \dots, 100\}$ ). Computation was performed in parallel using a cluster of 150 MATLAB workers for 310 hours. We occasionally reset the seed so as to avoid surpassing the periodicity of the random number generator.

<sup>21</sup> Note that it is the value of  $\Delta$  itself that directly controls the degree of deviation, with no other indirect effects since monotone bidding strategies do not by themselves violate the SIA.

each value  $k_l$  contained in the vector  $\tilde{\mathbf{n}}$ , compute

$$P_{k_l} \equiv \frac{\sum_{t=1}^T \mathbf{1}(\mathcal{F}_t = 1 \cap \tilde{n}_t = k_l)}{\sum_{t=1}^T \mathbf{1}(\tilde{n}_t = k_l)}, \quad l = 1, \dots, L,$$

which estimates the conditional Bernoulli probability that a first-price rule was triggered, given  $\tilde{N} = k_l$ . Second, we compute an adjusted empirical mass function for  $\tilde{N}$ , call it  $\hat{\rho}_N^{\text{adj}}$ , by resampling from the data with replacement  $S > T$  times, and building an adjusted sample  $\{\tilde{n}_s^{\text{adj}}\}_{s=1}^S$  where whenever  $\tilde{n}_t = k_l$  is sampled in the  $s$ th draw, we assign a value

$$\tilde{n}_s^{\text{adj}} = \begin{cases} k_l + 1, & \text{with probability } P_{k_l}, \\ k_l, & \text{otherwise.} \end{cases} \quad (16)$$

The adjusted empirical mass function then becomes  $\hat{\rho}_N^{\text{adj}}(k_l) = \sum_{s=1}^S \mathbf{1}(\tilde{n}_s^{\text{adj}} = k_l)/S$  for each  $k_l \in \{\tilde{\mathbf{n}} \cup (k_L + 1)\}$ . Note that if  $P_{k_l} = 0 \forall k_l$ , then the adjusted mass function will be the same as the original for large  $S$ . Third, rather than using the raw empirical mass function  $\hat{\rho}_N$  in the objective function (14), we substitute  $\hat{\rho}_N^{\text{adj}}$  instead. For practical purposes, since the resampling step happens only once during runtime, it is not terribly costly to choose  $S$  fairly large. For our implementation, we chose  $S = 10^7$ , which means that the simulation error will have little or no effect on the first four digits of  $\hat{\rho}_N^{\text{adj}}$ .

Intuitively, this adjustment recognizes that the raw empirical distribution of  $\tilde{N}$  is stochastically dominated by the one we would observe if the SIA were never violated, and it tends to push the former toward the latter. However, it only partially offsets the overelimination of bidders in reality, since the conditional Bernoulli probabilities  $P_{k_l}$  only inform us on the tendency for the top order statistics to be close together. Therefore, we also propose and execute a robustness check (see Figure A.6 in the Supplemental Appendix) so as to assess the magnitude of the remaining problem. Briefly, we generated a sample of data from our point estimates in which for each simulated auction we knew the true  $n$ , valuations, and equilibrium bids for all  $n$  potential bidders, as well as the  $\tilde{n}$  that would result under the SIA and the  $\tilde{n}^{\text{raw}}$ , which accounts for the overelimination due to  $\Delta$ . We then performed estimation using three hypothetical scenarios: (i) with an ideal data set in which the true  $n$  was known for each auction and can therefore be directly estimated (for a baseline comparison); (ii) an estimator that maintains the SIA, even though the generated  $\tilde{n}^{\text{raw}}$  data are subject to overelimination; (iii) an estimator that employs the same set of generated  $\tilde{n}^{\text{raw}}$  data but also incorporates a corresponding simulated  $\mathcal{F}$  variable to perform our proposed correction. We also provide, for comparison, a plot of the private value CDF that results from misspecifying EAs as simple, second-price auctions. As is clear from Figure A.6, the overall bias in our baseline estimator of  $F_V$  is small relative to the effect of pricing-rule misspecification. Moreover, our proposed adjustment eliminates nearly all of the bias due to deviations from the SIA. We take this as evidence that our estimator does not unduly oversimplify the actual EA price dynamics, and achieves an appropriate balance of statistical accuracy and computational tractability.

4.5.2 *Second stage: B-spline estimator of  $F_V$*  Our identification argument above states that knowledge of  $G_M$  is sufficient to recover the private value distribution. A seemingly natural way to proceed then would be to estimate  $G_B(b)$  directly from the observables and then reverse-engineer  $F_V$ . Implementing this approach can be difficult, as  $\hat{G}_B$  must also ensure that the mappings implied by equation (7) and/or (8) are monotone so as to be consistent with equilibrium bidding. These requirements rule out kernel density estimators in favor of sieve estimation where a finite-dimensional parametric restriction is imposed and then gradually relaxed as the sample size increases. Even then, parameterizing  $G_B$  and then optimizing some empirical criterion function subject to the constraints mentioned above, along with enforcing monotonicity and boundary conditions of  $G_B$  itself, poses a formidable numerical challenge. The resulting constraint set becomes very complicated and highly nonconvex, making it difficult to compute admissible initial guesses and converge to global optima afterward.

We propose an alternative approach whereby we directly parameterize the private value distribution  $F_V$  as a flexible B-spline function. Still, since  $G_B(b) = F_V[\xi(b)]$ , building and optimizing an empirical criterion function of the observables—order statistics of bids—requires finding the solution to a differential equation based on  $F_V$  given in (7) (or, equivalently, (8)). To solve this piece of the puzzle, we employ the *Galerkin method*, which is commonly used to solve differential equations numerically in physical sciences applications.<sup>22</sup> This approach involves parameterizing the inverse bid function  $\xi$  as a B-spline as well, and afterward enforcing its adherence to the conditions of the boundary value problem defined by the equilibrium first-order conditions (FOCs). This is done by defining a grid of points on the domain (where the number of grid points is at least as large as the number of free parameters in the B-spline function), and then augmenting the estimator objective function with extra terms that penalize it for deviations from equilibrium. Our approach of augmenting an extremum estimator with the Galerkin method has the added benefit of being relatively inexpensive to compute: rather than repeatedly updating parameter values for  $F_V$  and then solving a differential equation in sequence, modeling  $\xi$  as a B-spline allows us to fit  $F_V$  to the data while *simultaneously* adjusting  $\xi$  to conform to the equilibrium conditions required by theory. We explain our approach concretely below, but first a brief word on B-splines is in order.

B-splines have many attractive properties that are well adapted to our application. First, they behave identically to piecewise splines, so by the Stone–Weierstrass theorem they can fit a broad class of nonparametric curves to arbitrary precision, given a fine enough partition of the domain. They also mimic attractive properties of piecewise splines: being locally low-dimensional they are numerically stable, and adjusting parameters to improve model fit at one point will have little or no effect on behavior of the B-spline function at points outside of a relatively small neighborhood. This is in contrast to global polynomials, where adjusting parameters to improve fit at one point may have drastic consequences for the behavior of the function at points far away.

<sup>22</sup>See Zienkiewicz and Morgan (2006) for details on Galerkin method approximation. Hubbard and Paarsch (2014) discuss and compare various ways in which researchers have solved the differential equation(s) characterizing equilibrium behavior at auctions.

On the other hand, like global polynomials, B-spline basis functions are globally defined—in fact, the name “B-spline” is short for *basis spline* because of this—making their functional values, derivatives, and integrals less cumbersome to compute than piecewise splines. In addition, B-splines are more adaptable in applications where the researcher may have a priori information about regions of the domain where the function is likely to display a high degree of curvature or complexity. For global polynomials, if the researcher wishes to infuse extra flexibility at a single point, she must increase flexibility on the *entire* functional domain. In contrast, B-splines allow for surgical targeting of an arbitrary degree of curvature within a small neighborhood of a particular point, up to and *including* infinite curvature at that point if kinks or discontinuities are known to occur. This property will be particularly useful in dealing with transition points where the bid increment discretely shifts and the bid function must therefore adjust itself to account for the abrupt change in the strategic environment just above  $b^*$ . A brief primer on construction of our B-spline functions is included in the Supplemental Appendix.

To outline our estimator, let  $\mathbf{j}_J = \{j_1 < j_2 < \dots < j_J < j_{J+1}\}$  denote a set of unique “knots” that partition the private value support into  $J$  subintervals, with  $j_1 = \underline{v}$  and  $j_{J+1} = \bar{v}$ , and let  $\mathbf{k}_K = \{k_1 < k_2 < \dots < k_K < k_{K+1}\}$  denote a partition of the bid support into  $K$  subintervals, with  $k_1 = \underline{b} = \underline{v}$  and  $k_{K+1} = \bar{b}$ . These knot vectors uniquely define a set of  $(J + 3)$  and  $(K + 3)$  fourth-order (cubic) B-spline basis functions  $\{\mathcal{V}_j(\cdot)\}_{j=1}^{J+3}$  and  $\{\mathcal{B}_k(\cdot)\}_{k=1}^{K+3}$ , where  $\mathcal{V}_j : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}$  for each  $j$  and  $\mathcal{B}_k : [\underline{b}, \bar{b}] \rightarrow \mathbb{R}$  for each  $k$ .

The basis functions are uniquely defined from their respective knot vector through the Cox–de Boor recursion relation formula with concurrent boundary knots (see the Supplemental Appendix for details). The resulting basis behaves the same as a set of piecewise cubic splines that are constrained to be  $\mathcal{C}^2$  at the endpoints of the subintervals defined by the knot vector. Each basis function is  $\mathcal{C}^2$  everywhere on the global domain, with the  $\mathcal{C}^2$  conditions at the knots being built into the recursion formula. Moreover, each one is nonzero on at most four of the  $K$  subintervals (though some are nonzero on fewer than four), and exactly four of the cubic basis functions are nonzero on each of the  $K$  subintervals, which explains the name “fourth-order B-spline,” and also why there are  $(K + 3)$  basis functions in total. Figure 4 depicts a basis on the interval  $[-1, 1]$  partitioned by a uniform knot vector with  $K = 3$ .

Letting  $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_{J+3}\} \subset \mathbb{R}^{J+3}$  and  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_{K+3}\} \subset \mathbb{R}^{K+3}$  denote sets of weights, we can now parameterize  $F_V$  and  $\xi$  as

$$\hat{F}_V(v; \boldsymbol{\mu}) = \sum_{j=1}^{J+3} \mu_j \mathcal{V}_j(b),$$

$$\hat{\xi}(b; \boldsymbol{\alpha}) = \sum_{k=1}^{K+3} \alpha_k \mathcal{B}_k(b).$$

Given this parameterization, the bid distribution is  $\hat{G}_B(b; \boldsymbol{\mu}, \boldsymbol{\alpha}) = \hat{F}_V(\xi(b; \boldsymbol{\alpha}); \boldsymbol{\mu})$ , and, in a slight change of notation, we can redefine the model-generated distribution of the

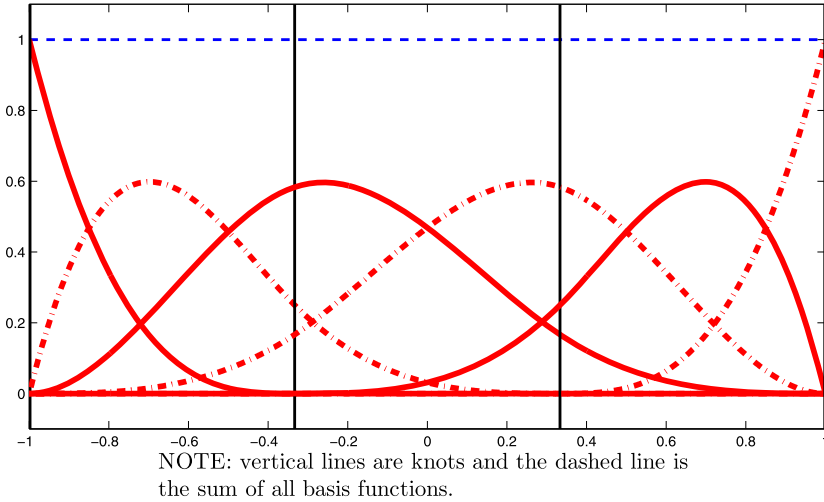


FIGURE 4. Fourth-order (cubic) B-spline basis functions.

highest losing bid as

$$\begin{aligned}
 H(b; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) &= \sum_{n=2}^{\bar{N}} \rho_N(n|N \geq 2; \hat{\boldsymbol{\lambda}}) \\
 &\times (\hat{G}_B(b; \boldsymbol{\mu}, \boldsymbol{\alpha})^n + n\hat{G}_B(b; \boldsymbol{\mu}, \boldsymbol{\alpha})^{n-1}[1 - \hat{G}_B(b; \boldsymbol{\mu}, \boldsymbol{\alpha})]).
 \end{aligned}
 \tag{17}$$

This function can now be directly compared to its empirical analog,  $\hat{H}(b) \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{1}(y_t \leq b)$ , so as to build a criterion function that will form the basis of an estimator. While optimizing the empirical criterion function, we must also impose the equilibrium conditions to ensure that  $\hat{F}_V$  and  $\hat{\xi}$  will be jointly consistent with the theory model. In other words, our final estimate for  $\hat{\xi}$  must constitute a valid solution to the boundary value problem in (7). To accomplish this, we specify a uniform grid of checkpoints  $\{b_l\}_{l=1}^L \subset [b, \bar{b}]$ , with  $L \geq K + 3$ , and we introduce four additional parameters,  $\varepsilon_0$ ,  $\varepsilon_1$ ,  $\varepsilon_2$ , and  $\varepsilon_3$ , and a residual function based on equations (12) and (8):

$$\begin{aligned}
 R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) &= (\hat{\xi}(b_l; \boldsymbol{\alpha}) - b_l)\hat{g}_M(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) - \hat{G}_M(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) \\
 &+ \hat{G}_M(\tau(b_l); \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}),
 \end{aligned}
 \tag{18}$$

where  $\hat{G}_M(b; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) = \sum_{n=2}^{\bar{N}} \rho_N(n|N \geq 2; \hat{\boldsymbol{\lambda}})\hat{G}_B(b; \boldsymbol{\mu}, \boldsymbol{\alpha})^{(n-1)}$ . With that, we can now define a constrained NLS estimator as

$$\begin{aligned}
 (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}}) &= \underset{(\boldsymbol{\mu}, \boldsymbol{\alpha}) \in \mathbb{R}^{J+K+6} \times \mathbb{R}_+^3}{\operatorname{argmin}} \left\{ \sum_{t=1}^T [H(x_t; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) - \hat{H}(x_t)]^2 \right. \\
 &\left. + P_0\varepsilon_0 + P_1\varepsilon_1 + P_2\varepsilon_2 + P_3\varepsilon_3 \right\}
 \end{aligned}$$

subject to

$$\begin{aligned}
 \hat{F}_V(\underline{v}; \boldsymbol{\mu}) &= 0, & \hat{F}_V(\bar{v}; \boldsymbol{\mu}) &= 1, \\
 \hat{f}_V(v; \boldsymbol{\mu}) &> 0, & v &\in [\underline{v}, \bar{v}], \\
 \hat{\xi}(\underline{b}; \boldsymbol{\alpha}) &= \underline{v}, \\
 \hat{\xi}'(b; \boldsymbol{\alpha}) &> 0, & b &\in [\underline{v}, \bar{b}], \\
 \mathcal{N}[\{R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})\}_{l=1}^L] &\leq \varepsilon_0, & b_l &\leq \underline{v} + \Delta_1, \\
 \mathcal{N}[\{R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})\}_{l=1}^L] &\leq \varepsilon_1, & \underline{v} + \Delta_1 &< b_l < b^* + \Delta_1, \\
 \mathcal{N}[\{R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})\}_{l=1}^L] &\leq \varepsilon_2, & b^* + \Delta_1 &\leq b_l < b^* + \Delta_2, \\
 \mathcal{N}[\{R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})\}_{l=1}^L] &\leq \varepsilon_3, & b^* + \Delta_2 &\leq b_l, \\
 d(\hat{\boldsymbol{\mu}}, \varepsilon_0 \geq 0, & \varepsilon_1 \geq 0, & \varepsilon_2 \geq 0, & \varepsilon_3 \geq 0,
 \end{aligned} \tag{19}$$

where  $\mathcal{N}[\cdot]$  is a norm function such as the  $L_\infty$  norm (the sup-norm),  $\sup_{l \in \{1, 2, \dots, L\}} |R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})|$ , or the  $L_2$  norm,  $(\sum_{l=1}^L R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}})^2)^{1/2}$ , and  $(P_0, P_1, P_2, P_3)$  are prespecified penalty parameters.<sup>23</sup>

Intuitively, the residual function equals zero when the private-value distribution  $\hat{F}_V$  together with the numerical bid function  $\hat{\xi}$  exactly conforms to the first-order conditions of a bidder's decision problem, and the constraint that  $\hat{\xi}(\underline{b}; \boldsymbol{\alpha}) = \underline{v}$  enforces the boundary condition. Thus, the vector  $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3)$  controls the degree of numerical error in the approximated solution to the piecewise differential equation. Moreover, by fitting the parameterized distribution  $F_V$  to the data while simultaneously searching over parameter values for  $\hat{\xi}$  that conform to the first-order conditions, given  $F_V$ , we avoid the computational cost involved in repeatedly solving differential equations *each time* the objective function is evaluated, which can easily number into the thousands. Instead we essentially only solve the equilibrium *once*, and our NLS estimator with Galerkin ordinary differential equation (ODE) solution is in the spirit of the MPEC (mathematical programming with equilibrium constraints) method pioneered by [Su and Judd \(2012\)](#): we choose the parameters of the parent distribution  $G_B$  so that the model-generated order statistic quantiles match the empirical order statistic quantiles as closely as possible, while penalizing the objective function for deviations from the equilibrium conditions in (12) and (8). Although we tailor our estimator here to the equilibrium of electronic auctions, our approach may be applicable to broader structural contexts outside of auctions where the distribution of a latent, structural random variable is linked to an observable distribution through an equilibrium mapping that must otherwise be repeatedly solved during estimation runtime.

<sup>23</sup>Note that the sup-norm variant of the estimator may be implemented without violating differentiability of the Lagrangean objective function: one can replace the absolute value operator with a two-part, nonlinear constraint where  $R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) < \varepsilon_i$  and  $-R(b_l; \boldsymbol{\mu}, \boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}) < \varepsilon_i$  must both be satisfied for each relevant  $(l, i)$  pair, for  $l = 1, 2, \dots, L$  and  $i = 0, 1, 2, 3$ . In our experience we have found the  $L^\infty$  (sup) norm to be much more stable and accurate than the  $L^2$  (Euclidean) norm. See Figure A.8 in the Supplement for a comparison.



The main difference between our proposal here and the way Galerkin methods are typically implemented stems from how the equilibrium conditions are enforced. The most common implementation is to define the residuals  $R(b_l)$  on a grid of checkpoints  $\{b_1, \dots, b_L\}$  with  $L = K + 2$ . This, plus a boundary condition, produces a square system of  $K + 3$  equations in  $K + 3$  unknowns, which can be solved exactly on the grid of checkpoints (see, e.g., [Hulme \(1972\)](#) for further discussion). However, following this standard approach during estimator runtime would once again necessitate solving the differential equation once for every objective function evaluation, which can number well into the thousands. This is why we opt for the objective function penalization approach described above (with overfitting, or  $L \geq K + 2$ ), which allows for the parameters of  $\hat{F}_V$  and  $\hat{\xi}$  to move independently during runtime. This essentially means that the differential equation need only be solved once. As a check on output though, having obtained the point estimate  $\hat{F}_V(v; \hat{\mu})$ , the researcher can easily verify the solution for  $\hat{\xi}(b; \hat{\alpha})$  by resolving the equilibrium boundary value problem in (12) and (8) using the standard Galerkin solution method based on a square system of nonlinear equations, *holding  $\hat{\mu}$  fixed*, and even with a finer knot vector than  $\mathbf{j}_J$ , if desired.

**4.5.3 Enforcing boundary conditions and shape restrictions** Before moving on, a final attractive property of B-splines is their ability to facilitate known shape restrictions on the latent function they parameterize. Note from [Figure 4](#) that, by construction, the extremal basis functions are the only ones to attain nonzero values at the boundaries, and they both equal 1 exactly at their respective endpoints. This makes enforcement of boundary conditions very easy:  $\hat{F}_V(\underline{v}; \hat{\mu}) = 0 \Leftrightarrow \mu_1 = 0$ ,  $\hat{F}_V(\bar{v}; \hat{\mu}) = 1 \Leftrightarrow \mu_{J+3} = 1$ , and  $\hat{\xi}(b; \hat{\alpha}) = \underline{v} \Leftrightarrow \alpha_1 = \underline{v}$ , which also reduces the number of free parameters by three. Finally, B-splines also have some other remarkable properties that make enforcement of monotonicity conditions (and hence, concavity/convexity also, if needed) quite simple: it turns out that for any B-spline function  $\hat{F}_V(v; \mu) = \sum_{j=1}^{J+3} \mu_j \mathcal{V}_j(b)$ , we have that  $\hat{F}'_V(v) \geq (>)0 \Leftrightarrow \alpha_{j+1} \geq (>)\alpha_j$ , for each  $j = 1, \dots, J+2$  (see [de Boor \(2001, p. 115\)](#)). Therefore, enforcing the shape and boundary conditions required by the theory reduces to adding a set of equality constraints and linear inequality constraints directly on the parameter values themselves. This presents another substantial numerical advantage over global polynomials, where enforcing shape conditions can require imposition of complicated, nonlinear constraints on the functional values, which may lead to nonconvex constraint sets and increase the tendency of a solver to get stuck at local optima.

**4.5.4 Accommodating transition points** Given the above derivations, we know that the primitives of the strategic environment remain fixed for bids below  $b^* + \Delta_1$  and above  $b^* + \Delta_2$ , but between these two points there is an abrupt (though continuous) increase in the probability that a first-price rule will be triggered, which means that the degree of demand shading is also expected to abruptly increase on this segment of the domain. Fortunately, the Cox–de Boor recursion formula (with concurrent boundary knots; see the Supplemental Appendix), which we used to compute our B-spline basis functions, was developed with such a contingency in mind. Specifically, it allows for arbitrary curvature at a particular point by adding additional knots that are closely spaced together.

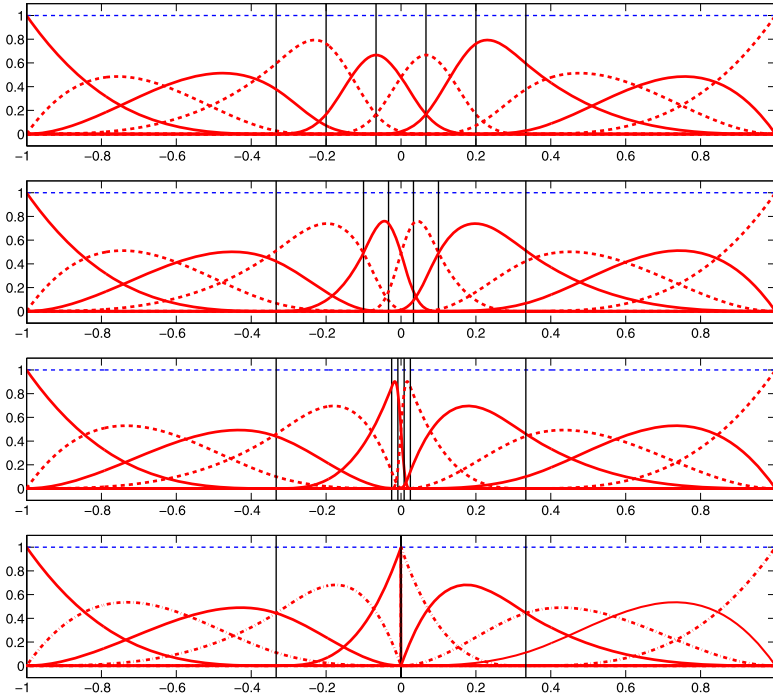


FIGURE 5. Fourth-degree (cubic) B-spline basis functions with additional knots.

Recall from the above discussion that we defined knot vector  $\mathbf{k}_K = \{k_1 < \dots < k_{K+1}\}$  in bid space, which in turn uniquely defines  $(K + 3)$ ,  $C^2$  basis functions. For simplicity of discussion, let  $K > 2$  and suppose that  $(b^* + \Delta_1, b^* + \Delta_2) \subset [k_2, k_3]$ . We now modify the knot vector by inserting four additional knots between  $k_2$  and  $k_3$  so that we now get a knot vector with  $(K + 5)$  elements:

$$\mathbf{k}'_K = \{k_1 < k_2 < \kappa_1^* \leq \kappa_2^* \leq \kappa_3^* \leq \kappa_4^* < k_3 < \dots < k_{K+1}\}.$$

This insertion will have several effects on the recursion formula, which we briefly outline here; the interested reader is once again directed to the Supplemental Appendix for additional details. First, we increase the number of B-spline basis functions by four; in particular, exactly eight basis functions will be nonzero on the subinterval  $[k_2, k_3]$  now. Second, each set of basis functions—uniquely determined by a particular configuration of the knot vector—places an implicit bound on the second derivative of the B-spline function, but this bound is relaxed within a certain neighborhood as the additional knots are spaced more closely together. Figure 5 illustrates this phenomenon: each of the four panes depicts the knot vector from Figure 4 with four additional knots inserted between  $k_2$  and  $k_3$  and placed at varying distances from one another. As these knots get closer together, the amount of curvature displayed nearby in the four new basis functions becomes ever more extreme, and in the limit as the four knots are stacked on top of one another, two of the basis functions become kinked and two are even discontinuous. Although the region with increased flexibility collapses as the bound on the

second derivative is relaxed, the researcher may spread the increased flexibility over a wider area by including more knots, if desired. This is useful because usually the conditions that require extra flexibility within a particular neighborhood are not meant to profoundly impact the behavior of the B-spline function at points far away. Once again, the intuitive properties of B-splines come to bear; recall that each basis function is nonzero on at most four subintervals defined by the knot vector. Suppose then that the researcher wishes to locate four new knots within a small neighborhood of the middle subinterval, so that  $k_2 < \kappa_1^* \leq \kappa_2^* \leq \kappa_3^* \leq \kappa_4^* < k_3$ , but she does not wish the influence of the new corresponding basis functions to extend outside of  $[k_2, k_3]$  either. In this case, inclusion of four additional knots in  $[k_2, \kappa_1^*)$  and four additional knots in  $(\kappa_4^*, k_3]$  (for a total of 12 new knots) would suffice.

A final change induced by the inclusion is that the adjacent basis functions also change shape as well, relative to Figure 4. This highlights an important difference between B-splines and global polynomials. With the latter, inclusion of new terms leads to a nested model structure, wherein new basis functions are simply being added to the preexisting set. For B-splines, however, as we pack the functional domain with an increasing number  $K$  of knots, we get a sequence of *nonnested* models for the underlying function we wish to approximate, since adding one additional knot will cause the preexisting basis functions with nonzero domain nearby that knot to change form. Numerically, this is of little consequence, but econometrically, this fact will have bearing on interpretation of varying model fit as  $K \rightarrow \infty$ .

## 5. EMPIRICAL APPLICATION

### 5.1 Data

We now consider a sample of laptop auctions collected between April and June of 2008 from the eBay website, which provides extensive information on item characteristics and bid histories. This application highlights various challenges present in real-world data, including those addressed in the previous section. The largest seller during the data collection period was a firm by the name of CSR Technologies (henceforth CSRT), which purchased large quantities of second-hand laptops for resale on eBay. CSRT's product line was mostly made up of Dell Latitude laptops, which come in several different configurations. Because laptop and auction characteristics can be important determinants of the price a computer will receive, we attempt to homogenize our sample as much as possible.

CSRT's most common laptop configuration, comprising 733 total auction listings, included an Intel Pentium 4M processor with a clock speed of 1.4 GHz, 512 MB of random access memory (RAM), 30-GB hard drive, a DVD-ROM (digital versatile disc-read-only memory) optical drive, a 14.1-inch screen, and with the Windows XP Professional operating system installed. All laptops in this sample were described by the seller as either "refurbished" or "used," and all corresponding auctions lasted for 24 hours. We restrict our sample to only laptops sold by CSRT, which ensures constant seller reputation, exchange and upgrade policy, flat shipping rate (\$36), auction setup, and so on. Moreover, CSRT used a template format for the display of each of its auction listings on

TABLE 3. Descriptive statistics.

Variable	Mean	Median	St. dev.	Min	Max	# Obs.
<i>Time remaining (minutes)</i>						
Winning bid submission	24.04	1.43	97.50	0.00	1392.65	733
High loser bid submission	24.29	3.10	61.76	0.00	633.40	733
<i>Observed participation</i>						
$\tilde{N}$ (serious bidders only)	4.69	4	1.70	2	11	733
<i>Monetary outcomes</i>						
Sale price (w/o shipping)	\$299.63	\$300	\$31.31	\$202.50	\$405	733
Highest losing bid	\$295.42	\$299	\$31.20	\$200	\$400	733
First-price frequency	22.24%	–	–	–	–	–

eBay, making appearance during the auction uniform as well. After these restrictions, the only auction characteristics that vary are the bidding data themselves. This unique and homogeneous sample of auctions will allow us to abstract away from further complications (such as unobserved heterogeneity) as we develop an empirical methodology to tackle the already formidable challenges inherent in ideal eBay data.

## 5.2 Sample paring

For each auction, we have detailed bidding information, including the timing and amount of each bid submitted to the eBay server, as well as the identity of the bidder who submitted it. Looking into the bid history allows us to calculate the number of observed bid submissions and the number of observed unique bidders. The first empirical challenge we encounter is that bidders may play different, potentially nonequilibrium actions at various points in the auction; for example, submitting low, cheap-talk bids early on and then later bidding based on best-response calculations resembling those in a sealed-bid auction. Empirically, a significant fraction of observed bid amounts, particularly those submitted early on in the life of the auction, fall too far below realistic transaction prices to be taken seriously.

On the other hand, in the vast majority of cases, the top two bids arrive within the final 30 minutes of the auction. Table 3 shows that the average time to end when the winning bid arrives is 24.04 minutes, and the median time to end is 1.43 minutes. For the highest losing bid, the mean and median time remaining are 24.29 and 3.10 minutes, respectively. Previous empirical work on eBay has established these phenomena as empirical regularities and our data are no different.

**5.2.1 Intra-auction dynamics** This discussion hints at a need to deal with the issue of intra-auction dynamics on eBay in some way. We adopt an approach similar to that of [Bajari and Hortaçsu \(2003\)](#) by partitioning the auction runtime into two stages.<sup>24</sup> Taking

<sup>24</sup>[Bajari and Hortaçsu \(2003\)](#) were motivated by the fact that eBay does not fit within [Milgrom and Weber's \(1982\)](#) "open-exit" ascending auction format where bidders' exit decisions are observable to their competitors. On the contrary, eBay bidders may rejoin the auction at any time after an initial proxy bid is surpassed.

the total auction time to be  $T$ , we treat the first stage as an open-exit ascending auction played until  $T - \varepsilon$ ; the final (terminal) stage, of length  $\varepsilon < T$ , is treated as a sealed-bid auction.<sup>25</sup> During the first stage of the auction bidders submit cheap-talk bids that convey little, if any, information on the likely sale price that will result from the auction.

As in [Bajari and Hortaçsu \(2003\)](#), bidders formulate a strategic sealed bid for submission during the terminal stage. We assume that this strategic bid ignores the outcome of the initial cheap-talk stage and the terminal stage is like a sealed-bid auction in the sense that bidders do not reoptimize their strategic bids in response to observed price-path dynamics during the terminal stage. Bidders may, however, submit these bids directly to the server to take advantage of eBay's automated proxy bidding capability, or they may choose to incrementally increase their bid submissions up to the level of their strategic bid on their own. In other words, we assume bidders formulate strategic bids based on the distribution of private values and their expectation of  $N$  at the beginning of the terminal period, and afterward they stick to their planned strategic bid through the end of the auction. Finally, consistent with the previous section, price-path dynamics within the terminal period are assumed to be simple in the sense that ordering of bidders' submission times is random rather than coordinated.

**5.2.2 *Serious bidders and strategic bids*** This partitioning of the auction into an initial cheap-talk period followed by a terminal sealed-bid auction leads to the following definition: a *serious bid* is one that affects the price path within the terminal period; likewise, a *serious bidder* is one who is observed to submit at least one serious bid. This distinction allows for the possibility that some observed bidders early on in the life of the auction were merely "fishing for a steal" or casually dabbling, rather than seriously vying to win a laptop like other bidders who remain active close to the end. Of course, the possibility always exists that some bidders who are determined to be nonserious by the above criterion had serious intent to compete for the item, but were priced out before submitting a serious bid. This, however, is just part of the problem that our proposed estimator for  $\lambda$  based on our explicit model for the filter process is meant to solve: recall that observed participation is merely treated as a lower bound on actual participation when identifying the arrival rate of bidders within an auction.

In our empirical application, we specify the terminal period as the last 30 minutes of an auction. During this period, we see an average of 4.69 observed serious bidders. [Figure 6](#) provides a justification for this choice. It shows two empirical distribution functions: one for time remaining when the highest losing bid is submitted, and one for time remaining across all serious bids. Note that the highest losing bid arrives with fewer than 30 minutes left in over 80% of the auctions in our sample. Note also that 20.98% of all serious bids are submitted *prior* to the final 30 minutes of the auction. This possibility

---

<sup>25</sup>Some other work within the related literature, including [Nekipelov \(2007\)](#), has attempted to develop an explicit empirical model for determining both levels *and* timing of bids within an auction. One challenge to such an undertaking is that fully formed equilibrium theories of bid timing within an auction are rare. Such an exercise would introduce considerable complexity and is beyond the scope of our current purposes. Thus, for tractability, we follow the standard approach of adopting assumptions that simplify intra-auction dynamics.

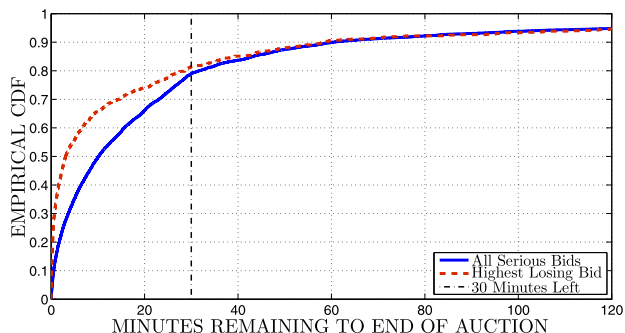


FIGURE 6. Empirical distributions: time remaining when bids are submitted.

is naturally built into our serious bidder criterion so as to avoid making too sharp a distinction between the two stages of the auction. Whenever there are at least two bidders over the life of an auction, the two highest bids during the cheap-talk phase are the ones that set the initial price for the final 30 minutes. Therefore, the individuals who submitted them will be logged as serious bidders, even if they are not observed to actively bid during the final 30 minutes.

The majority of serious bidders only submit a single proxy bid, but just under one-third of serious bidders are observed to raise their own bid levels incrementally with multiple submissions during the terminal stage. This discrepancy motivates our final data restriction: so as to be conservative we assume only that the highest losing submission (that is, the second-highest overall bid) is reflective of a strategic equilibrium bid consistent with the model from Section 4.1 above. Within the context of our simplified model of intra-auction dynamics, if some bidders log multiple bid submissions, we interpret such behavior as raising bid levels manually up to the point of their planned strategic bid (which we assume is not updated during auction runtime), rather than availing themselves of the automated proxy bidding system.<sup>26</sup> Thus, many submissions by serious bidders may still represent only lower bounds on their equilibrium strategy. In defense of our implicit trust in the highest losing bid, we appeal to the [Haile and Tamer \(2003\)](#) assumption that bidders will never allow an opponent to win at a price they would be willing to pay, where willingness to pay is interpreted here as a planned strategic bid rather than a private valuation. By this logic one can at least be confident that the highest losing submission is reflective of a full strategic bid, and it will always be available since it is always recorded by Nature as the filter process runs its course.<sup>27</sup>

<sup>26</sup>Our assumption that bidders choose not to update their planned strategic bids during auction runtime is based in the idea that electronic auctions exist in larger marketplaces where a loser today may return tomorrow to bid again. [Bodoh-Creed, Boehnke, and Hickman \(2016\)](#) find empirical evidence in support of this view; in Section 6.1 we provide a brief discussion of intra-auction dynamics.

<sup>27</sup>Nature also always records the attendance of the winner, but the eBay bidding system only reports exact amounts for bids that were surpassed by the next lowest bid plus  $\Delta$ . Thus, the winning bid itself is only observable to the econometrician for auctions where price is generated by a first-price rule, which occurs in roughly 22.24% of the auctions in our sample. In principle, inclusion of these bids may improve the statistical efficiency of the estimator, but it would come with added complexity and computational cost.

Of course, if the researcher is able to incorporate more bidding data per auction in a reliable way, then it would improve the statistical precision of resulting estimates.

In total, our various sample restrictions leave us with 733 observed highest loser bids (that is, all auctions in our data had at least two serious bidders).<sup>28</sup> Table 3 displays descriptive statistics on bid timing, observed participation, sale prices, and highest losing bids. With probability 0.2224 the sale price takes on a value of exactly the highest bid for our sample. Note, too, that bids and sale prices do not include the flat shipping fee of \$36.

**5.2.3 Practical matters** As a final practical matter, we had to, choose two knot vectors  $\mathbf{j}_J$  and  $\mathbf{k}_K$  for the B-spline basis functions  $\mathcal{V}_j, j = 1, \dots, J + 3$ , and  $\mathcal{B}_k, k = 1, \dots, K + 3$  to be well defined. The first challenge is to specify the endpoints. Since the sample extrema are superconvergent estimators of the support endpoints, and since theory requires that  $\beta(\underline{v}) = \underline{v}$ , we set  $j_1 = k_1 = \min_t \{y_t\}$  and  $k_{K+1} = \max_t \{y_t\}$ . Because the bid function does not touch the 45° line at the upper end, we can only bound  $\bar{v}$  from below and must, therefore, make a guess at an appropriate value. We chose  $j_{J+1} = \bar{v} + 2\Delta_2$  as an approximate value for the upper bound of the private-value support. Our point estimates, discussed below, suggest that this was a reasonable choice that does not seem to drive results in any meaningful way.

The next step is to select the number and placement of the knots. The primary concern in selecting  $K$  is to minimize numerical error in the approximate solution to the equilibrium differential equation. We chose a grid of 48 knots to be placed uniformly over the set  $[\underline{b}, b^* + \Delta_1 - 4\epsilon] \cup [b^* + \Delta_1 + 8\epsilon, \bar{b}]$ , with  $\epsilon = (\Delta_2 - \Delta_1)/4$ . We also inserted an intermediate knot grid  $\{b^* + \Delta_1 - 3\epsilon, b^* + \Delta_1 - 2\epsilon, \dots, b^* + \Delta_1 + 7\epsilon\}$ , with 12 additional subintervals, to provide extra flexibility for the transition region where the bid increment changes from  $\Delta_1$  to  $\Delta_2$ , but without projecting the influence of this region to points far away. This gives us  $K = 60$  total knots and 63 total basis functions for the inverse bid function approximant  $\hat{\xi}$ . The vector of points where equilibrium conditions were enforced was a uniform grid of 500 points on  $[\underline{b}, \bar{b}]$ . From experimentation, this configuration of knots and checkpoints seemed to deliver a reasonable trade-off between computational cost and numerical accuracy: larger values of either  $K$  or  $L$  do not improve model fit or numerical accuracy in any meaningful way.

As for the principal knot vector  $\mathbf{j}_J$ , the primary concern is goodness of fit to the distribution of the observables. One challenge is first to reduce the dimensionality of the decision problem in a data-driven way, if possible. Our proposal is to specify a uniform grid of  $J$  quantile ranks  $\{q_1, \dots, q_{J-1}, q_{J+1}\}$  spanning  $[0, 1]$ , and then to map them into bid space using the empirical quantile function  $\hat{H}^{-1}(q)$ . We then replace the uppermost knot  $j_{J+1}$  with a value of  $(\bar{v} + 2\Delta_2)$ , as mentioned above, so as to account for the fact that these knots will govern the behavior of  $\hat{F}_V$  in private-value space. Finally, because our

---

For simplicity sake, we ignore them here. Despite this additional data loss, the confidence bounds we get on our estimates are still remarkably tight.

<sup>28</sup>In addition, we used one final sample restriction so as to avoid numerical instability issues in the upper tail of the private-value distribution. The original data set contained 736 observations fitting the description above, but for three of these, the highest losing bid was 4 standard deviations or more above the mean. We drop these three auctions from the sample, leaving us with 733 total auctions.



highest loser bid distribution is skewed to the right with a long, curved upper tail, we insert an additional knot  $j_j = (j_{j-1} + j_{j+1})/2$  to provide some added flexibility in this region. The advantage of this approach is that it reduces the  $(J + 1)$ -dimensional decision of knot choice to a single decision of  $J$ , the number of knots, while letting the data determine the placement of the knots. On the other hand, this method concentrates knots more densely within higher quantiles of  $\hat{F}_V$ , since we are choosing knots at the quantiles of highest losing bids. However, this is a natural problem inherent in estimating a parent distribution from any data set consisting of order statistics: observations selected from the within-auction sample extremes will always be more informative of the higher quantiles of the parent distribution. Thus, in finite samples, knot choice involves the familiar bias–variance trade-off. We propose the above method because it allows the observables to determine knot location. Statistically optimal knot choice, while an interesting problem, is beyond the scope of the current exercise and is therefore left to future work.<sup>29</sup>

Finally, when enforcing the equilibrium conditions, we found the  $L_\infty$  norm most effective. The problem with the other  $L_p$  norms (with finite  $p$ ; for example, the  $L_2$  norm) is that they allow for a small number of drastic deviations from the FOCs, as long as the solution to  $\hat{\xi}$  is well behaved at most domain points. This can lead to poor performance of the solution method. Although both methods produce a spline estimate with the same overall trend, the  $L_\infty$  norm, by directly disciplining the worst-behaved segments of the spline, is able to achieve a better overall fit, while avoiding wild oscillations around the bidder optimality condition. As an illustration of this point, Figure A.8 in the Supplemental Appendix plots the relative approximation error of the B-spline approximation,  $R(b)/\hat{\xi}(b; \hat{\alpha})$  under the  $L_\infty$  and  $L_2$  norms. The plot shows that the approximation error under the  $L_\infty$  norm is several orders of magnitude smaller along much of the functional domain. Moreover, few remedies exist to improve this comparison without drastically increasing computational load; for example, the picture remains virtually unchanged if we double the number of domain checkpoints. Thus, we recommend the  $L_\infty$  norm for practical use. We chose values of  $P_0 = 10$  and  $P_1 = P_2 = P_3 = 200$  as this allowed for a good balance of least squares fit and small numerical error.

### 5.3 Empirical results

We first separately estimated two parametric models of the bidder arrival process—a generalized Poisson model indexed by  $(\lambda_1^{\text{GP}}, \lambda_2^{\text{GP}})$  and a standard Poisson model indexed by  $\lambda^{\text{P}}$  (where the second parameter in the generalized model is restricted to be zero). Under the generalized Poisson model we get both a higher mean, 11.569, and standard deviation, 6.832, of the random variable  $N$  as compared to the Poisson model with 10.023 and 3.166, respectively. Figure 7 depicts a comparison of the empirical distribution of  $\tilde{N}$  versus the two parametric models. Both fit the data fairly well, but the generalized Poisson model emerges as the clear winner: on the majority of the sample domain

<sup>29</sup>By experimenting with alternatives where knots are placed uniformly in  $V$  space instead, we found that the shape of the point estimate for  $\hat{f}_V$  changed somewhat near the lower end, but variance of the estimator increased significantly. The simple intuition is that with uniform knots in  $V$  space there are several model parameters whose values are being determined by sparse data near the lower extreme of the sample, which is likely leading to higher mean squared error.

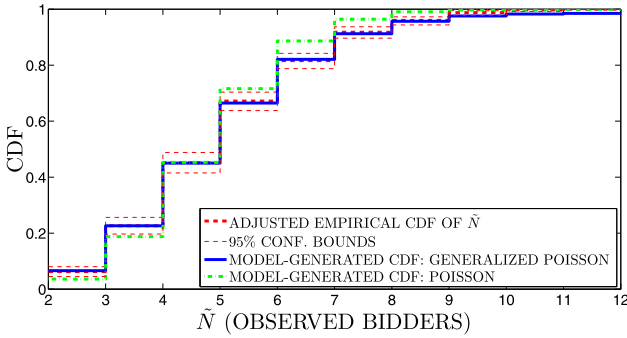
FIGURE 7. Model fit: empirical CDF of  $\tilde{N}$  versus parametric model-generated CDFs.

TABLE 4. Parameter estimates and standard errors.

$\hat{\lambda}_1$	$\hat{\lambda}_2$	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$	$\hat{\mu}_1 = 0$	$\hat{\mu}_{14} = 1$
5.760 (0.424)	0.502 (0.049)	200.00	255.08	267.86	278.18	287.19	294.99	$\hat{\mu}_2$	$\hat{\mu}_3$
		$j_7$	$j_8$	$j_9$	$j_{10}$	$j_{11}$	$j_{12}$	0.0291 (0.0513)	0.3443 (0.0547)
		301.87	309.10	318.90	337.78	373.89	410.00		
$\hat{\mu}_4$	$\hat{\mu}_5$	$\hat{\mu}_6$	$\hat{\mu}_7$	$\hat{\mu}_8$	$\hat{\mu}_9$	$\hat{\mu}_{10}$	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$
0.6345 (0.0201)	0.7300 (0.0152)	0.7775 (0.0126)	0.8206 (0.0108)	0.8621 (0.0079)	0.8994 (0.0064)	0.9374 (0.0052)	0.9593 (0.0049)	0.9897 (0.0050)	0.9989 (0.0016)
$\varepsilon_0$	$\varepsilon_1$	$\varepsilon_2$	$\varepsilon_3$	$SSR _{J=11}$					
$9.8 \times 10^{-6}$	$3.4 \times 10^{-6}$	$8.6 \times 10^{-5}$	$9.9 \times 10^{-5}$	0.1194					

the model-generated distribution of observables under the generalized Poisson model is within the nonparametric 95% confidence bounds of the adjusted empirical distribution of observed bidders. This implies that, although there still may be asymptotic gains from a more flexible functional form, the current sample size precludes further improvement without more data.

For the private-value distribution we settled on a value of  $J = 11$ , meaning 12 knots in  $V$  space and 14 total parameters, with 12 of them being free parameters after we enforce the boundary conditions on  $\hat{F}_V$ . This choice seemed to provide a high degree of flexibility for fitting patterns in the data, while producing a high degree of numerical accuracy. Table 4 displays our chosen knots  $\{j_1, \dots, j_{12}\}$ , parameter point estimates, and bootstrapped standard errors in parentheses. Figure 8 conveys how the model specifications  $J = 9, 10, \dots, 13$  all fit the data remarkably well. In selecting  $J = 11$  we conducted a bootstrap exercise to compare the mean integrated squared error across these specifications.<sup>30</sup> All of them implied virtually identical pointwise mean values for the CDF,

<sup>30</sup>A broader question concerns the optimal rate at which model complexity should increase as the sample size grows. Given our rule for knot placement (uniform in quantile rank space), could one construct a consistent estimator that includes a rule for choice of  $J \rightarrow \infty$  as a function of sample size  $T$ ? Given that the Stone–Weierstrass theorem applies to B-splines as it does to global polynomials, the answer is likely yes. However, it is beyond the scope of this work, as is the question of optimal rate of increase for  $J$ , and is left to future research.

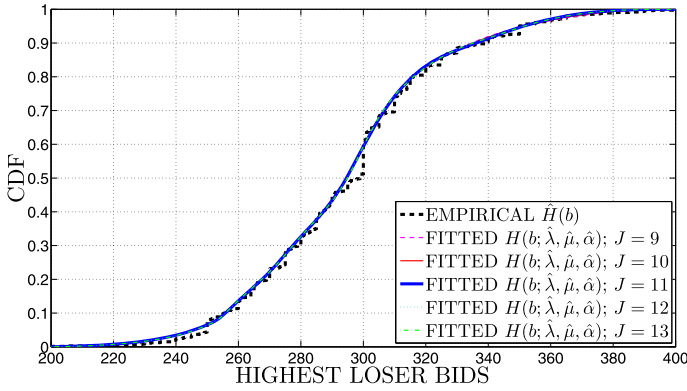


FIGURE 8. Model fit: empirical  $H(b)$  and estimated  $H(b)$  with  $J = 9$ –13.

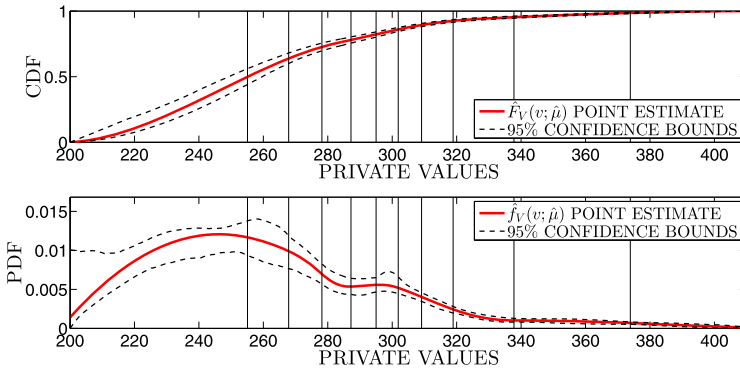


FIGURE 9. Estimated valuation distribution and density with knots.

meaning that there should be no meaningful difference in bias. Therefore, the integrated mean squared error comparison reduces to one of integrated variance, which is lowest for specification  $J = 11$ . Figure A.7 in the Supplemental Appendix also shows that the relative numerical errors under the different model specifications are all quite small and roughly the same.

Figure 8 illustrates the fit between the empirical CDF  $\hat{H}(b)$  (thick dashed line), and the model-generated version under varying choices for  $J$ . The thick solid line corresponds to our preferred specification with  $J = 11$ . Figures 9 and 10 depict the estimated private-value distribution and inverse bid function with bootstrapped 95% confidence bounds.<sup>31</sup> In each figure, the thin vertical lines indicate knots. The inverse bid function plot is zoomed in below the 90th percentile of the private-value distribution so that the features of the function can be seen more clearly. Note the abrupt transition in the degree of demand shading above the point  $b^* = \$250$ . The confidence bounds on the inverse bid function are remarkably tight because within the relevant ranges for  $\Delta$ , private-

<sup>31</sup>To compute our bounds we estimated the model parameters  $\lambda$ ,  $\mu$ , and  $\alpha$  on 1000 bootstrapped samples, holding the knot vectors  $\mathbf{k}_K$  and  $\mathbf{j}_J$  fixed.

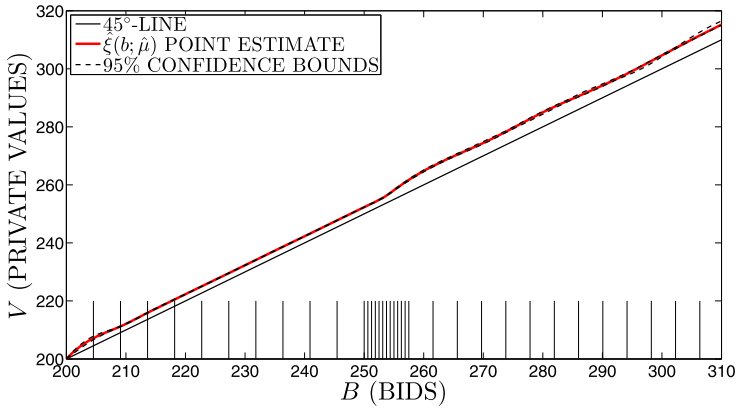


FIGURE 10. Estimated equilibrium inverse bidding function.

value distributions that significantly differ tend to still largely agree on the appropriate degree of demand shading. For completeness, Figure A.7 in the Supplemental Appendix presents the relative approximation error, or the ratio of the residual function  $R(b)$  to the units of  $\hat{\xi}(b; \hat{\alpha})$ . As suggested in Section 4.5.2 above, we also compute a final check on our solution for  $\hat{\xi}(b; \hat{\alpha})$  by holding the structural parameters  $\hat{\lambda}$  and  $\hat{\mu}$  fixed, and computing a standard Galerkin approximant, call it  $\tilde{\xi}(b; \tilde{\alpha})$ , by solving a square system of nonlinear equations on an appropriately chosen grid of checkpoints. This we do with a knot vector  $\tilde{\mathbf{j}}_J$ , where  $\tilde{J} = 5J$  for improved accuracy (under the fixed-point estimates for distributional parameters), and we then compare the two solutions; see Figure A.9 in the Supplemental Appendix. The knot vector we selected followed the same rule as the one used in estimation, just with more knots; the grid of checkpoints included the midpoint of each subinterval, the endpoints of the knot vector, and a final point chosen halfway between the lowermost knot and the lowermost midpoint. The sup-norm distance between our runtime ODE solution  $\hat{\xi}(b; \hat{\alpha})$  and the ex post, standard Galerkin implementation  $\tilde{\xi}(b; \tilde{\alpha})$  (using a square system of equations) turns out to be 20 cents, but over the majority of the domain, the two differ by less than 5 cents. Figure A.9 displays the runtime solution with a thick, dark line and the ex post solution with a thin, light line so that the two can be visually distinguishable. This comparison indicates that our point estimates for the private-value distribution are based on a differential equation solution with an acceptable level of numerical error.

It is interesting to note that both the point estimate  $\hat{\xi}(b; \hat{\alpha})$  and the lower confidence bound lay significantly above the 45° line, which represents the hypothetical inverse bid function under a second-price equilibrium. To put the picture into context, the upper panel of Table 5 summarizes the degree of estimated demand shading for bids occurring at various quantiles of the observable distribution  $H(b)$ . Within the 90–10 range, the difference between the estimated EA bid function and the second-price bid function ranges from \$3.41 to \$5.29. Note that the information on bid shading in the table implicitly represents the magnitude of the misspecification problem from ignoring the hybrid EA pricing rule as well: under a second-price assumption, the econometrician would estimate the parent distribution of bids using our model of the filter process, and then

TABLE 5. Absolute and relative differences between private values and bids.

	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
<i>High loser bid:</i>	\$255.50	\$275.00	\$299.00	\$311.86	\$339.00
<i>Bid shading: <math>V - \beta(V)</math></i>	\$3.41	\$4.61	\$4.54	\$5.29	\$4.61
	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
<i>Buyer Information rents</i>					
$\mathcal{I} = V_{(1:N)} - \min\{B_{(2:N)} + \Delta, B_{(1:N)}\}$	\$5.01	\$9.19	\$22.51	\$44.46	\$69.36

incorrectly interpret  $G_B$  as being the same as  $F_V$  since bids and private values are the same under a second-price rule. Our approach is to incorporate equations (12) and (8) into the estimation step as a correction factor. The numbers from Tables 3 and 5 indicate that failing to do so would result in misspecifying point estimates of private valuations by between 11% and 17% of a standard deviation of the observable distribution  $H$ . It is also worth mentioning that if we compute an alternative B-spline estimator, call it  $\tilde{F}_V$ , using the same knot vector as our preferred specification (with  $J = 11$ ) and the generalized Poisson point estimate  $\hat{\lambda}^{\text{gp}}$ , but under the second-price assumption (where we just interpret observed bids as private values), then we get a CDF estimate that is strongly rejected by the EA bidding model. Specifically, the estimator  $\tilde{F}_V$  is shifted to the left and parts of it lay outside of the bootstrapped 99% confidence bounds of  $\hat{F}_V(v; \hat{\mu})$ .<sup>32</sup>

#### 5.4 Model simulations and counterfactual analysis

We begin by exploring auction winners' market rents arising from private information on their willingness to pay. We define information rents as the random variable

$$\mathcal{I} \equiv V_{(1:N)} - P = V_{(1:N)} - \min\{B_{(2:N)} + \Delta, B_{(1:N)}\},$$

or the winner's private valuation minus the price she pays. Note that the randomness in  $\mathcal{I}$  comes from it being a function of three separate random variables:  $N$ ,  $V_{(1:N)}$ , and  $B_{(2:N)}$ . In our case, where we only occasionally observe the highest bid, and where we never directly observe  $N$ , our structural model estimates are needed so as to simulate the distribution of  $\mathcal{I}$ .

To do so, we proceed in three steps: first, we generate a random value of  $N$  from the generalized Poisson distribution with our point estimate  $\hat{\lambda}^{\text{gp}}$ ; second, we generate  $N$  random draws from the private-value distribution  $\hat{F}_V(v; \hat{\mu})$  and map them into bid space using a cubic interpolant of the inverse of  $\hat{\xi}(b; \hat{\alpha})$ ; finally, we collect the highest private valuation and the two highest bids to compute a simulated value for  $\mathcal{I}$ . We followed this process 1,000,000 times to get a large enough sample to reliably represent the moments of the distribution of  $\mathcal{I}$ . The lower panel of Table 5 summarizes information rents at several quantiles of the winner distribution. On average, the winner in an

<sup>32</sup>See Figure A.6 in the Supplemental Appendix for a robustness check in which we compare the bias induced by pricing-rule misspecification with the bias resulting from deviations in the scale invariance assumption.

auction from our sample retained an information rent of \$30.57, but the distribution is highly skewed to the right with a large standard deviation of \$26.36. From a seller's perspective, one of the main concerns is how to maximize their own revenues by extracting as much of these information rents as possible through auction design. We now conclude our discussion with a counterfactual exploration of optimal reserve prices.

One puzzling regularity within eBay data is that the vast majority of sellers set reserve prices at or near zero. In the 733 auctions from our empirical application, a reserve price is never used, but this could represent default seller policy (given that all of these auctions come from the same seller, CSRT). Nevertheless, in a broader set of used laptop data from the same time period, with 13,193 separate auctions, only 724 of them, or 5.5% overall, involved a reserve price that exceeded \$1; 359, or 2.6%, had a reserve price that exceeded \$10; and 156, or 1.1%, had a reserve price that exceeded \$100. The sparse use of reserve prices might be surprising given the literature on optimal mechanisms in which a positive reserve price is suggested even for a seller who values the item at \$0. Equipped with our estimates of the private-value distribution and the parameters of the generalized Poisson distribution, we can consider counterfactual experiments to understand why.

First though, following McAfee and McMillan (1987) and Harstad, Kagel, and Levin (1990), observe that because we have a symmetric, independent private-values model with an unknown number of risk-neutral bidders, standard auctions will be revenue equivalent as long as bidders have the same beliefs about the number of potential bidders. Our EA model has the same equilibrium allocation rule as the canonical first- and second-price formats—the bidder with the highest valuation always wins the object (efficiency)—and the expected payoff of the lowest possible type  $\underline{v}$  is zero; therefore, the revenue equivalence principle applies. Second, optimal auction design does not depend on the distribution of  $N$ ; see McAfee and McMillan (1987). Taken together, these facts imply that the optimal auction can be implemented using a first- or second-price format or via an EA.

This simplifies computation of counterfactuals: we can generate the correct expected revenues by considering a second-price auction, which avoids the need for solving the EA bid function using our estimated valuation distribution for every possible choice of reserve or value of the participation parameters. Let  $R$  denote revenue and let  $r$  denote reserve price. Expected revenues are given by

$$E(R|r) = r[1 - F_V(r)] \sum_{n=1}^{\infty} \rho_N(n; \boldsymbol{\lambda}) n F_V(r)^{n-1} + \int_r^{\bar{v}} v[1 - F_V(v)] f_V(v) \sum_{n=1}^{\infty} \rho_N(n; \boldsymbol{\lambda}) n(n-1) F_V(v)^{n-2} dv, \quad (20)$$

where the integral is solved using an adaptive recursive Simpson's rule. The maximizer,  $r^*$ , of this expression satisfies the equation

$$r^* = v_0 + \frac{1 - F_V(r^*)}{f_V(r^*)}, \quad (21)$$

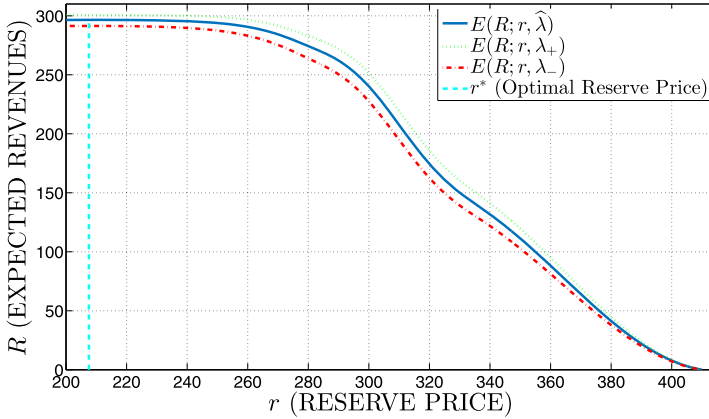


FIGURE 11. Counterfactual revenue curve comparisons.

first derived by Myerson (1981), where  $v_0$  is the seller's valuation. For the purpose of this discussion, we shall assume throughout that  $v_0 = 0$ .

Figure 11 depicts three expected revenue curves as a function of reserve price  $r$ : one deriving from model point estimates (solid line), and two hypothetical curves where the generalized Poisson parameters are altered. For both of these we compute an alternative  $\lambda$  vector so that  $\text{Var}[N]$  is held constant while for one, call it  $\lambda_+$ ,  $E[N]$  increases by 1 relative to the point estimates, and for the other, call it  $\lambda_-$ ,  $E[N]$  decreases by 1. A well known result concerning reserve prices in auctions by Bulow and Klemperer (1996) is that the benefit from attracting an additional bidder to an auction exceeds any possible gain by optimizing the reserve price. This result plays out strongly in the figure: the impact on expected revenues from changing  $E[N]$  by 1 is hundreds of times greater than the impact of moving from  $r = 0$  to  $r = r^*$ . However, our empirical results say something even stronger about optimal reserve prices on eBay: they are, in fact, almost entirely irrelevant to begin with! Although the expected revenue curves in the figure have an interior global optimum, under our model estimates, the maximum benefit a seller may reap by optimizing the reserve price is estimated to be \$0.0288. To put this number in perspective, if we assume that an eBay seller values her time at a conservative \$10/hour, then if it takes her longer than a mere 11 seconds to decide what the optimal reserve price should be, she would be better off by simply setting  $r = 0$  and allocating her time toward some other, more profitable use. This resolves the puzzle of why the vast majority of eBay sellers choose nonbinding reserve prices: the online auction house has done its job of attracting buyers to the market well enough so there is no longer any need to worry about this aspect of auction design.

## 6. DISCUSSION AND CONCLUSION

### 6.1 Model extensions

Two possible extensions are worth comment before we conclude. Our estimation framework presented here assumes away two common aspects of online auction markets:



binding reserve prices and inter-auction, dynamic incentives. Specifically, we have assumed here that sellers set reserve prices at a value of zero, as this was consistent with our empirical application to laptop data. We have also shown in our counterfactual analysis that if sellers have reservation values of zero, then they have little to gain by optimizing this particular aspect of auction design. However, in many data sets involving online auctions, nontrivial fractions of sellers do choose binding reserve prices. From an estimation standpoint, this addition introduces another source of sample selection: some actual bidders who would otherwise have been observed are withheld from the econometrician's view because they happen to have low private valuations. Incorporating binding reserve prices into the filter process involves significant complications of both the model and the estimator—it is no longer possible to estimate  $\lambda$  separately from  $G_B$  since both objects influence the link between the distributions of  $N$  and  $\tilde{N}$ —and is explored in ongoing work by Bodoh-Creed, Boehnke, and Hickman (2016). Another important assumption underlying our framework presented here is the idea that individual auctions may be treated in isolation as a series of static, one-shot games. In reality though, online auction markets for many consumer products are more complicated environments where bidders' opportunity costs of losing are not the full amount of their private valuations, since they can always return in a future period if they lose an auction today. This positive continuation value gives rise to an additional source of demand shading from intertemporal incentives. Still, much can still be learned from a framework like ours that characterizes only purely static strategic demand shading. Bodoh-Creed, Boehnke, and Hickman (2016) develop a model of dynamic platform markets like eBay. They characterize equilibrium bidding behavior and show that in the dynamic platform setting there is an intuitive layering of the intertemporal and static demand shading incentives. Their results imply that our identification and estimation strategy, based on a static, one-shot bidding model, can be applied in broader settings where bidding is influenced by the future outside option.

## 6.2 Conclusion

EAs are important market mechanisms in the world today, with eBay in 2010 accounting for total sales of \$25 billion from auction listings alone.<sup>33</sup> In this paper, we have made four contributions to research on EAs: first, we documented how failing to account for the nonstandard EA pricing rule can bias estimates of the latent valuation distribution. Second, we demonstrated that a realistic EA bidding model is nonparametrically identified by observables readily available. As part of this exercise, we solved another, independent problem by proposing a new identification strategy for inferring the distribution of  $N$  using observable lower bounds on bidder participation. Third, we proposed an estimation strategy to recover model parameters from actual eBay data in a flexible yet computationally tractable way. Although conceptually complex, the method we employed is easily implementable using a MATLAB toolbox that is available in the Supplemental

---

<sup>33</sup>See the eBay Annual Report 2010 available at <http://investor.ebayinc.com/annuals.cfm>. Our calculation is based on the gross merchandise value (GMV) multiplied by the reported share of GMV obtained under an auction-style format.



Code and Data. Thus, this research fills two important methodological gaps in the literature: (i) researchers no longer need to assume (incorrectly) that eBay bids are private valuations; (ii) having estimates of both the latent valuation distribution and the ex ante participation rate permits model simulations, where previous work only managed to identify the latent valuation distribution. Finally, we used our framework to shed light on empirically relevant questions in online auction design. For example, our model estimates explain why reserve prices are commonly not used by sellers: given the number of bidders who typically participate, they play very little role in determining revenues. On the other hand, varying the expected number of bidders who participate can play a significant role under observed market conditions. This suggests that the most relevant aspects of online market design take place at the level of the auction house itself, where policy levers exist to produce movements on the margin of marketwide, buyer–seller mix. Future research to investigate this question further will surely yield fruitful new insights and will benefit from the methodological advances we developed here.

#### REFERENCES

- Adams, C. P. (2007), “Estimating demand from eBay prices.” *International Journal of Industrial Organization*, 25, 1213–1232. [509]
- Bajari, P. and A. Hortaçsu (2003), “The winner’s curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions.” *RAND Journal of Economics*, 34, 329–355. [537, 538]
- Bajari, P. and A. Hortaçsu (2004), “Economic insights from Internet auctions.” *Journal of Economic Literature*, 42, 457–486. [509]
- Bodoh-Creed, A., J. Boehnke, and B. R. Hickman (2016), “How efficient are decentralized auction platforms?” Working Paper No. 2016-22, Becker Friedman Institute for Research in Economics. [539, 548]
- Bulow, J. and P. Klemperer (1996), “Auctions versus negotiations.” *American Economic Review*, 86 (1), 180–194. [519, 547]
- Consul, P. C. and G. C. Jain (1973), “A generalization of the Poisson distribution.” *Technometrics*, 15 (4), 791–799. [524]
- de Boor, C. (2001), *A Practical Guide to Splines*, Revised edition. Springer, New York. [534]
- Guerre, E., I. Perrigne, and Q. H. Vuong (2000), “Optimal nonparametric estimation of first-price auctions.” *Econometrica*, 68, 525–574. [513]
- Haile, P. A. and E. Tamer (2003), “Inference with an incomplete model of English auctions.” *Journal of Political Economy*, 111, 1–51. [510, 539]
- Harstad, R. M., J. H. Kagel, and D. Levin (1990), “Equilibrium bid functions for auctions with an uncertain number of bidders.” *Economics Letters*, 33 (1), 35–40. [546]
- Hickman, B. R. (2010), “On the pricing rule in electronic auctions.” *International Journal of Industrial Organization*, 28, 423–433. [507, 511, 512, 516, 525]

- Hickman, B. R. and T. P. Hubbard (2015), “Replacing sample trimming with boundary correction in nonparametric estimation of first-price auctions.” *Journal of Applied Econometrics*, 35, 739–762. [513, 514]
- Hickman, B. R., T. P. Hubbard, and Y. Sağlam (2015), “Structural econometric methods in auctions: A guide to the literature.” *Journal of Econometric Methods*, 1, 67–106. [508]
- Hortaçsu, A. and E. R. Nielsen (2010), “Commentary: Do bids equal values on eBay?” *Marketing Science*, 29, 994–997. [510]
- Hubbard, T. P. and H. J. Paarsch (2014), “On the numerical solution of equilibria in auction models with asymmetries within the private-values paradigm.” In *Handbook of Computational Economics*, Vol. 3 (K. Schmedders and K. L. Judd, eds.), 37–115. [530]
- Hulme, B. L. (1972), “One-step piecewise polynomial Galerkin methods for initial value problems.” *Mathematics of Computation*, 26 (118), 415–426. [534]
- Lucking-Reiley, D. (2000), “Auctions on the Internet: What’s being auctioned, and how?” *Journal of Industrial Economics*, 48, 227–252. [509]
- McAfee, R. P. and J. McMillan (1987), “Auctions with a stochastic number of bidders.” *Journal of Economic Theory*, 43 (1), 1–19. [546]
- McAfee, R. P. and J. McMillan (1996), “Analyzing the airwaves auction.” *Journal of Economic Perspectives*, 10, 159–175. [507]
- Milgrom, P. R. and R. J. Weber (1982), “A theory of auctions and competitive bidding.” *Econometrica*, 50, 1089–1122. [537]
- Myerson, R. B. (1981), “Optimal auction design.” *Mathematics of Operations Research*, 6, 55–73. [507, 519, 547]
- Nekipelov, D. (2007), “Entry deterrence and learning prevention on eBay.” Report. [538]
- Riley, J. G. and W. F. Samuelson (1981), “Optimal auctions.” *American Economic Review*, 71, 381–392. [507, 519]
- Roth, A. E. and A. Ockenfels (2002), “Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet.” *American Economic Review*, 92, 1093–1103. [509]
- Scholz, F. W. and M. A. Stephens (1987), “ $K$ -Sample Anderson–Darling tests.” *Journal of the American Statistical Association*, 82, 918–924. [517]
- Song, U. (2004a), “Nonparametric estimation of an eBay auction model with an unknown number of bidders.” Typescript, Department of Economics, University of British Columbia. [508]
- Song, U. (2004b), “Nonparametric identification and estimation of a first-price auction model with an uncertain number of bidders.” Typescript, Department of Economics, University of British Columbia. [508, 521]

Su, C.-L. and K. L. Judd (2012), “Constrained optimization approaches to estimation of structural models.” *Econometrica*, 80 (5), 2213–2230. [509, 533]

Vickrey, W. S. (1961), “Counterspeculation, auctions, and competitive sealed tenders.” *Journal of Finance*, 16, 8–37. [507]

Zeithammer, R. and C. P. Adams (2010), “The sealed-bid abstraction in online auctions.” *Marketing Science*, 29, 964–987. [509]

Zienkiewicz, O. C. and K. Morgan (2006), *Finite Elements and Approximation*. Dover, Mineola, NY. [530]

---

Co-editor Elie Tamer handled this manuscript.

Manuscript received 21 October, 2011; final version accepted 1 August, 2016; available online 14 September, 2016.