RAND Journal of Economics
Vol. 51, No. 3, Fall 2020
pp. 868–904

# Estimation in English auctions with unobserved heterogeneity

**Cristián Hernández***

**Daniel Quint****

**and**

**Christopher Turansick*****

*We propose a framework for identification and estimation of a private values model with unobserved heterogeneity from bid data in English auctions, using variation in the number of bidders across auctions, and extend the framework to settings where the number of bidders is not cleanly observed in each auction. We illustrate our method on data from eBay Motors auctions. We find that unobserved heterogeneity is important, accounting for two thirds of price variation after controlling for observables, and that welfare measures would be dramatically misestimated if unobserved heterogeneity were ignored.*

## 1. Introduction

■ In many settings where auctions are used, unobserved auction-level heterogeneity has a significant impact on valuations. For example, unobserved heterogeneity has been found to be economically significant in highway procurement and US Forest Service timber auctions, and structural estimation that ignored such heterogeneity would yield misleading estimates and policy conclusions.[1] The same holds for consumer products as well. Bodoh-Creed, Boehnke, and

* NERA Economic Consulting; cristian.hernandez@nera.com.
** University of Wisconsin-Madison; dquint@ssc.wisc.edu.
*** Georgetown University; cmt152@georgetown.edu.
The authors thank Danielle Labruzzo and Hannah O'Leary for capable research assistantship, and Aaron Bodoh-Creed, Joachim Freyberger, Brad Larsen, Mikkel Sølvsten, seminar participants at Johns Hopkins and Penn State, the Editor, and two anonymous referees for valuable comments. This article supplants an earlier working article, Quint (2015), "Identification in Symmetric English Auctions with Additively Separable Unobserved Heterogeneity."

[1] Working with bid data from Michigan highway procurement auctions, Krasnokutskaya (2011) finds that variation in private information accounts for only one third of bid variation, and that ignoring unobserved heterogeneity would lead to estimates of bidder markups that were more than double their actual level. Working with data from US Forest Service timber auctions, Athey, Levin, and Seira (2011) note that allowing for unobserved heterogeneity in estimation "appears crucial," as they find "implausibly high bid margins when we fail to account for [it]." Working with data from timber auctions in a different region, Aradillas-López, Gandhi, and Quint (2013) find positive correlation among bidder valuations—possibly due to unobserved heterogeneity—"even conditional on the rich vector of available covariates (the presence of which is often used to defend the IPV assumption)." They find optimal reserve prices and expected seller profit to be significantly misestimated when it is ignored: for example, they find the Forest Service's actual reserve price

Hickman (2018a) recently assembled an extremely detailed dataset on eBay sales of unopened first-generation Amazon Kindle Fire tablets, and found that by combining this rich data (the entire .html page of the listing) with sophisticated machine learning techniques, they could explain 42% of the variation in prices—more than three times what is explained by simpler analysis of a more-typical subset of the variables in the dataset. Thus, even when heterogeneity across listings is not truly unobservable, standard analysis fails to fully account for it, leaving residual variation that is inconsistent with an independent private values model.

Although there are well-established techniques for dealing with unobserved heterogeneity in the estimation of first-price auction models, this is much less true for English auctions. As we discuss below, the techniques used in first-price auctions do not translate to the English auction setting. Until recently, the empirical literature on English auctions ignored both correlation and unobserved heterogeneity. Three recent advances offer ways to account for it, each with significant limitations. One approach (see Aradillas-López, Gandhi, and Quint (2013)) identifies only bounds on measures of interest, as they are not point identified, and requires wide exogenous variation in participation across auctions for the bounds to be narrow; further, the approach works only for certain counterfactuals and not others. A second approach (Roberts (2013)) relies on an assumption that the unobserved heterogeneity is observed by the seller, and that the reserve price is set as a strictly increasing function of this heterogeneity. A third approach (Mbakop (2017), Freyberger and Larsen (2017)) depends on the assumption that at least two (and, depending on the other assumptions, as many as five) losing bidders bid up to their valuations. In the absence of one of the latter two assumptions, we are not aware of any positive results on point identification of an English auction model with unobserved heterogeneity. Indeed, Athey, Levin, and Seira (2011) had data from both first-price and English auctions, but chose to estimate the structural model using only the first-price data for exactly this reason.[2]

This article aims to fill this void. Focusing on a model of independent private values with one-dimensional, separable unobserved heterogeneity, we show that the model is point identified if there is *any* exogenous variation in the number of bidders across auctions—in the absence of information-revealing reserve prices, and with only a single bidder's bid (the highest losing bid) being assumed to reveal her valuation. We extend this result in two ways to account for settings where the number of bidders in each auction is not perfectly observed. To illustrate the approach, we apply it to data from eBay Motors car auctions. We find that after controlling for observable covariates, auction-level unobserved heterogeneity still accounts for 67% of price variation, and ignoring this heterogeneity would lead to a drastic (230%) overestimate of bidder surplus.

## 2. Related literature

■ The literature distinguishes the case where bidders perceive their valuations as being correlated (typically modeled as affiliated) from the case where bidder valuations are independent conditional on variables they can see but the analyst cannot (unobserved heterogeneity). In first-price auctions, equilibrium bidding depends on both a bidder's valuation and her belief about others' valuations, so these are distinctly different models. In English auctions with private values, bidding is effectively in dominant strategies, so the two models are observationally equivalent.

As noted in the Introduction, the auction literature contains well-established techniques that allow for either unobserved heterogeneity or correlated valuations in first-price auctions. Li, Perrigne, and Vuong (2000), Krasnokutskaya (2011), and Hu, McAdams, and Shum (2013) build on the "measurement error" approach of Li and Vuong (1998) to estimate a model of

levels to be about as high as they could be to meet its stated policy goal of selling at least 85% of offered tracts, whereas these reserves would seem overly cautious by a substantial margin in the absence of unobserved heterogeneity.

[2] Aradillas-López, Gandhi, and Quint (2013, footnote 7) note: "Referring to an earlier version of Athey, Levin, and Seira (2011), Athey and Haile (2006, p. 33) write: 'To account for this correlation [of bids within a first-price auction], ALS select a model of independent private values with unobserved heterogeneity. . . . This model is not identified in data from ascending auctions; thus, ALS focus their structural estimation on first-price auctions.' "

conditionally independent values or values with unobserved auction-level heterogeneity. In a separate approach, Li, Perrigne, and Vuong (2002) extend the estimation technique of Guerre, Perrigne, and Vuong (2000) to affiliated private values. Compiani, Haile, and Sant'Anna (2019) allow for both unobserved heterogeneity and affiliation of signals/interdependent values, and discuss other approaches. However, none of these approaches work for English auctions, as they rely on observation of multiple informative bids from each auction—either as independent noisy estimates of the unobserved variable, or to account for the competition a bidder faces conditional on her own valuation—which is not available in an English auction.

As discussed in Aradillas-López, Gandhi, and Quint (2013) and Roberts (2013), most of the empirical literature on English auctions has assumed that bidder valuations are independent (conditional on observables), ignoring both correlation and unobserved heterogeneity. Early work modeled bidding as a button auction, where bidding revealed the exact price at which each losing bidder stopped wanting to win. Haile and Tamer (2003) introduced a more realistic but "incomplete" model of bidding in open-outcry ascending auctions, based on two relatively weak assumptions about the relationship between valuations and bids: a bidder never bids more than her valuation, and never loses an auction when she still could have bid less than her valuation. Still assuming independent private values, they show how these assumptions lead to set identification of the underlying primitives from bid data and estimate useful bounds.

Three recent strands of literature have moved away from the assumption of independent private values, allowing for either correlation of values or unobserved heterogeneity. Aradillas-López, Gandhi, and Quint (2013) use variation in the number of bidders across auctions to construct bounds on relevant counterfactual measures—expected profit and bidder surplus at different reserve price levels, and the seller-optimal reserve. Although the model of valuations is extremely general (except for assuming ex ante symmetry), this method requires accurate observation of the number of bidders in each auction, and the resulting bounds can be fairly wide unless the number of bidders varies a lot. Coey, Larsen, Sweeney, and Waisman (2017) demonstrate similar bounds for a model with asymmetric bidders. Another limitation of this method is that it makes no attempt to fully recover underlying model primitives, making it useful for certain counterfactuals but not for others.

A second approach, introduced by Roberts (2013), assumes that the *seller* in each auction has access to the same information the bidders do—a one-dimensional variable that is unobserved to the analyst—and sets a reserve price that is strictly increasing in this variable. He notes that such behavior will often be optimal, but does not require that sellers set the optimal reserve, just one that is monotonic in the unobserved variable. The reserve price and the transaction price then give two separate noisy observations of the unobserved variable, identifying the model. The assumption that reserve prices essentially reveal the unobserved characteristic of the object (and that all sellers set them in the same way) is plausible in the environment Roberts studies, used car auctions, but may be less applicable in online auctions, where reserve prices are often set very low and there is great heterogeneity in seller sophistication.

A third approach does not rely on variation in either the number of bidders or the reserve price, but depends on the assumption that several bidders' valuations in each auction are revealed by their bids. Mbakop (2017) shows that a fairly general symmetric model with finite unobserved heterogeneity is identified if *five* order statistics of bidders' valuations are observed in each auction, corresponding to five losing bidders bidding up to their exact valuations; this requirement can be reduced to three order statistics in the presence of an instrument like a varying reserve price. Luo and Xiao (2019) show identification using three consecutive order statistics without an instrument.[3] Freyberger and Larsen (2017) combine ideas from Decarolis (2018) and Song (2004)[4] to show identification of a model with unobserved heterogeneity (in a separable model

[3] Konstantopoulos and Yuan (2019) show that an additively separable model like ours is identified from the distributions of the two highest valuations, but bids in English auctions do not typically reveal the highest valuation.

[4] The former extends the measurement error approach discussed earlier to the case of a first-price auction where only the winning bid is observed, using a non-binding reserve price as the second instrument. The latter uses two order

similar to ours) and unobserved participation; the cost is that they need to assume bidding precisely reveals *two* order statistics of bidders' valuations (the second and third highest), in addition to a reserve price that is correlated with the unobserved value component. Assuming several bidders' valuations are revealed by bids is natural in a button auction or in a sealed-bid second-price auction, but much harder to believe in open-outcry or online auctions, where bidders have more freedom in when to bid.[5] Komarova (2013) works with a general asymmetric model of valuations that is not point identified and focuses on obtaining bounds on the joint and marginal distributions of valuations. Although her approach is compatible with the bidding assumptions of Haile and Tamer (2003) and various different assumptions about what outcomes are observable, she focuses on the case where all losers' valuations are exactly learned, and obtains bounds that are quite wide without this assumption.

In contrast with the existing literature on English auctions, then, we show point identification of a model of English auctions that allows for unobserved heterogeneity, with no assumptions about reserve price setting (other than that reserves do not bind) and no need to observe more than the number of bidders and the winning bid in each auction; and although we do rely on exogenous variation in the number of bidders, we require only minimal variation to get point identification, and can deal with imperfect observation of the number of bidders.[6]

## 3. Model and nonparametric identification

□ **Model.** We first lay out our model, then discuss its key assumptions; after that, we will show different sets of conditions under which the model is identified from plausible data.

We assume the analyst has data from a series of auctions with the same set of primitives, which includes the transaction price $T$ in each auction, and that the number of bidders $N$ varies (at least somewhat) across auctions. We maintain the following two assumptions throughout the article:

*Assumption 1.* Auctions in the data do not have binding reserve prices, and the transaction price $T$ in each auction is equal to the second-highest valuation.

*Assumption 2.* Any variation in the number of bidders $N$ across auctions is exogenous, that is, independent of bidder valuations.

We work with a model of symmetric, independent private values with additively separable unobserved heterogeneity. Specifically, a bidder's private value is

$$v_i = \theta + \epsilon_i, \tag{1}$$

where $\{\epsilon_i\}$ are independent and identically distributed random variables (across bidders and auctions) drawn from a distribution $F_\epsilon$, and $\theta$ is a random variable (independent of $\{\epsilon_i\}$ and across auctions) drawn from a distribution $F_\theta$. We assume $F_\epsilon$ and $F_\theta$ have convex support in $\mathbb{R}^+$ and admit density functions $f_\epsilon$ and $f_\theta$, and we will assume these two densities satisfy a strong smoothness condition almost everywhere. Specifically, we will assume that they are *piecewise real analytic*, as we will define below.

*Definition 1.* Given an open set $U \subseteq \mathbb{R}$, a function $f : U \to \mathbb{R}$ is **real analytic** if at every point $x \in U$, there is an open set containing $x$ on which $f(z) = \sum_{d=0}^{\infty} a_d (z - x)^d$.

statistics—the highest and second-highest losing bids in an English auction—to identify an IPV English auction model when the number of bidders is not observable.

[5] Under the bidding assumptions of Haile and Tamer (2003), the highest losing bidder's valuation must be close to her bid, but the valuations of other losing bidders need not be.

[6] On the other hand, our model of valuations is more restrictive than some others in the literature, and we do rely on other substantive assumptions—see the discussion in Section 3.

This is equivalent to the function being $C^\infty$ (infinitely many times continuously differentiable) and locally equal to a convergent Taylor series.[7] For an unknown function, however, this is a very strong assumption to impose: if two real analytic functions $f$ and $g$ on a compact domain agree on an open interval, they agree everywhere, and thus a real analytic function is defined globally by its behavior in the neighborhood of a single point. This is a much stronger requirement than we need; we need only the local properties of real analytic functions, not this global extrapolability. We therefore define a particular notion of piecewise smoothness that will suffice for our purposes:

*Definition 2.* Given a set $X \subseteq \mathbb{R}$ (not necessarily open), a function $f : X \to \mathbb{R}$ is **piecewise real analytic** if for every $x \in X$, there is a $\delta > 0$ and a real analytic function $g : (x - \delta, x + \delta) \to \mathbb{R}$ such that $f(z) = g(z)$ for $z \in [x, x + \delta)$.

Thus, a piecewise real analytic function can be thought of as a pasting together of various real analytic functions on different domains, with no assumption of differentiability at these pasting points. Although $f$ need not be differentiable everywhere, this does require that it be *right* differentiable infinitely many times at every point, and that at every point $x$, it is equal to its Taylor expansion at $x$ (based on right- derivatives) on some neighborhood $[x, x + \delta)$. This will turn out to be enough to prove identification, so we will maintain this assumption:

*Assumption 3.* The density functions $f_\epsilon$ and $f_\theta$ are continuous and piecewise real analytic.

□ **Discussion of model.** Given the setting of private-value English auctions, our focus on unobserved heterogeneity rather than correlated values is somewhat arbitrary, as the two models are observationally equivalent. (They have different implications for certain counterfactuals, such as changing the auction format; but bid data does not allow us to distinguish between the two models, and recovery of $F_\epsilon$ and $F_\theta$ is unaffected by the choice.[8]) Our assumption of additive separability, however, is quite substantive; our model of valuations is much more restrictive than those of Aradillas-López, Gandhi, and Quint (2013), Hu, McAdams, and Shum (2013), and Mbakop (2017), in line with those used by Krasnokutskaya (2011), Roberts (2013), and Freyberger and Larsen (2017) and in a wide range of empirical studies of first-price auctions. Though we focus on the case of additive separability, as long as valuations are bounded away from zero, everything we do extends trivially to the multiplicative model $v_i = \theta\epsilon_i$, simply by taking the log of transaction prices in the data and using the techniques below to identify the distributions of $\log\theta$ and $\log\epsilon_i$.

Our local smoothness assumption—piecewise real analytic distributions—is admittedly fairly strong. However, we have also proved analogs to our identification results (Theorems 1 and 2) for the case where $\theta$ and $\epsilon_i$ are drawn from distributions with discrete support, with no further restrictions on the distributions. Thus, the identification result holds both when the distributions are "very smooth" and "very chunky," giving us hope that it might be true more generally.

In real English auctions, as opposed to button auctions or sealed-bid second-price auctions, our assumption that transaction price perfectly matches the second-highest valuation is a substantive one. Under the bidding assumptions of Haile and Tamer (2003) mentioned above, in an auction with bid increment $\Delta$, the second-highest valuation must be in the interval $[T - \Delta, T + \Delta]$ if there was no jump bid at the end of the auction. In a setting where bidders were choosing which

[7] Most common empirical specifications use real analytic functions: for example, Fox, Kim, Ryan, and Bajari (2012) note that conditional choice probabilities in a random-coefficients logit model are real analytic, and use this to prove identification of that model. Krantz and Parks (2002) offer more details on real analytic functions.

[8] If we assume affiliated or otherwise correlated private values, then if they are exchangeable (symmetric) and drawn from an infinite sequence of potential valuations, conditional independence is without loss of generality by de Finetti's theorem; and because bidding is essentially in dominant strategies, bids would be the same if bidders observed $\theta$ and $\epsilon_i$ separately (unobserved heterogeneity) or only $v_i$ (conditionally independent values).

auction to enter and jump bids could potentially deter entry, this might be a strong assumption; in a setting like eBay, with automated proxy bidding, jump bids are effectively impossible, and this assumption seems reasonable.

Although we assume the second-highest valuation is exactly revealed by bids, we do not make the analogous assumption about other losing bids—that is, unlike Freyberger and Larsen (2017) and Mbakop (2017), we do not assume that other losing bidders (besides the second-highest) bid up to their valuations. We could, as in Haile and Tamer (2003), interpret each losing bid as a lower bound on the bidder's valuation, which would still be informative—for example, multiple losing bids clustered close to the transaction price would indicate a tighter distribution of $\epsilon_i$, suggesting more of the variation in prices could be attributed to variation in $\theta$. To keep estimation simple, we have chosen to ignore other losing bids completely, as our model is point identified from just transaction prices; although in principle, losing bids could potentially be useful in getting sharper estimates or testing the other assumptions.

Finally, the assumption of exogenous $N$, though not unique to this article, is a substantial one. Athey and Haile (2002) refer to this as "exogenous participation." This could occur due to exogenous variation in the number of available bidders, or in a setting with endogenous entry where bidders do not learn their valuations until after sinking their entry costs.[9] Aradillas-López, Gandhi, and Quint (2013) make this assumption to identify bounds on certain counterfactuals in an English auction setting, and Aradillas-López, Gandhi, and Quint (2016) offer a partial test of this assumption using bid data from English auctions. This same assumption has also been made in several empirical articles on first-price auctions: Guerre, Perrigne, and Vuong (2000, 2009), Haile, Hong, and Shum (2003), Gillen (2010), and Aryal, Grundl, Kim, and Zhu (2018), for example, make this assumption to improve the efficiency of their estimator, to identify a risk aversion parameter, to distinguish private from common values, to identify a "level-*k*" model of strategic sophistication, and to estimate a model with ambiguity aversion, respectively, and Liu and Luo (2017) present a test of the assumption on first-price auction data.[10] Although we vary (below) what we assume about the *observability* of $N$, we will maintain the assumption it is exogenous.

□ **Nonparametric identification.** We will build up our identification results in pieces, starting with the simplest case. Proofs of Theorems 1, 2, 3, and 4 are in the Appendix. Note that all identification results are up to offsetting horizontal shifts in the two distributions.[11]

*Case 1: N observed.* If (along with the transaction price) the number of bidders is directly observed in each auction, then any variation in $N$ suffices to point identify the model:

*Theorem 1.* If $N$ varies exogenously and takes at least two values, then the model is nonparametrically identified from observation of $(T, N)$.

For intuition about why transaction prices from two realizations of $N$ should identify the model, consider a simple example where $F_\theta$ and $F_\epsilon$ are known to be normal distributions and $N$ takes the values 3 and 4. As noted above, identification is only up to offsetting horizontal shifts in the two distributions, so one of the two mean parameters—say, $\mu_\theta$—is a normalization. $\mu_\epsilon$

[9] In such models, entry costs are interpreted as information acquisition costs. Levin and Smith (1994) present one such model, where bidders have identical entry costs and play a mixed-strategy equilibrium at the entry stage. Lu (2010), Moreno and Wooders (2011), and Lu and Ye (2013) consider a different model with heterogeneous (private) entry costs that are independent of valuations, where bidders play cutoff strategies at the entry stage. In both these models, the realized number of bidders is either random or determined by the realization of entry costs, and therefore independent of valuations, so either of these models would satisfy our Assumption 2.

[10] Bajari and Hortaçsu (2005) also make this assumption in estimation of a first-price auction model with risk aversion, but they are analyzing experimental data from Dyer, Kagel, and Levin (1989) in which it holds by construction.

[11] That is, because all observations are functions of $\theta + \epsilon_i$, the distributions $\tilde{F}_\theta$ and $\tilde{F}_\epsilon$ defined by $\tilde{F}_\theta(t) = F_\theta(t - K)$ and $\tilde{F}_\epsilon(t) = F_\epsilon(t + K)$ for any constant $K$ would be observationally equivalent to the distributions $F_\theta$ and $F_\epsilon$.

is then determined by the mean of the distribution of transaction prices when $N = 3$ (because the second highest of three *i.i.d.* draws from a symmetric distribution has the same mean as the underlying distribution). What remains is to recover the two variance parameters.

Let $F_{T|n}$ denote the observed distribution of transaction prices conditional on $N = n$. To be consistent with our model, it must be that $F_{T|4} \succsim_{FOSD} F_{T|3}$; the magnitude of $\sigma_\epsilon$ is determined by the distance between the two distributions. In the case of normal distributions, this is particularly clean, as the difference in mean between the second highest of four and the second highest of three independent draws from the same normal distribution is linear in the standard deviation of that distribution[12], so $\sigma_\epsilon$ can be calculated directly from the difference in the means of $F_{T|3}$ and $F_{T|4}$. $\sigma_\theta$ is then pinned down by the variance of either $F_{T|3}$ or $F_{T|4}$ that is not explained by $\sigma_\epsilon$.

Of course, the proof that $F_\epsilon$ and $F_\theta$ are nonparametrically identified (given in the Appendix) is more complicated. We want to show that only a single pair of density functions $(f_\theta, f_\epsilon)$ can be consistent with the data, so we suppose there were two such pairs, $(f_\theta, f_\epsilon) \neq (g_\theta, g_\epsilon)$, which would both lead to the two observed bid distributions $F_{T|n}$ and $F_{T|n'}$, and show that this leads to a contradiction. Suppose 0 is the bottom of the support of all the distributions. First, we show that if both pairs of densities give the observed bid distributions $F_{T|n}$ and $F_{T|n'}$ in a neighborhood of 0, then they must be equal and have identical derivatives of every order at 0, that is, $f_\theta(0) = g_\theta(0)$, $f'_\theta(0) = g'_\theta(0)$, $f''_\theta(0) = g''_\theta(0)$, and so on (and likewise for $f_\epsilon$ and $g_\epsilon$). If the densities are piecewise real analytic, this implies that $f_\theta = g_\theta$ and $f_\epsilon = g_\epsilon$ on some interval $[0, \delta)$. Second, then, we let $t^* > 0$ be the point at which the two pairs of densities $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ begin to diverge. If they are equal up to $t^*$ but not to the right of $t^*$ and both pairs of densities generate $F_{T|n}$, we show that either $f_\theta > g_\theta$ and $f_\epsilon < g_\epsilon$ or vice versa on some interval $(t^*, t^* + \delta)$; and because $F_{T|n}$ and $F_{T|n'}$ depend differently on the distribution of $\epsilon$, we then show that both pairs of densities cannot generate $F_{T|n'}$ if they both generate $F_{T|n}$.

Thus, the key properties of the distributions that we rely on in the proof are (i) that if two distributions have identical derivatives of all orders at a point, then they are equal to each other in a neighborhood to the right of that point; and (ii) that if two distributions are identical up to $t^*$ and then begin to diverge, then one is strictly less than the other on a neighborhood to the right of $t^*$. Both of these follow easily from the assumption that distributions are piecewise real analytic. However, our intuition is that the identification result holds even more generally. To illustrate our intuition that smoothness is not required for identification, we have proved the exact analog to Theorem 1 (as well as of Theorem 2) for the case where $F_\epsilon$ and $F_\theta$ have discrete, rather than continuous, support, with no restrictions on the distributions.[13]

*Case 2: N unobserved, but drawn from multiple known distributions.* If $N$ is not directly observable in the data, we cannot apply Theorem 1. However, we can still proceed if we have access to a "participation shifter"—an auction-level covariate that affects $N$ but is independent of valuations. Let $X$ denote such a variable, and let $\mathbf{p}(n|x) = \Pr(N = n|X = x)$ be the probability distribution of $N$ given a particular realized value of $X$. For example, $X$ could be the day of the week, if weekend auctions were known to be better attended than mid-week auctions but similar objects were sold at both.

*Definition 3.* Two distributions $p$ and $q$ with common support in $\mathbb{Z}^+$ satisfy the monotone likelihood ratio property (MLRP) if the ratio $p(n)/q(n)$ is weakly increasing in $n$.

[12] It is approximately $0.297\sigma$—see, for example, the table at the end of Harter (1961).

[13] Suppose $F_\theta$ and $F_\epsilon$ both have support $\mathbb{Z}^+$, and let $t_k$ and $e_k$ denote $\Pr(\theta = k)$ and $\Pr(\epsilon = k)$, respectively. A transaction price of 0 occurs when both $\theta$ and the second-highest $\epsilon_i$ are equal to 0, and so $\Pr(T = 0|N) = t_0(Ne_0^{N-1} - (N-1)e_0^N)$. Observing $\Pr(T = 0|N)$ for two values of $N$ gives two equations for two unknowns ($t_0$ and $e_0$), and it is not hard to show we can recover them. Similarly, if we observe $\Pr(T = 1|N)$ for two values of $N$, this gives us two more equations, now for four unknowns ($t_0$, $t_1$, $e_0$, and $e_1$); as we already know two of them, we can recover the other two. We can show inductively that at every step, observation of $\Pr(T = k|N = n)$ and $\Pr(T = k|N = n')$ allows recovery of $e_k$ and $t_k$, proving identification in the discrete case.

If the distribution of $N$ (as a function of $X$) is known, and this distribution at different values of $X$ satisfies the MLRP, then once again, given any variation, the model is identified:

*Theorem 2.* If $\mathbf{p}(\cdot|X)$ is known and takes at least two values $\mathbf{p}(\cdot|x') \neq \mathbf{p}(\cdot|x)$ satisfying the MLRP, then the model is nonparametrically identified from observation of $(T, X)$.

Once again, we prove the result when the distributions are piecewise real analytic, but the same result holds when $\theta$ and $\epsilon_i$ are discrete valued. If the densities $f_\theta$ and $f_\epsilon$ are real analytic (rather than piecewise real analytic), then Theorem 2 (and therefore Theorem 4) does not require the MLRP, and identification holds as long as $\mathbf{p}(\cdot|x') \neq \mathbf{p}(\cdot|x)$ for two realizations of $X$.[14] Analogs to Theorems 1 and 2 also hold in an asymmetric model where there is one strong bidder in each auction and a varying number of ex ante identical weak bidders, provided we also observe the identity of the winner (that is, whether the strong bidder won) in each auction.

*Case 3: N partly observed, but drawn from unknown distributions.* Finally, we consider the case where the distribution of $N$ (conditional on $X$) is not known ex ante, but the number of bidders in each auction is partly observed. For example, in eBay auctions, we can observe the number of bidders who submitted bids in each auction; but we cannot expect this number to match the relevant, "true" $N$, for the following reason. In online auctions, bidders arrive to the auction over time, and can see the standing high bid (essentially the second-highest bid submitted so far, given eBay's proxy bidding system) at the time they arrive. Bidders who arrive after the auction price has risen past their own valuation will have no reason to cast a bid, and will not be recorded in the number of observed bids $N_{\text{obs}}$. To correctly interpret the transaction price as the "second-highest out of $N$ valuations," however, these bidders should be counted in $N$.

We can deal with this partial observability of $N$ by explicitly positing an "entry model," which determines how bidders arrive at the auction and choose whether or not to bid based on the actions of the bidders who arrived before them. For example, bidders might arrive sequentially in random order, and on arrival, submit proxy bids equal to their valuations as long as that is above the standing high bid at the time.[15] Any such entry model will yield a mapping from the "true" number of potential bidders $N$ in an auction to the probability distribution of the number of observed bids, $N_{\text{obs}}$, giving a set of probabilities $\Pr(N_{\text{obs}} = k|N = j)$ for each $k \leq j$. Under very general conditions, the mapping determined by any such entry model is invertible, allowing us to infer the distribution of $N$ from the distribution of $N_{\text{obs}}$:

*Theorem 3.* If the distribution of $N$ has bounded support and the entry model implies that $\Pr(N_{\text{obs}} = n|N = n) > 0$ for every $n$ in the support of $N$, then the distribution of $N$ is nonparametrically identified from the distribution of $N_{\text{obs}}$.

Note, however, that Theorem 3 need not hold when the support of $N$ is unbounded; we give a counterexample in the Appendix. In terms of nonparametric identification, requiring bounded support seems without loss, as the support of $N$ is trivially bounded above by, say, the Earth's population. That said, bounded support is at odds with many obvious choices for parameterization, including the one we use in our empirical application—we discuss this further in the next section.

Combining Theorems 2 and 3 gives a final result:

[14] Specifically, part 1 of the proof of Theorem 2 (in the Appendix) never invokes the MLRP, and suffices to prove identification when the densities are real analytic.

[15] Hickman, Hubbard, and Paarsch (2017) consider this entry model (they refer to it as a "filter process"), and show that the distribution of $N$ is identified from the distribution of $N_{\text{obs}}$; Platt (2017) further develops identification of this model when $N$ is drawn from a Poisson distribution. Bodoh-Creed, Boehnke, and Hickman (2018b) use it in their empirical application, having calculated the terms $\Pr(N_{\text{obs}} = k|N = j)$ by simulation. In the Appendix of our article, we offer a formula (Lemma A7) for iteratively calculating closed-form expressions for $\Pr(N_{\text{obs}} = k|N = j)$ for this entry model; however, for reasons we discuss there, we use a different entry model in our empirical application.

*Theorem 4.* Suppose $\mathbf{p}(\cdot\,|X)$ is unknown but has bounded support, and $X$ has two realizations $x$ and $x'$ such that $\mathbf{p}(\cdot\,|x)$ and $\mathbf{p}(\cdot\,|x')$ are different and satisfy the MLRP. Given knowledge of the correct entry model, the rest of the model—$\mathbf{p}(\cdot\,|X)$, $f_\theta$, and $f_\epsilon$—is nonparametrically identified from observation of $(T, X, N_{\text{obs}})$.

## 4. Empirical application

□ **Setting and data.** To illustrate how one can operationalize these results, we apply them to a rich dataset of online auctions for used cars that took place on eBay between February and October 2006, previously studied by Lewis (2011).[16] For each auction, the dataset contains the starting bid (public reserve price), transaction price, the highest bid placed by each bidder (bid value, timing, and identity of the bidder) except for the winning bid, a rich set of covariates with characteristics of the car being sold (e.g., model, make, year, mileage, and book value), information about the eBay listing and the seller (e.g., number of photos, seller's feedback score, and negative feedback), and other information about the auction (e.g., end date and time and auction length in days).[17]

The eBay site uses an automatic proxy bidding system. With this system, a bidder basically tells eBay her maximum willingness to pay, and the system then bids on her behalf, always bidding the minimal amount to exceed the other existing bids. Thus, in theory, a bidder could arrive at the auction, submit a proxy bid in the amount of her willingness to pay, and never bid again—and in fact, the eBay website advises bidders to do this. However, in actuality many bidders bid multiple times in the same auction.

We begin with a sample of 26,781 auctions with two or more bids that have all the relevant covariates populated. To account for auction-specific observable heterogeneity in our framework, we normalize the transaction price through the homogenization procedure of Haile, Hong, and Shum (2006): we estimate a linear regression of the log-transaction price on observables, and calculate a normalized log-transaction price by subtracting the predicted log-transaction price from the actual. Our regression equation is

$$\log T = z'\delta + I'\eta + \zeta, \tag{2}$$

where $z$ is a vector of auction-specific observables and $I$ a vector of dummies for the different values of $N_{\text{obs}}$. $z$ contains the same observables as the final specification in Lewis (2011, Table 2, column 6)—log miles, number of photos, number of photos squared, number of options, log-seller feedback, percentage of negative feedback, log-book value, and fixed effects for model, year, and week. Fixed effects for the number of observed bidders are included in the regression because we expect participation to affect transaction price and we wish to avoid omitted variable bias if other observables are correlated with $N_{\text{obs}}$. Only the $z$ observables are used to normalize transaction prices, however, as our identification strategy relies on the variation in price due to participation; normalized log-transaction prices are therefore defined as $\log T - z'\delta$. (If we drop $I$ from the normalization regression, we get nearly identical estimates for $\delta$ and normalized prices.) Table 6 in the Appendix shows the results of the normalization regression.

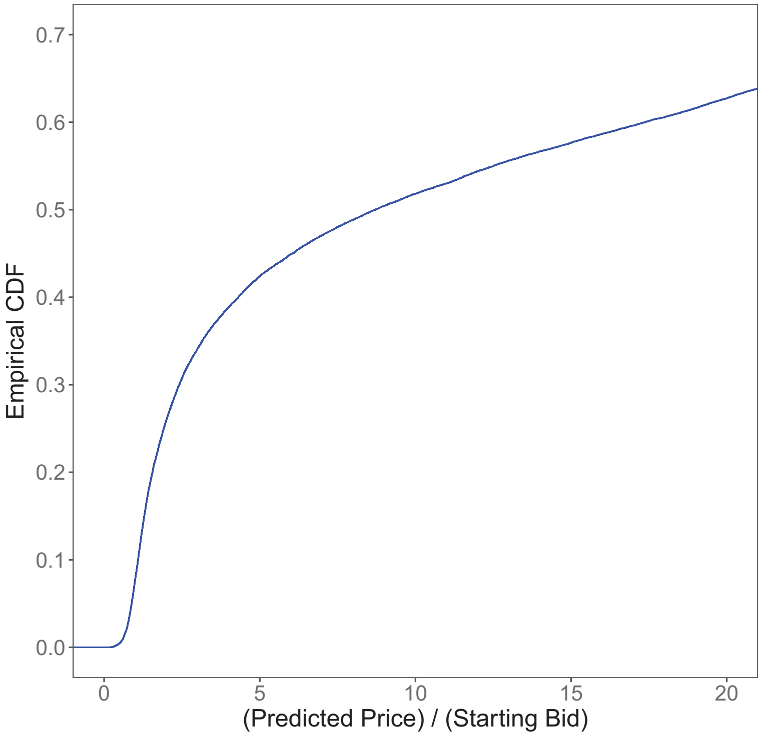
Because our identification results assume that reserve prices are not binding, we focus on the auctions in which the predicted price is several times higher than the opening bid. Figure 1 shows the empirical distribution of the ratio of predicted transaction price (from the normalization regression) to starting bid. The graph shows that about 40% of sellers set starting bids that are more than one fifth of the predicted transaction price (predicted price/starting bid below 5), which (to be conservative) we interpret as potentially binding. The remaining sellers do not

[16] The data are publicly available through the *American Economic Review* website, at http://www.aeaweb.org/articles?id=10.1257/aer.101.4.1535; see Lewis (2011) for a thorough description.

[17] Lewis (2011) focuses on asymmetric information, and the use of detailed photos to create an enforceable "contract" to deliver a car that matches those photos. Thus, for him, the photographs in the listing are a key choice of the seller; we, on the other hand, take them as given, and estimate buyer preferences given the products as offered.

FIGURE 1

EMPIRICAL DISTRIBUTION OF PREDICTED PRICE OVER STARTING BID [Color figure can be viewed at wileyonlinelibrary.com]



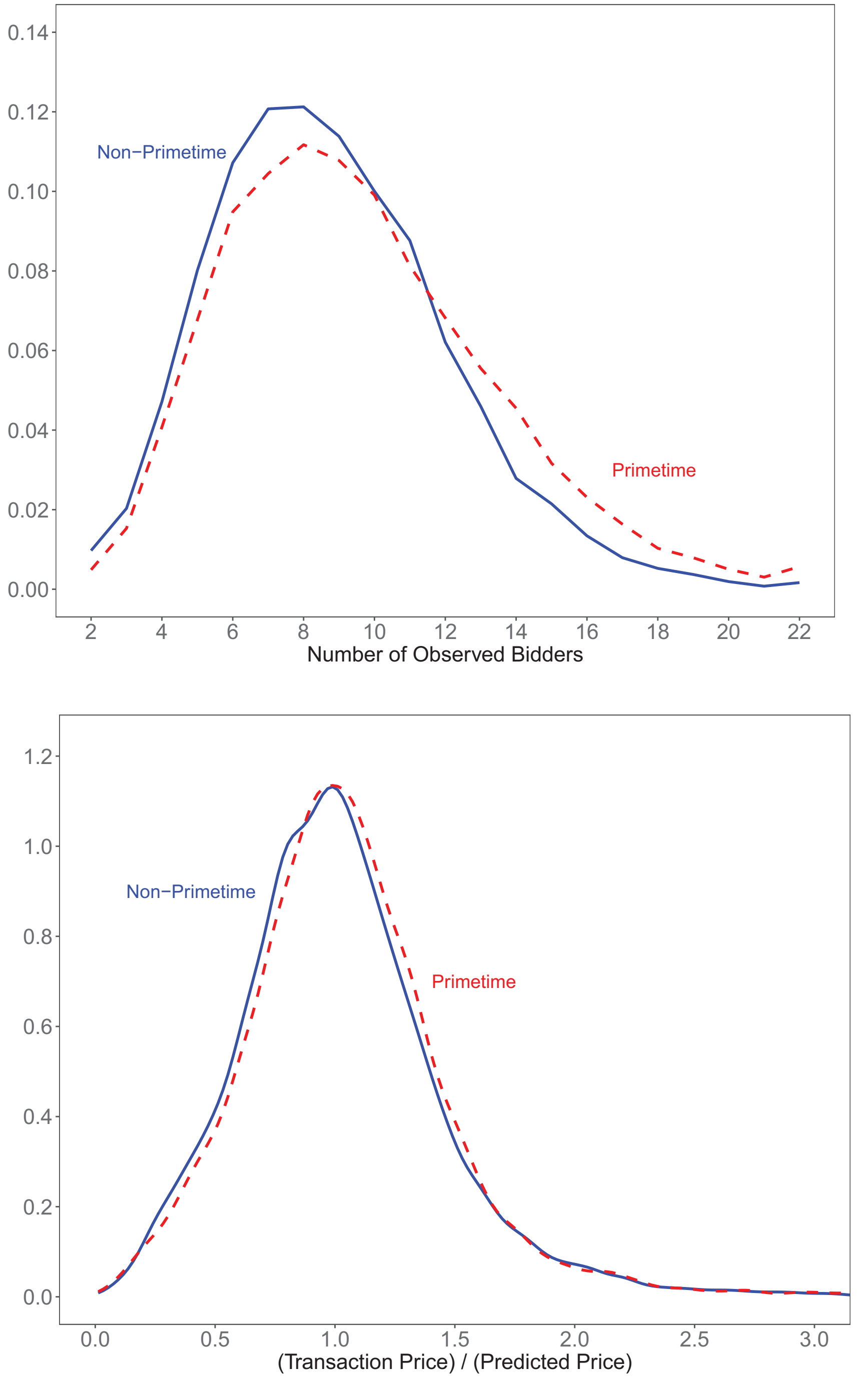

appear to be using the starting bid in a meaningful way. We drop auctions with predicted price less than five times the opening bid, which leaves 15,424 auctions in the sample. Table 7 in the Appendix compares summary statistics of our sample compared to the auctions dropped. (On average, the dropped listings—those auctions with potentially binding reserve prices—had 13% fewer miles, a book value 15% higher, and sold at a price 27% higher.)

To apply our identification results, we also need an exogenous participation shifter, that is, a variable that affects the number of bidders but does not otherwise affect the value of the object. A plausible one is the end time of the auction. We define an auction as "primetime" if it ends between 4 p.m. and 9 p.m. Pacific time. Figure 2 shows that an auction ending in primetime is associated with a small but noticeable right shift in the distribution of the number of observed bidders (top pane); and with a (quite small but still noticeable) right shift in the distribution of normalized transaction price (bottom pane).

To be a valid participation shifter, "primetime" should affect transaction prices only through the number of bidders. Because $N$ is not observable, this is difficult to check directly; but we can verify that "primetime" does not appear to affect prices once we control for the number of *observed* bidders. Table 1 shows that when we regress normalized transaction price on an auction's primetime status, we get a positive and significant coefficient; but when we include dummy variables for each value of $N_{\text{obs}}$, the coefficient on the primetime variable becomes small and not statistically significant. (Table 1 gives just the coefficient on prime-time; Table 8 in the Appendix shows the full regression result.) This at least supports the possibility that whether an auction ends in primetime might affect participation, but not the price conditional on participation (hence not valuations). Primetime status also conveniently splits

FIGURE 2

EMPIRICAL DISTRIBUTION OF OBSERVED BIDDERS AND NORMALIZED PRICES
[Color figure can be viewed at wileyonlinelibrary.com]




Note: top pane is empirical probability mass of number of observed bidders by primetime status; bottom pane is kernel density estimate of distribution of (transaction price) over (predicted price), by primetime status.

**TABLE 1 Primetime Predicts Transaction Price, but Only Through Participation**

| | *Dependent Variable*: Normalized log Transaction Price (1) | (2) |
|---|---|---|
| Primetime | $0.025^{***}$ | 0.005 |
| | (0.008) | (0.007) |
| Controls for observed $N$? | NO | YES |

Note: Robust standard errors in parentheses. $^{*}$p <0.1; $^{**}$p <0.05; $^{***}$p <0.01.

**TABLE 2 Summary Statistics by Primetime Status**

| | Non-Primetime | | | | Primetime | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum |
| Miles | 85,394 | 68,263 | 1.00 | 500,000 | 96,868 | 70,101 | 1.00 | 500,000 |
| Number of photos | 16.66 | 9.79 | 1.00 | 75.00 | 18.54 | 11.49 | 1.00 | 105.00 |
| Number of options | 6.72 | 4.95 | 0.00 | 20.00 | 6.73 | 4.87 | 0.00 | 21.00 |
| Feedback score | 131.47 | 598.85 | 1.00 | 27,575 | 177.84 | 481.87 | 1.00 | 11,552 |
| Negative feedback | 1.49 | 4.10 | 0.00 | 50.00 | 1.30 | 3.29 | 0.00 | 42.90 |
| Book value ($) | 10,421 | 8,182 | 889 | 45,097 | 9,247 | 7,200 | 889 | 45,097 |
| Transaction price ($) | 10,609 | 9,503 | 102 | 80,600 | 9,324 | 8,337 | 102 | 75,301 |
| Observed bidders | 8.78 | 3.31 | 2.00 | 22.00 | 9.53 | 3.74 | 2.00 | 22.00 |
| Observations | 7,832 | | | | 7,592 | | | |

our sample almost exactly in half. We thus proceed with primetime status as our participation shifter.[18]

Table 2 gives some comparative statics on auction covariates, broken down by primetime status. Whether an auction ends in primetime is not highly correlated with any of the covariates used in the normalization regression.

□ **Estimation.** As noted above, we estimate our model using primetime status as a participation shifter. We first posit an entry model (below), and estimate the distribution of $N$ for primetime and non-primetime auctions from the number of observed bidders in each. We then use these estimates, along with the observed normalized transaction prices, to estimate the remaining primitives, the distributions of $\theta$ and $\epsilon_i$.[19] Rather than the additive structure presented above, for the empirical application, we instead use the multiplicative structure

$$v_i = e^{z'\delta}\theta\epsilon_i, \tag{3}$$

where $z$ is a vector of observable covariates and $\delta$ a vector of coefficients. The translation between the two models is straightforward: as $\delta$ was already estimated in the normalization regression, we can define each bidder's normalized log valuation

$$\log \tilde{v}_i = \log v_i - z'\hat{\delta}, \tag{4}$$

and the normalized log-transaction price

$$\log \widetilde{T} = \log T - z'\hat{\delta}, \tag{5}$$

[18] Results using auction length (how many days the auction ran for) as the participation shifter are similar.

[19] In a sense, working in two steps like this does not "use all the data": we make use of the marginal distributions of $N_{\text{obs}}$ for each $X$, and the marginal distributions of $T$ for each $X$, but we do not use the available information on the correlation between $N_{\text{obs}}$ and $T$ for each $X$. However, working in two steps greatly simplifies estimation.

for each auction. Aside from the difference between the estimated parameter $\hat{\delta}$ and the unknown true value $\delta$,[20] $\log \widetilde{v}_i = \log \theta + \log \epsilon_i$, and $\log \widetilde{T}$ is the second highest of the normalized log valuations; Theorem 2 therefore says that the distributions $f_{\log \theta}$ and $f_{\log \epsilon}$ of $\log \theta$ and $\log \epsilon_i$ are identified from observed distributions $f_{\log \widetilde{T}}(\cdot|X)$ of log normalized transaction prices for two values of the participation shifter.

*Estimating the distribution of N.* Let $X \in \{p, n\}$ denote the type of auction, where $X = p$ indicates primetime and $X = n$ non-primetime. Our first step is to estimate the two distributions $\mathbf{p}(\cdot|X = p)$ and $\mathbf{p}(\cdot|X = n)$ based on an entry model and the observed distributions of the number of observed bidders $N_{\text{obs}}$. By Theorem 3, these distributions are nonparametrically identified (given the correct entry model), but we estimate them parametrically. Specifically, we assume they follow a negative binomial distribution, truncated by dropping realizations of 0 and 1.[21] Thus, we need only estimate four parameters, $p_p$, $r_p$, $p_n$, and $r_n$, using the observed distributions of $N_{\text{obs}}$ given $X$ and an entry model.

The entry model we have chosen is as follows. An auction consists of a primetime status $X$, a vector of observed covariates $z$, and an unobserved characteristic $\theta$. The number of potential bidders is drawn according to the distribution $\mathbf{p}(\cdot|X)$. These $N$ potential bidders then arrive sequentially, each receiving an independent idiosyncratic valuation term $\epsilon_i$. On arrival, a bidder sees the existing standing high bid $B$, which is the second-highest bid submitted so far (or the minimum opening bid, whichever is higher). Let $V$ be the valuation of the arriving bidder, which is equal to $e^{z'\delta}\theta\epsilon_i$. If $V < B$, she does not bid. If $V > B$, she immediately submits a proxy bid somewhere in the interval $[B, V]$, randomly drawn on that interval according to the interim distribution of $e^{z'\delta}\theta\epsilon_i$ conditional on the realization of $\theta$ and conditional on $e^{z'\delta}\theta\epsilon_i \in [B, V]$.[22] Any bidder whose value is still above the standing high bid as the end of the auction approaches, submits a second proxy bid equal to her valuation just before the auction ends, so that the transaction price is equal to the second-highest valuation.[23] Though somewhat arbitrary, this formulation of the entry model makes the relationship between the distributions of $N$ and $N_{\text{obs}}$ independent of the distributions of $\epsilon$ and $\theta$, allowing us to simulate the entry process as a first step, rather than needing to estimate the distributions of $N$ and the value distributions simultaneously. Our choice of this entry model over another alternative is discussed in the Appendix.

We calculate the probabilities $\Pr(N_{\text{obs}} = k|N = j)$ by simulation. Specifically, for each $j \in \{2, 3, \ldots, 1000\}$, we simulate the entry model one million times, and estimate $\Pr(N_{\text{obs}} = k|N = j)$ as the observed frequency. (We stop at $N = 1000$ because for reasonable values of $(p, r)$, $\Pr(N = n|p, r)$ drops to within machine accuracy of 0 before $n = 1000$.) Given these simulated probabilities, we then estimate $(p_p, r_p)$ and $(p_n, r_n)$ via maximum likelihood, that is, taking $(p_x, r_x)$ to be

$$\arg\max_{p,r} \frac{1}{M_x} \sum_{m=1}^{M_x} \ln \left( \sum_{n=2}^{\infty} \Pr(N = n|p, r) \widetilde{\Pr}(N_{obs} = n_{m,x}|N = n) \right)$$

[20] The use of $\hat{\delta}$ for $\delta$ will not affect consistent estimation of the distributions $f_{\log \theta}$ and $f_{\log \epsilon}$, only inference. To simplify exposition, except where we discuss inference, we proceed as if we knew the true value $\delta$.

[21] The untruncated negative binomial distribution with parameters $p \in (0, 1)$ and $r > 0$ gives $\Pr(N = n) = \frac{\Gamma(r+n)}{n!\Gamma(r)} p^n (1 - p)^r$ for $n \geq 0$, so the truncated gives $\Pr(N = n) = \frac{1}{1-(1-p)^r - rp(1-p)^r} \frac{\Gamma(r+n)}{n!\Gamma(r)} p^n (1 - p)^r$ for $n \geq 2$.

[22] To put it another way, she bids $e^{z'\delta}\theta F_\epsilon^{-1}(u)$, where $u$ is a random variable drawn uniformly from the interval $[\underline{u}, \overline{u}]$, where $[e^{z'\delta}\theta F_\epsilon^{-1}(\underline{u}), e^{z'\delta}\theta F_\epsilon^{-1}(\overline{u})] = [B, V]$.

[23] Thus, our entry model implies that there should be both substantial bidding throughout the auction and near the end, and that some bidders bid more than once. This is consistent with our data: on average, auctions receive 2.4 bids per observed bidder, 12% of observed bidders in an auction place a bid in the last hour, and 6% place a bid in the last 5 minutes; the highest losing bid arrives in the last hour in 50% of our auctions. Although our entry model takes a stand on how a bidder will bid conditional on her valuation, we do not use the entry model to recover valuations from bids; the entry model is only used to establish the relationship between $N$ and $N_{\text{obs}}$, to allow us to recover the distribution of $N$ from the observed distribution of $N_{\text{obs}}$.

$$= \arg\max_{p,r} \frac{1}{M_x} \sum_{m=1}^{M_x} \ln \left( \sum_{n=2}^{\infty} \frac{1}{1-(1-p)^r - rp(1-p)^r} \frac{\Gamma(r+n)}{n!\Gamma(r)} p^n (1-p)^r \widetilde{\Pr}(N_{obs} = n_{m,x} | N = n) \right), \tag{6}$$

where $n_{m,x}$ is the observed number of bidders in the $m$th auction (out of $M_x$) of primetime status $x \in \{p, n\}$ and $\widetilde{\Pr}$ refers to the simulated probabilities calculated in the previous step.

*Estimating the distributions of $\theta$ and $\epsilon_i$.* Given Theorem 2, once $(p_p, r_p)$ and $(p_n, r_n)$ are known, the densities $f_{\log\epsilon}$ and $f_{\log\theta}$ can be consistently estimated via semi-nonparametric maximum likelihood (Gallant and Nychka (1987)). In this estimation method, the density of a random variable is flexibly approximated by a finite dimensional parametric density, or sieve, where the number of parameters grows with sample size. Let $\xi_0 = (f_{\log\theta}, f_{\log\epsilon}) \in \Xi$, where $\Xi$ is the parameter space. Then, instead of searching over the entire set of density functions $\Xi$, semi-nonparametric maximum likelihood estimation searches over a finite-dimensional sieve space of $\Xi$, denoted $\Xi_M$, that depends on the sample size $M$. We can therefore estimate $\xi_0$ as

$$\widehat{\xi} = \arg\max_{\xi\in\Xi_M} \frac{1}{M} \sum_{m=1}^{M} \ell(\xi, \widehat{p}_m, \widehat{r}_m, \log \widetilde{T}_m), \tag{7}$$

where $\ell(\xi, \widehat{p}_m, \widehat{r}_m, \log \widetilde{T}_m) = \log[f_{\log \widetilde{T}}(\log \widetilde{T}_m; \xi, \widehat{p}_m, \widehat{r}_m)]$ is the log likelihood of the observed normalized log-transaction price $\log \widetilde{T}_m$ given $\xi$ and $(\widehat{p}_m, \widehat{r}_m)$ is the plug-in estimate of the negative binomial distribution for auction $m$.

The assumption that $N$ follows a (truncated) negative binomial distribution simplifies this significantly, by eliminating the need to sum over different possible realizations of $N$. For a general distribution of $N$, the density function of normalized log-transaction price $\log \widetilde{T}$ would be

$$f_{\log \widetilde{T}}(t|x) = \sum_{n=2}^{\infty} \Pr(N = n | X = x) f_{\log \widetilde{T}}(t | N = n), \tag{8}$$

where $\log \widetilde{T}$ is the sum of $\log\theta$ and the second highest of $N$ independent draws of $\log\epsilon_i$, giving

$$f_{\log \widetilde{T}}(t|N = n) = n(n-1) \int_{-\infty}^{\infty} f_{\log\theta}(t-s) F_{\log\epsilon}^{n-2}(s)(1 - F_{\log\epsilon}(s)) f_{\log\epsilon}(s) ds. \tag{9}$$

The presence of an integral within an infinite sum in the log-likelihood function could make the problem very computationally demanding, but given the particular structure of our problem, we are able to simplify things to eliminate the sum. In particular, when $N$ follows a truncated negative binomial distribution with parameters $(p, r)$, calculations in the Appendix show that[24]

$$f_{\log \widetilde{T}}(t) = \frac{(r+1)rp^2(1-p)^r}{1-(1-p)^r - rp(1-p)^r} \int_{-\infty}^{\infty} f_{\log\theta}(t-s)(1 - F_{\log\epsilon}(s))(1 - pF_{\log\epsilon}(s))^{-(r+2)} f_{\log\epsilon}(s) ds. \tag{10}$$

For estimation of $f_{\log\epsilon}$ and $f_{\log\theta}$, we choose the set of orthonormal Hermite polynomials as our sieve space. Thus, the density of a random variable $Z$ is approximated with the functional form

$$f_Z(z) \approx \frac{1}{\sqrt{2\pi\sigma_Z^2}} \left( \sum_{k=0}^{K_M} \beta_{Z,k} H_k\left(\frac{z-\mu_Z}{\sigma_Z}\right) \right)^2 e^{-\frac{1}{2}\left(\frac{z-\mu_Z}{\sigma_Z}\right)^2}, \tag{11}$$

[24] A similar simplification also works when $N$ follows a truncated Poisson distribution. Both simplifications are shown in the Appendix, as Theorem 5.

where $K_M$ is a smoothing parameter that increases with sample size and $\mu_Z$, $\sigma_Z$, and $\beta_Z$ are parameters to be estimated.[25,26] $H_k(z)$ corresponds to the sequence of orthonormal Hermite polynomials defined by the recurrence relation $H_0(z) = 1$, $H_1(z) = z$, and

$$H_k(z) = \frac{1}{\sqrt{k}}\Big(zH_{k-1}(z) - \sqrt{k-1}H_{k-2}(z)\Big), \tag{12}$$

for $k > 1$. With these polynomials, the vector $\beta_Z$ must satisfy $\sum_{k=0}^{K_M} \beta_{Z,k}^2 = 1$, so that $f_Z(z)$ is a density.[27] We explicitly enforce this constraint in the optimization routine with $\beta_{Z,0} = \sqrt{1 - \sum_{k=1}^{K_M} \beta_{Z,k}^2}$. This is without loss of generality because $f_Z(z; \beta_Z, \mu_Z, \sigma_Z) = f_Z(z; -\beta_Z, \mu_Z, \sigma_Z)$.

We proceed with a model where the densities of $\log\epsilon$ and $\log\theta$ are approximated with a Hermite polynomial of degree three, to provide enough flexibility for estimating these distributions given our sample size. In addition, we normalize $E[\log\theta]$ to 0 because (as noted above in footnote 11) the means of $\log\epsilon$ and $\log\theta$ are not separately identified. We enforce this constraint explicitly in the optimization routine, by setting $\mu_{\log\theta} = -\sigma_{\log\theta}\int_{-\infty}^{\infty} \frac{s}{\sqrt{2\pi}}(\sum_{k=0}^{K_M} \beta_{\log\theta,k} H_k(s))^2 e^{-\frac{1}{2}s^2} ds$. With this model specification, the convolution integral in the likelihood function runs from minus infinity to plus infinity because the densities have full support, so we compute the convolution integral in the likelihood function using a Gauss–Hermite quadrature rule.

We should note that the empirical specification we have chosen does not satisfy all the assumptions of our identification results above. Specifically, our formulation for $f_{\log\theta}$ and $f_{\log\epsilon}$ has full support on $\mathbb{R}$, whereas Theorems 2 and 4 assume the distributions are bounded below. Similarly, our parameterization for the distribution of $N$ has unbounded support, whereas Theorems 3 and 4 assume the support is bounded above; and we do not restrict the two estimated distributions of $N|X$ to satisfy the MLRP. Thus, if this parameterization were the true model, our results as stated above would not guarantee it was nonparametrically identified. We do not see these gaps between our functional form assumptions and our identification results as being problematic, for the following reasons. First, our density functions have unbounded support in theory, but in practice, the densities will be within machine accuracy of 0 outside of a bounded range, so for purposes of estimation, they effectively have bounded support.[28] As for the MLRP, as noted above, when the distributions of $\theta$ and $\epsilon$ are real analytic (as they are in our empirical specification) rather than piecewise real analytic, Theorems 2 and 4 do not require the MLRP, and nonparametric identification holds as long as the distribution of $N$ varies in any way with $X$. We therefore do not see the gaps between our empirical formulation and the conditions of Theorem 4 to be a problem for identification, and view our parameterizations as expedient choices that let us fit the data well with relatively few parameters.[29]

Though we estimate our model in three steps—first estimating the parameters $\delta$ of the normalization regression, then the parameters $\lambda = (p_p, r_p, p_n, r_n)$ of the distribution of $N$ given $X$, and finally the parameters $\alpha = (\mu_{\log\epsilon}, \sigma_{\log\epsilon}, \beta_{\log\epsilon,1}, \beta_{\log\epsilon,2}, \beta_{\log\epsilon,3}, \sigma_{\log\theta}, \beta_{\log\theta,1}, \beta_{\log\theta,2}, \beta_{\log\theta,3})$

[25] Consistent estimation of $f_Z$ requires that $K_M \to \infty$ as sample size $M \to \infty$.

[26] Note that $\mu_Z$ and $\sigma_Z$ are not necessary for consistent estimation, but they improve the numerical performance of the estimator.

[27] $\sum_k \beta_{Z,k}^2 = 1$ ensures that $f_Z$ integrates to 1; given its functional form, $f_Z$ is always non-negative.

[28] Although the simplification in (10) means we are using the exact (estimated) distributions of $N$ rather than approximated ones when we estimate $f_{\log\theta}$ and $f_{\log\epsilon}$, we are using numerically approximated distributions of $N$ when we estimate the parameters of the distributions of $N$.

[29] We could alternatively estimate truncated distributions, but this would greatly complicate estimation with little gain. Equation (10), in particular, shows how much using the "full" negative binomial distribution (truncated below but not above) for $N$ simplifies estimation.

**TABLE 3 MLE Results for Entry Model**

| | Primetime Auctions | Non-Primetime Auctions |
|---|---|---|
| $p_x$ (s.e.) | 0.876 (0.004) | 0.821 (0.005) |
| $r_x$ (s.e.) | 3.023 (0.089) | 3.848 (0.121) |
| Observations | 7,592 | 7,832 |
| −2 Log likelihood | 41,035.82 | 40,506.32 |

of the value distributions $f_{\log\theta}$ and $f_{\log\epsilon}$—our three-step semi-nonparametric MLE procedure is numerically equivalent to parametric GMM estimation using the moments

$$E\left[\begin{bmatrix} z \\ I \end{bmatrix}(\log T - z'\delta - I'\eta)\right] = 0, \qquad E\left[\frac{d}{d\lambda}\log\left(\Pr\left(N_{obs} = n|\lambda\right)\right)\right] = 0,$$
$$E\left[\frac{d}{d\alpha}\log\left(f_{\log\widetilde{T}}(\log T - z'\delta, \alpha, \lambda)\right)\right] = 0. \tag{13}$$

This means that under the appropriate regularity conditions, one could perform inference on functionals of $f_{\log\theta}$ and $f_{\log\epsilon}$ as if they were estimated via GMM estimation of a correctly specified parametric model.[30]

## □ Results.

*Entry model.* As discussed above, we assume that the (unobserved) number of bidders follows a negative binomial distribution, truncated by dropping realizations of 0 and 1, and that the number of observed bidders is then generated from the number of bidders via the "entry game" described above. We estimated the entry model via maximum likelihood estimation given the distribution of the number of observed bidders; Table 3 shows the results of this estimation. A likelihood ratio test of whether the two parameters are the same across primetime and non-primetime auctions returns a test statistic of 217.5; the critical value to reject the hypothesis at the 0.1% level is 13.8, so we can easily reject the hypothesis that the distribution of $N$ is the same. Figure 3 shows the fit of the entry model, comparing the distribution of the number of observed bids across auctions in the data (solid blue curves) and the corresponding distribution generated by the model (dashed red curves). With two parameters, we are able to fit the observed distributions quite well. The two estimated distributions of $N$ given $X$ do not satisfy the MLRP; as noted above, this is not a problem for identification of $f_{\log\theta}$ and $f_{\log\epsilon}$ because the distributions are real analytic.

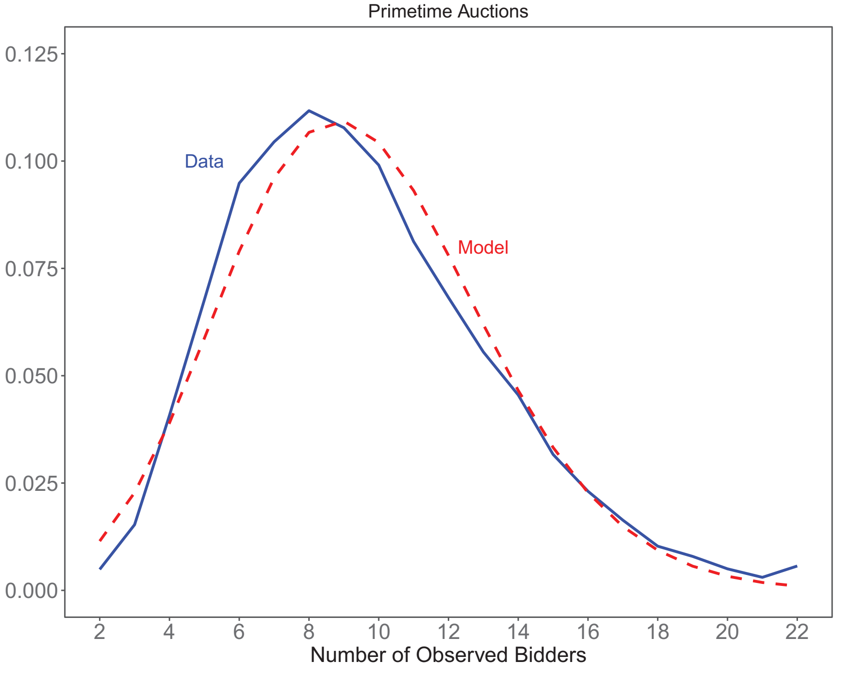
*Valuations.* We plug in our estimated parameters for the entry model and estimate the distributions of $\theta$ and $\epsilon$. We estimate the densities of both $\log\theta$ and $\log\epsilon$ as third-order Hermite polynomials. The estimated density functions and CDFs are shown in Figure 4, with $\log\theta$ on the left and $\log\epsilon$ on the right. The estimated density of $\log\epsilon$ is bimodal, suggesting there appears to be two distinct "types" of bidders, high valuation and low valuation.

*Model fit.* Figure 5 shows model fit, comparing the probability density functions of transaction prices implied by our estimated model (solid blue curves) to the smoothed densities of the actual transaction price data (using kernel density estimation, dashed red curves). The top pane contains non-primetime auctions, and the bottom pane primetime auctions. The graph shows that the sieve maximum likelihood estimator fits the density of transaction prices quite well for both samples.

[30] See Chen and Liao (2014) for regularity conditions for non-parametric models, or Ackerberg, Chen, and Hahn (2012) for semi-parametric models.

FIGURE 3

PROBABILITY MASS FUNCTION OF NUMBER OF OBSERVED BIDDERS, ESTIMATED MODEL VERSUS OBSERVED [Color figure can be viewed at wileyonlinelibrary.com]



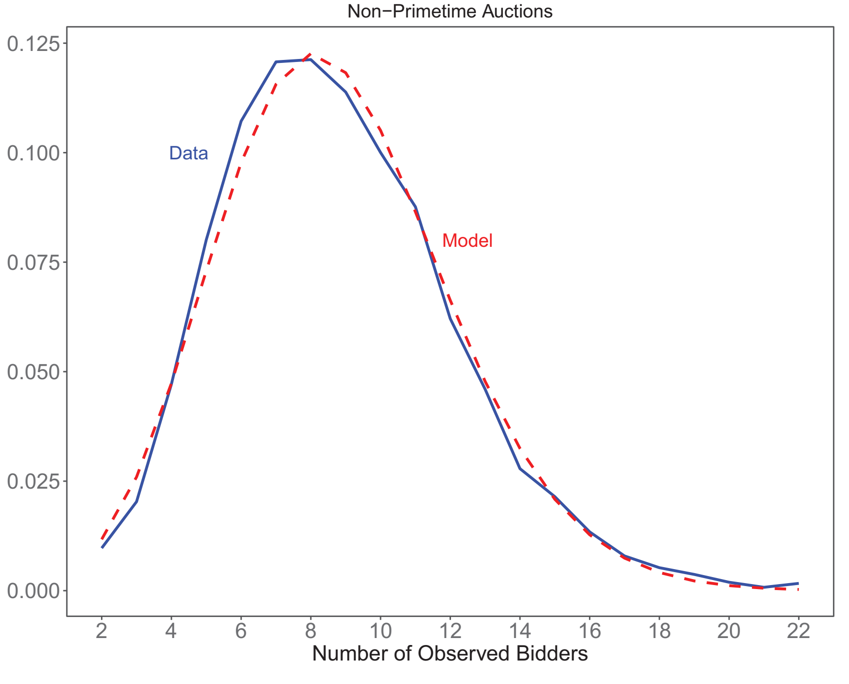
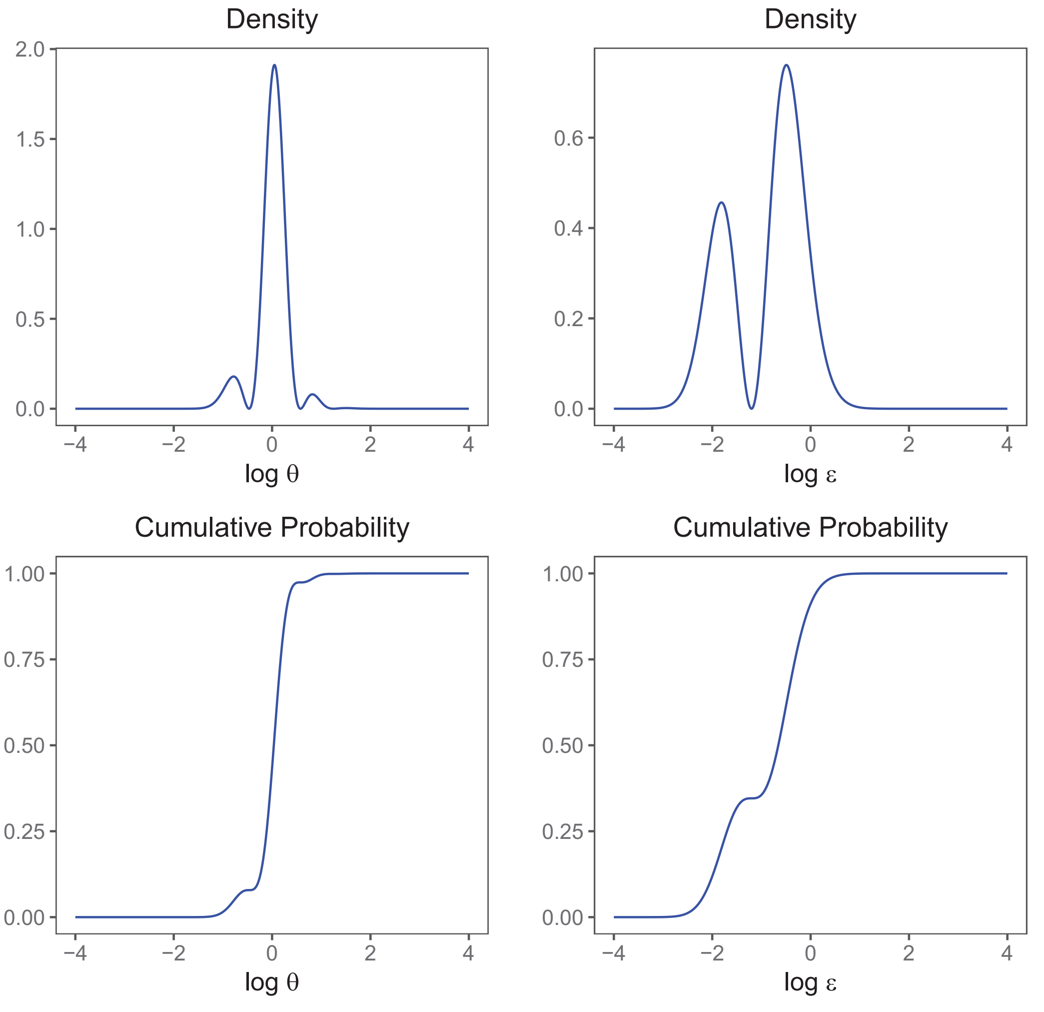



FIGURE 4

ESTIMATED DISTRIBUTIONS OF $\log\theta$ AND $\log\epsilon$ [Color figure can be viewed at wileyonlinelibrary.com]




□ **Analysis.**

*Variance decomposition.* As noted above, we model the valuation of bidder $i$ in a particular auction as $v_i = e^{z'\delta}\theta\epsilon_i$, where $z$ is a vector of observable characteristics of the auction, $\delta$ a vector of parameters, $\theta$ the unobserved term corresponding to the auction, and $\epsilon_i$ bidder $i$'s idiosyncratic value term; this leads to a transaction price of $T = e^{z'\delta}\theta\epsilon^{(2)}$, where $\epsilon^{(2)}$ is the second-highest $\epsilon_i$ among the bidders present. Taking logs, we can write $\log T = z'\delta + \log\theta + \log\epsilon^{(2)}$; by assumption, $(z, \theta, \epsilon_i)$ are mutually independent, so we can decompose the variance of log-transaction prices as

$$\mathrm{Var}\log T = \mathrm{Var}(z'\delta) + \mathrm{Var}\log\theta + \mathrm{Var}\log\epsilon^{(2)}. \tag{14}$$

The variance of $\log\epsilon^{(2)}$ depends on $N$, of course, but taking the expectation over the distributions of $N$ implied by our estimates, it is a straightforward calculation to decompose the variance of transaction prices into its three components: variation in observables, variation in unobservables, and variation in idiosyncratic valuations. Table 4 shows this decomposition: the fraction of the variance of log-transaction prices attributable to variance in $z'\delta$, $\log\theta$, and $\log\epsilon_i$, respectively. The

FIGURE 5

OBSERVED AND ESTIMATED PROBABILITY DENSITIES OF NORMALIZED LOG-TRANSACTION PRICE
[Color figure can be viewed at wileyonlinelibrary.com]

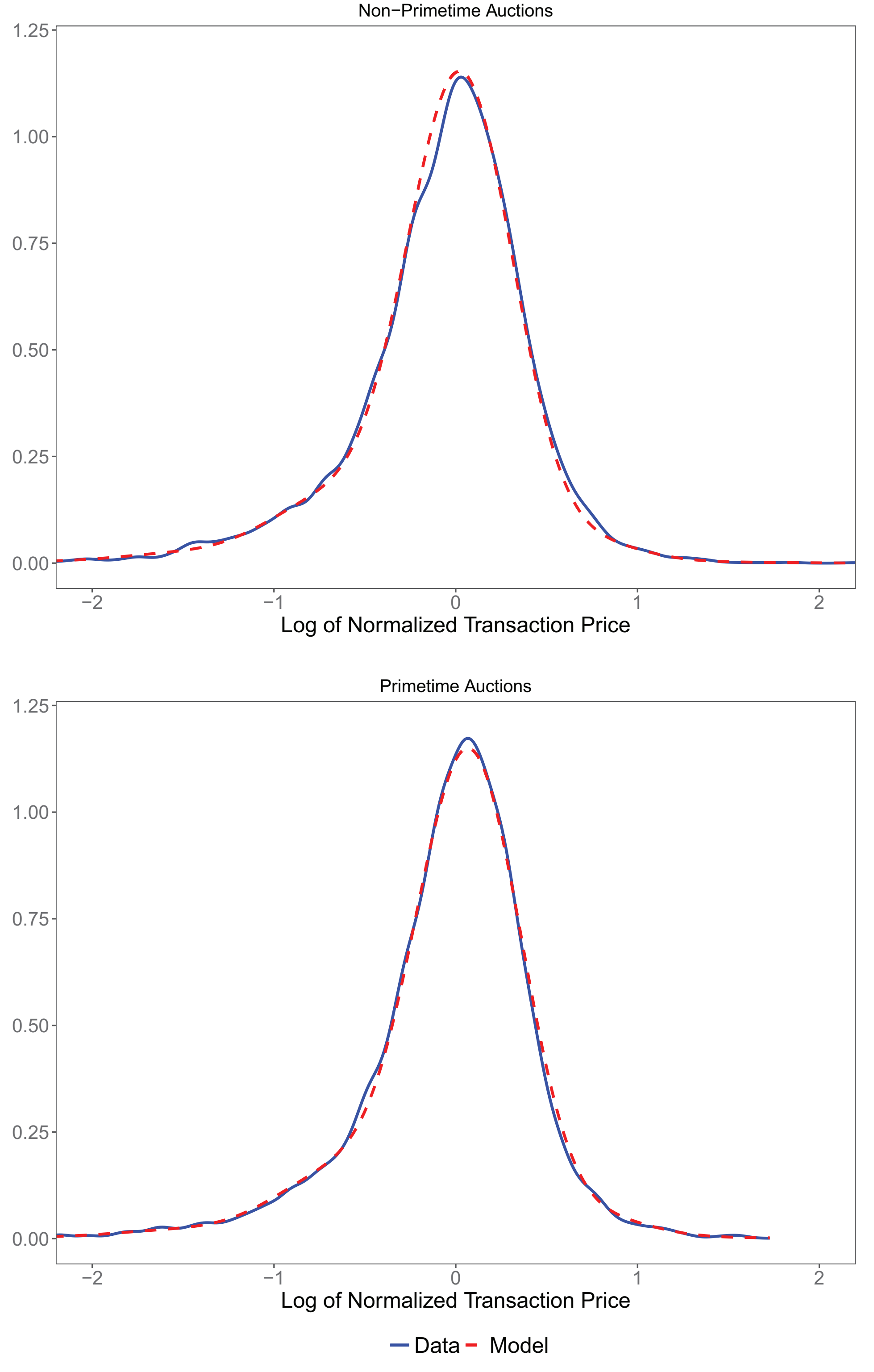


**TABLE 4 Decomposing the Variance of Log-Transaction Price**

| Auction type | Observations | $\frac{\text{Var}(z'\delta)}{\text{Var}\log T}$ | $\frac{\text{Var}\log\theta}{\text{Var}\log T}$ | $\frac{\text{Var}\log\epsilon^{(2)}}{\text{Var}\log T}$ | $\frac{\text{Var}\log\theta}{\text{Var}\log\theta+\text{Var}\log\epsilon^{(2)}}$ |
|---|---|---|---|---|---|
| Non-primetime | 7,832 | 83.5% | 10.9% | 5.6% | 66.0% |
| Primetime | 7,592 | 82.1% | 12.1% | 5.8% | 67.4% |
| Combined | 15,424 | 82.9% | 11.4% | 5.7% | 66.7% |

final column shows the fraction of the variance not explained by observables that is attributed to unobserved heterogeneity. The main takeaway is twofold:

- *Observables explain a lot*. 83% of the variance in log-transaction prices is explained by observable covariates.
- *Unobserved heterogeneity is important*. Of the variance not explained by observables, 67% is from unobserved heterogeneity, only 33% from variation in bidder-specific private values.

A few points are worth making about the variance decomposition in Table 4. First, of course, this depends on having correctly specified the normalization regression. Any misspecification in controlling for observables would leave a residual which would be "picked up" as part of $\theta$. The estimate of $f_{\log\epsilon}$ (the key to calculating bidder surplus below and other counterfactuals) would still be correct, but we would be misattributing to $\theta$ some of the variation that is actually due to observables. (For example, if the true model of valuations were $v_i = e^{g(z)}\theta\epsilon_i$ and we erroneously used the linear model $v_i = e^{z'\delta}\theta\epsilon_i$, what we measured as $f_{\log\theta}$ would actually be the distribution of $\log\theta + (g(z) - z'\delta)$, and the variance of $\log\theta$ in Table 4 would be overstated.) Second, a key insight of Lewis (2011) is that sellers control how much information to disclose about the car being sold, and adverse selection is mitigated via more detailed listings, because a detailed description forms a contract to deliver a car matching the description. In our setting, this would correspond to bidders receiving only noisy signals about $\theta$, with the noisiness of the signal depending on $z$. A common values model like this would require a different empirical approach than ours. Even with private values, if the distribution of $\theta$ (and/or of $\epsilon$) varied with $z$, our nonparametric identification results would still be valid—they hold separately for each value of $z$; but our parametric specification would not be the right one, and one would want to proceed with a formulation where parameters of $f_{\log\theta}$ and $f_{\log\epsilon}$ were explicitly allowed to depend on $z$. Finally, if the "idiosyncratic" part of bidder valuations $\epsilon_i$ were correlated, then we would be interpreting that correlation as being due to unobserved heterogeneity (and picking it up as part of $\theta$), although this is primarily a question of interpretation—we would be decomposing correlated $\epsilon_i$ into a part that was common across bidders and a part that was independent across bidders, and interpreting the former as $\theta$.

*Bidder surplus*. Table 4 suggests that unobserved heterogeneity is present in the auctions in our data, and of a magnitude that could be economically important. To demonstrate its importance for accurate estimation and analysis of counterfactuals, we also estimate a model of independent private values without unobserved heterogeneity,[31] to see how the implications of the two sets of estimates differ. Table 5 shows the two resulting estimates of average bidder surplus per auction. (This is for the combined sample of primetime and non-primetime auctions, but the numbers are comparable for the two subsamples.) At both the mean and the median, estimating a model without unobserved heterogeneity gives an estimate of bidder surplus about 3.3 times the value estimated with unobserved heterogeneity.

[31] That is, maintaining the same estimates from the normalization regression and entry model, we assume $v_i = e^{z'\delta}\epsilon_i$ and re-estimate $f_{\log\epsilon}$ without $\theta$, replacing the likelihood function (10) with

$$f_{\log\widetilde{T}}(t|x) = \frac{(r+1)rp^2(1-p)^r}{1-(1-p)^r-rp(1-p)^r}(1 - F_{\log\epsilon}(t))(1 - pF_{\log\epsilon}(t))^{-(r+2)}f_{\log\epsilon}(t) \tag{15}$$

with $(p, r)$ the entry model coefficient estimates for auctions of primetime status $x$.

**TABLE 5 Estimates of Bidder Surplus**

| | Our estimate | Estimated without $\theta$ |
|---|---|---|
| Mean | 3,028 | 10,065 |
| 25th Pct | 985 | 3,292 |
| Median | 2,155 | 7,176 |
| 75th Pct | 3,927 | 13,066 |

The intuition for the direction of this result is straightforward. In order to match the variation in transaction prices, an empirical model without unobserved heterogeneity would, by necessity, ascribe greater variance to idiosyncratic bidder tastes $\epsilon_i$. But bidder surplus is the difference between the highest and second highest of the $\epsilon_i$, and is thus increasing in the variance in $\epsilon_i$. Thus, it is inevitable that a model without unobserved heterogeneity would lead to a greater estimate of bidder surplus. What is striking about Table 5 is the magnitude: that ignoring the effect of $\theta$ would lead to an estimate of bidder surplus more than three times as large.

## 5. Conclusion

■ We have shown that a model of independent private values with separable unobserved heterogeneity is point identified from standard English auction data if there is any exogenous variation in the number of bidders, and that imperfect observability of the number of bidders can be overcome if one has access to a "participation shifter" and is willing to posit an entry model.

In our empirical application, the participation shifter is binary, so our model is just-identified. However, with a participation shifter taking more values, our model would be over-identified, and we could envision extending it in several directions. First, we could potentially develop an econometric test of the exogeneity of $N$—or really a joint test of all our key assumptions, including exogenous $N$ and additive separability. Second, a more flexible entry model could potentially be used, with the parameters of the entry model being estimated in parallel with those of the value distributions. Third, we rely on a model of additively or multiplicatively separable unobserved heterogeneity, which is not without loss. Hu, McAdams, and Shum (2013) for first-price auctions, and Mbakop (2017) and Luo and Xiao (2019) for English auctions, have shown that with enough observed bids in each auction, one can move beyond the additively separable model and identify a more general non-separable model of valuations with unobserved heterogeneity; with sufficiently many different realizations of the participation shifter (or of $N$ in settings where it is observed), similar results might follow in our setting as well.

In our application, we have used a simple (linear) model to control for observable covariates. One might wonder whether such a model were misspecified, and whether a "correct" model of covariates would account for more of the price variation, meaning we would be overstating the importance of truly unobserved heterogeneity. Indeed, as we noted earlier, Bodoh-Creed, Boehnke, and Hickman (2018a) find that by applying machine learning techniques to an extremely rich dataset, they could triple the amount of price variation explained by observables, relative to linear analysis of a more standard set of covariates. One might even conjecture that with careful enough conditioning on a rich enough set of observables, allowing for unobserved heterogeneity might be unnecessary, and an IPV model sufficient. To this, we have two responses. First, what we do is consistent with empirical approaches taken by many researchers in many settings—the vast majority of empirical auction articles do *not* apply cutting-edge machine learning techniques to unusually rich datasets. But more importantly, accounting for unobserved heterogeneity in estimation is consistent, even desirable, with a misspecified model of covariates: as noted earlier, in addition to unobserved heterogeneity, the $\theta$ term captures the residual variation due to observables that were omitted or mismodeled. With mismodeled observables, our model would of course not be perfectly specified—what we measured as $\theta$ would not be independent of observables, so the variance decomposition we showed in Table 4 would not be correct. Nonetheless, the estimates of $f_{\log \epsilon}$—the key to estimating bidder surplus and other counterfactuals—would still be correct, and quite different from the results if we were ascribing to $\epsilon$ this residual variation due to misspecification.

We conclude with the point we started with: unobserved auction-level heterogeneity is economically meaningful in many auction settings, including ours; ignoring it in estimation can lead to misleading conclusions, so allowing for it in estimation is important. This article is intended as a step in that direction.

## Appendix A

□ **Normalization regression and included versus excluded sample.** Table 6 shows the bid normalization regression estimates. Specification (1) includes fixed effects for the number of observed bidders and Specification (2) does not. A comparison of the two specifications indicates that, in this case, if we drop $I$ from the normalization regression, we get nearly identical estimates for $\delta$. Furthermore, normalized prices are also nearly identical. Overall, the results in Table 6 are similar to the ones presented in Lewis (2011), Table 2, Specification (6).

Prior to estimation, we dropped auctions with a reserve price higher than 20% of the predicted transaction price based on observables. Table 7 shows summary statistics for the estimation sample (the auctions included) and the excluded sample (the auctions we dropped). Auctions in the estimation sample have on average about four more bidders than auctions in the excluded sample. This is consistent with the fact that the number of observed bidders in the excluded sample is more likely to be censored due to a binding reserve price. The summary statistics also indicate that, compared with the estimation sample, cars in the excluded sample have 13% fewer miles, have a book value 15% higher, and are sold at a price 27% higher (which is more than 15%, possibly due to the use of a reserve price), on average. Although there are some observable differences between the estimation and excluded samples, the bid normalization regression cannot be affected by them because it was run on the whole sample.

**TABLE 6 Bid Normalization Regression Table**

| | Log Transaction Price | |
|---|---|---|
| | (1) | (2) |
| Log miles | $-0.183^{***}$ (0.009) | $-0.183^{***}$ (0.009) |
| Number of photos | $0.011^{***}$ (0.001) | $0.011^{***}$ (0.001) |
| Photos squared/100 | $-0.012^{***}$ (0.002) | $-0.012^{***}$ (0.002) |
| Number of options | $0.009^{***}$ (0.001) | $0.010^{***}$ (0.001) |
| Log feedback | $-0.012^{***}$ (0.002) | $-0.010^{***}$ (0.002) |
| Negative feedback | $-0.003^{***}$ (0.001) | $-0.003^{***}$ (0.001) |
| Log-book value | $0.576^{***}$ (0.019) | $0.584^{***}$ (0.019) |
| Model/year/week FE | Yes | Yes |
| Number of observed bidders FE | Yes | No |
| $R^2$ | 0.815 | 0.814 |
| Observations | 26,780 | 26,780 |

Note: Robust standard errors in parentheses clustered by seller. $^{***}$Significant at the 1% level.

**TABLE 7 Summary Statistics for Estimation Sample and Excluded Sample**

| | Estimation Sample | | | | Excluded Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum |
| Miles | 91,042 | 69,409 | 1.00 | 500,000 | 79,249 | 67,425 | 2.00 | 500,000 |
| Number of photos | 17.58 | 10.70 | 1.00 | 105.00 | 16.62 | 9.63 | 1.00 | 75.00 |
| Number of options | 6.72 | 4.91 | 0.00 | 21.00 | 6.53 | 5.15 | 0.00 | 25.00 |
| Feedback score | 154.29 | 544.90 | 1.00 | 27,575 | 106.04 | 469.39 | 1.00 | 33,368 |
| Negative feedback | 1.40 | 3.72 | 0.00 | 50.00 | 1.57 | 4.13 | 0.00 | 42.90 |
| Book value ($) | 9,843 | 7,737 | 889 | 45,097 | 11,324 | 8,932 | 889 | 45,097 |
| Transaction price ($) | 9,977 | 8,971 | 102 | 80,600 | 12,699 | 10,990 | 296 | 78,100 |
| Observed bidders | 9.15 | 3.55 | 2.00 | 22.00 | 5.32 | 2.78 | 2.00 | 22.00 |
| Observations | | 15,424 | | | | 11,356 | | |

□ **Full regression results behind Table 1 in text.**

**TABLE 8 Primetime Affects Transaction Price, but Only Through Participation**

| | Normalized Log-Transaction Price | |
|---|---|---|
| | (1) | (2) |
| Primetime | 0.025*** (0.008) | 0.005 (0.007) |
| Dummy for $N_{\text{obs}} = 3$ | | 0.140 (0.099) |
| Dummy for $N_{\text{obs}} = 4$ | | 0.246*** (0.093) |
| Dummy for $N_{\text{obs}} = 5$ | | 0.379*** (0.091) |
| Dummy for $N_{\text{obs}} = 6$ | | 0.429*** (0.091) |
| Dummy for $N_{\text{obs}} = 7$ | | 0.479*** (0.090) |
| Dummy for $N_{\text{obs}} = 8$ | | 0.526*** (0.090) |
| Dummy for $N_{\text{obs}} = 9$ | | 0.555*** (0.090) |
| Dummy for $N_{\text{obs}} = 10$ | | 0.575*** (0.090) |
| Dummy for $N_{\text{obs}} = 11$ | | 0.594*** (0.090) |
| Dummy for $N_{\text{obs}} = 12$ | | 0.602*** (0.091) |
| Dummy for $N_{\text{obs}} = 13$ | | 0.636*** (0.091) |
| Dummy for $N_{\text{obs}} = 14$ | | 0.636*** (0.091) |
| Dummy for $N_{\text{obs}} = 15$ | | 0.649*** (0.091) |
| Dummy for $N_{\text{obs}} = 16$ | | 0.663*** (0.093) |
| Dummy for $N_{\text{obs}} = 17$ | | 0.684*** (0.093) |
| Dummy for $N_{\text{obs}} = 18$ | | 0.684*** (0.093) |
| Dummy for $N_{\text{obs}} = 19$ | | 0.658*** (0.096) |
| Dummy for $N_{\text{obs}} = 20$ | | 0.729*** (0.100) |
| Dummy for $N_{\text{obs}} = 21$ | | 0.754*** (0.102) |
| Dummy for $N_{\text{obs}} = 22$ | | 0.615*** (0.097) |
| Constant | −0.065*** (0.005) | −0.572*** (0.090) |
| Observations | 15,424 | 15,424 |
| $R^2$ | 0.001 | 0.067 |

Note: Robust standard errors in parentheses. *** Significant at the 1% level.

□ **Proof of Theorem 1.** *Outline of proof.* We want to show that there can be only a single pair of continuous, piecewise real analytic distributions $(f_\theta, f_\epsilon)$ that could generate the observed distribution of transaction prices for two different values of $N$. Suppose this were false, and there were two pairs of distributions, $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$, which generated the observed distributions $f_{T|n}$ and $f_{T|n'}$ of transaction prices given $N = n$ and $N = n'$, respectively. Because each transaction price is the realization of $\theta + \epsilon^{(2)}$, where $\epsilon^{(2)}$ is the second highest of the $N$ independently drawn $\epsilon_i$, standard order statistic and convolution results would require that

$$\int_0^t f_\theta(s)d\psi_n(F_\epsilon(t-s)) = f_{T|n}(t) = \int_0^t g_\theta(s)d\psi_n(G_\epsilon(t-s)),$$

and

$$\int_0^t f_\theta(s)d\psi_{n'}(F_\epsilon(t-s)) = f_{T|n'}(t) = \int_0^t g_\theta(s)d\psi_{n'}(G_\epsilon(t-s)), \tag{A1}$$

where $\psi_n(x) = nx^{n-1} - (n-1)x^n$.

For any density $f$, let $f^{(k)}$ denote its $k$th derivative, with $f^{(0)} = f$; at points where a density is not $k$ times differentiable, let $f^{(k)}$ denote its $k$th right derivative. (If $f$ is piecewise real analytic, then at any point $x$, it is infinitely many times right differentiable, and equal to its Taylor expansion based on right derivatives at $x$ on some neighborhood $[x, x+\delta)$.) To be consistent with our model, $f_{T|n}$ and $f_{T|n'}$ must have the same support; assume without loss that the bottom of the support is 0, and normalize 0 to be the bottom of the supports of the distributions of both $\theta$ and $\epsilon_i$. We will prove Theorem 1 in two steps:

(1) If (A1) holds, then $f_\theta^{(k)}(0) = g_\theta^{(k)}(0)$ and $f_\epsilon^{(k)}(0) = g_\epsilon^{(k)}(0)$ for $k = 0, 1, 2, \ldots$
Because all distributions are assumed to be piecewise real analytic, there exists a $\delta > 0$ such that $f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for $x \in [0, \delta)$.

(2) Fix $t^* > 0$, and suppose that the two pairs of distributions $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ begin to diverge at $t^*$: that is, the statement "$f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for all $x \le t$" holds for $t = t^*$, but not for any $t > t^*$.

Then for some $t$ just above $t^*$, it is impossible for both parts of (A1) to hold.

This establishes that there cannot be two different primitive distributions that would generate the same distributions of observables, and therefore that the model is identified.

*Part 1: $f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for $x$ close to 0.* We begin by proving the following lemma, which will imply that if $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ are both compatible with the observed distributions $f_{T|n}$ and $f_{T|n'}$, then $f_\theta^{(k)}(0) = g_\theta^{(k)}(0)$ and $f_\epsilon^{(k)}(0) = g_\epsilon^{(k)}(0)$ for $k = 0, 1, 2, \ldots$. By the assumption of piecewise real analytic distributions, this implies there is an interval $[0, \delta)$ on which all four distributions are equal to their Taylor expansions at 0, and therefore that $f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for $x \in [0, \delta)$.

*Lemma A1.* Fix $n$ and $n'$, with $n' > n \geq 2$, and let $f_\theta$ and $f_\epsilon$ be analytic.

(1) $f_\theta(0)$ and $f_\epsilon(0)$ can be recovered from the derivatives $f_{T|n}^{(n-1)}(0)$ and $f_{T|n'}^{(n'-1)}(0)$.
(2) For any $k > 0$, the derivatives $f_\theta^{(k)}(0)$ and $f_\epsilon^{(k)}(0)$ can be recovered from the derivatives $f_{T|n}^{(n-1+k)}(0)$ and $f_{T|n'}^{(n'-1+k)}(0)$ and the lower derivatives $\{f_\theta^{(j)}(0), f_\epsilon^{(j)}(0)\}_{j<k}$.

*Preliminary calculations for proof of Lemma A1.* To prove Lemma A1, we first perform some preliminary calculations summarized in the following Lemma. As in the text, let $y$ denote the second highest of the $\{\epsilon_i\}$ in a given auction, and let $f_{y|n}$ denote its distribution conditional on a value $n$ of $N$.

*Lemma A2.* For any $n \geq 2$,

(1) For any $k \geq 0$, $f_{T|n}^{(k)}(0) = \sum_{i=0}^{k-1} f_{y|n}^{(i)}(0) f_\theta^{(k-1-i)}(0)$.
(2) For $m < n-2$, $f_{y|n}^{(m)}(0) = 0$.
(3) $f_{y|n}^{(n-2)}(0) = n! \cdot (f_\epsilon(0))^{n-1}$.
(4) For $m > n-2$, $f_{y|n}^{(m)}(0)$ contains no derivatives of $f_\epsilon$ higher than $f_\epsilon^{(m-n+2)}(0)$, and the only term containing $f_\epsilon^{(m-n+2)}(0)$ is

$$n! \sum_{i=1}^{n-1} i \binom{m-i}{m-n+1} (f_\epsilon(0))^{n-2} f_\epsilon^{(m-n+2)}(0),$$

and for any $k > 0$, $A_n(k) \equiv \sum_{i=1}^{n-1} i \binom{n-2+k-i}{k-1}$ is strictly increasing in $n$.

*Proof of Lemma* A2. *Part 1*. As $T = \theta + y$, the standard convolution equation gives

$$F_{T|n}(t) = \int_0^t F_{y|n}(t-s) f_\theta(s) ds,$$

and therefore

$$f_{T|n}(t) = F_{y|n}(0) f_\theta(t) + \int_0^t f_{y|n}(t-s) f_\theta(s) ds = \int_0^t f_{y|n}(t-s) f_\theta(s) ds.$$

It is easily shown by induction that

$$f_{T|n}^{(k)}(t) = \sum_{i=0}^{k-1} f_{y|n}^{(i)}(0) f_\theta^{(k-1-i)}(t) + \int_0^t f_{y|n}^{(k)}(t-s) f_\theta(s) ds.$$

Plugging in $t = 0$ and noting that the integral term vanishes proves part 1. □

Parts 2 and 3.. Because $F_{y|n}(t) = nF_\epsilon^{n-1}(t) - (n-1)F_\epsilon^n(t)$,

$$f_{y|n}(t) = n(n-1)F_\epsilon^{n-2}(t) f_\epsilon(t) - n(n-1)F_\epsilon^{n-1}(t) f_\epsilon(t).$$

Differentiating gives

$$f'_{y|n}(t) = n(n-1)(n-2)F_\epsilon^{n-3}(t) f_\epsilon^2(t) + n(n-1)F_\epsilon^{n-2}(t) f'_\epsilon(t)$$
$$-n(n-1)^2 F_\epsilon^{n-2}(t) f_\epsilon^2(t) - n(n-1)F_\epsilon^{n-1}(t) f'_\epsilon(t).$$

Specifically, we get two terms from differentiating $n(n-1)F_\epsilon^{n-2}(t) f_\epsilon(t)$ – one from taking the derivative of $F_\epsilon^{n-2}$ and one from taking the derivative of $f_\epsilon(t)$—and likewise two terms from differentiating $n(n-1)F_\epsilon^{n-1}(t) f_\epsilon(t)$. As we take subsequent derivatives of $f_{y|n}$, we keep getting additional terms, each corresponding to differentiating one "piece" of a term from the previous derivative.

Part 2 of the lemma ($f_{y|n}^{(m)}(0) = 0$ for $m < n-2$) stems from the fact that until we have taken at least $n-2$ derivatives, every term in $f_{y|n}^{(m)}$ still contains a nonzero power of $F_\epsilon(t)$, that vanishes at 0. Likewise, when we take exactly $n-2$ derivatives, the only term that does not vanish at 0 is the one that "used" all $n-2$ derivatives to differentiate the $F_\epsilon^j(t)$ piece of the first term; each time this happens, the term gets multiplied by $j$ ($j$ running from $n-2$ down to 1) and picks up another $f_\epsilon(t)$, so

$$f_{y|n}^{(n-2)}(0) = n \cdot (n-1) \cdot (n-2)! \cdot (f_\epsilon(0))^{n-1} + \text{terms that vanish}$$

proving part 3.

*Part 4.*Next, suppose we take $m > n-2$ derivatives of $f_{y|n}$. Any term that has a derivative $f_\epsilon^{(m')}$ with $m' > m - n + 2$ must have "used" more than $m-(n-2)$ derivatives differentiating $f_\epsilon(t)$ and its subsequent derivatives; this would have left strictly fewer than $n-2$ derivatives to differentiate either $F_\epsilon^{n-2}$ or $F_\epsilon^{n-1}$, leaving a positive power of $F_\epsilon$ that would therefore vanish at 0. Finally, the only way to have a nonvanishing term containing $f_\epsilon^{(m-n+2)}$ would be to start with the first term of $f_{y|n}$, $n(n-1)F_\epsilon^{n-2}(t)f_\epsilon(t)$, and use exactly $n-2$ derivatives differentiating the $F_\epsilon^{n-2}$ term and the remaining $m-(n-2)$ derivatives differentiating $f_\epsilon(t)$ and its subsequent derivatives. Each of the $n-2$ derivatives we take of $F_\epsilon^{n-2}$ generates an additional $f_\epsilon$ term, and we differentiate only one of these, so we are left with $f_\epsilon^{n-2}(0)f_\epsilon^{(m-(n-2))}(0)$. The coefficient on this term is the sum of the coefficients of all the different "ways" we can generate these terms—basically, all the different orders in which we can take $n-2$ derivatives of $F_\epsilon^{n-2}$ and $m-(n-2)$ derivatives of $f_\epsilon$.

Now, regardless of the order in which we take the derivatives, at some point, we need to differentiate $F_\epsilon^{n-2}$, generating an $(n-2)$ coefficient; then at some point we differentiate $F_\epsilon^{n-3}$, generating an $(n-3)$; and so on. Combined with the $n(n-1)$ we started with, this gives us a coefficient of $n!$ attached to every nonvanishing term. In addition, at some point, we differentiated $f_\epsilon^i(t)$, which would have generated an $i$ coefficient. The rest of our derivatives were applied to the $f_\epsilon^{(j)}(t)$ term, which never gave any additional multiplicative coefficients.

Now, if we take the derivative of $f_\epsilon$ first—when the coefficient on $f_\epsilon^i(t)$ is $i=1$—then the coefficient on our eventual nonvanishing term will be $1 \cdot n!$. How many terms like this are there? Well, we still have $m-1$ derivatives left to take, of which $n-2$ need to apply to $F_\epsilon^{n-2}$ and the rest to $f_\epsilon^{(j)}$, so there are $\binom{m-1}{n-2}$ different terms corresponding to the choice of differentiating $f_\epsilon$ first.

More generally, suppose we differentiate $f_\epsilon$ after we have already differentiated $F_\epsilon^j$ $i-1$ times, and therefore when the term we are differentiating is $F_\epsilon^{n-2-(i-1)}(t)f_\epsilon^i(t)$. This again provides a new $i$ coefficient. And in addition, we have $m-1-(i-1) = m-i$ derivatives left to take, of which $n-2-(i-1) = n-1-i$ need to be applied to $F_\epsilon^j$; so there are $\binom{m-i}{n-1-i}$ different terms that correspond to this case.

Finally, if we wait to differentiate $f_\epsilon^i$ until after we have already taken $n-2$ derivatives of $F_\epsilon^{n-2}$, then we are differentiating $f_\epsilon^{n-1}$, and we get an $i = n-1$ coefficient; but then all remaining derivatives have to be applied to $f_\epsilon^{(j)}$, and there is only one way to do that.

All told, then, the coefficient on $f_\epsilon^{n-2}(0)f_\epsilon^{(m-n+2)}(0)$ in $f_{y|n}^{(m)}$ will be

$$n! \cdot \sum_{i=1}^{n-1} i \cdot \binom{m-i}{n-1-i} = n! \cdot \sum_{i=1}^{n-1} i \cdot \binom{m-i}{m-n+1}.$$

Final part. Finally, to show that $A_n(k)$ (which is $\frac{1}{n!}$ times this coefficient evaluated at $m = n-2+k$) is increasing in $n$, fix $k$ and calculate

$$\begin{aligned}
A_{n+1}(k) - A_n(k) &= \sum_{i=1}^{(n+1)-1} i\binom{(n+1)-2+k-i}{k-1} - \sum_{i=1}^{n-1} i\binom{n-2+k-i}{k-1} \\
&= \sum_{i=1}^{n} i\binom{n-2+k-(i-1)}{k-1} - \sum_{i=1}^{n-1} i\binom{n-2+k-i}{k-1} \\
&= \sum_{i'=0}^{n-1} (i'+1)\binom{n-2+k-i'}{k-1} - \sum_{i=1}^{n-1} i\binom{n-2+k-i}{k-1} \\
&= \sum_{i=0}^{n-1} \binom{n-2+k-i}{k-1} > 0,
\end{aligned}$$

completing the proof of Lemma A2.

*Proof of Lemma A1 given Lemma A2.* That concludes the preliminaries, and we can now prove Lemma A1.

*Part 1.*Parts 1, 2, and 3 of Lemma A2 give

$$f_{T|n}^{(n-1)}(0) = \sum_{i=0}^{n-2} f_{y|n}^{(i)}(0) f_\theta^{(n-2-i)}(0) = n! \cdot (f_\epsilon(0))^{n-1} \cdot f_\theta(0),$$

and likewise $f_{T|n'}^{(n'-1)}(0) = n'! \cdot (f_\epsilon(0))^{n'-1} \cdot f_\theta(0)$. For $n' > n$, then, we can recover $f_\epsilon(0)$ as

$$f_\epsilon(0) = \left( \frac{\frac{1}{n'!} f_{T|n'}^{(n'-1)}(0)}{\frac{1}{n!} f_{T|n}^{(n-1)}(0)} \right)^{1/(n'-n)},$$

and from there, recover $f_\theta(0)$ as $f_{T|n}^{(n-1)}(0)/(n!(f_\epsilon(0))^{n-1})$.

Part 2. Now assume we already know $\{f_\epsilon^{(j)}(0), f_\theta^{(j)}(0)\}_{j<k}$. From Lemma A2 parts 1 and 2,

$$f_{T|n}^{(n-1+k)}(0) = \sum_{i=0}^{n-2+k} f_{y|n}^{(i)}(0) f_\theta^{(n-2+k-i)}(0) = \sum_{i=n-2}^{n-2+k} f_{y|n}^{(i)}(0) f_\theta^{(n-2+k-i)}(0),$$

because the first $n-3$ derivatives of $f_{y|n}$ are 0 at 0. Lemma A2 part 4 implies that for $i < n-2+k$, $f_{y|n}^{(i)}(0)$ contains derivatives no higher than $f_\epsilon^{(k-1)}$, so the only "unknowns" on the right-hand side are $f_\theta^{(k)}(0)$ and the $f_\epsilon^{(k)}(0)$ term contained in $f_{y|n}^{(n-2+k)}$. Let

$$B(n,k) = \sum_{i=n-1}^{n-3+k} f_{y|n}^{(i)}(0) f_\theta^{(n-2+k-i)}(0),$$

be all but the first and last terms of the sum, and let $C(n,k)$ denote all the terms of $f_{y|n}^{(n-2+k)}(0)$ other than the one containing $f_\epsilon^{(k)}(0)$, both of which depend only on the derivatives $\{f_\epsilon^{(j)}(0), f_\theta^{(j)}(0)\}_{j<k}$ and are therefore known. We can then calculate the value of

$$f_{T|n}^{(n-1+k)}(0) - B(n,k) - C(n,k) f_\theta(0) = f_{y|n}^{(n-2)}(0) f_\theta^{(k)}(0) + \left(n! \cdot A_n(k)(f_\epsilon(0))^{n-2} f_\epsilon^{(k)}(0)\right) f_\theta(0),$$

where

$$A_n(k) = \sum_{i=1}^{n-1} i \binom{(n-2+k)-i}{(n-2+k)-n+1} = \sum_{i=1}^{n-1} i \binom{n-2+k-i}{k-1}.$$

Dividing by $f_{y|n}^{(n-2)}(0) = n!(f_\epsilon(0))^{n-1}$ and then by $f_\theta(0)/f_\epsilon(0)$, we get

$$\frac{f_{T|n}^{(n-1+k)}(0) - B(n,k) - C(n,k) f_\theta(0)}{n!(f_\epsilon(0))^{n-1}} \cdot \frac{f_\epsilon(0)}{f_\theta(0)} = \frac{f_\epsilon(0)}{f_\theta(0)} f_\theta^{(k)}(0) + A_n(k) f_\epsilon^{(k)}(0). \tag{A2}$$

Given our inductive assumption that $f_{T|n}^{(n-1+k)}$, $f_{T|n'}^{(n'-1+k)}$, and $\{f_\epsilon^{(j)}(0), f_\theta^{(j)}(0)\}_{j<k}$ are known, we can calculate the value of the left-hand side of (A2) for both $n$ and $n'$ and subtract, giving us the value of $(A_{n'}(k) - A_n(k)) f_\epsilon^{(k)}(0)$. The last part of Lemma A2 says that $A_n(k)$ is strictly increasing in $n$; because $A_{n'}(k) - A_n(k) \neq 0$, knowing the value of $(A_{n'}(k) - A_n(k)) f_\epsilon^{(k)}(0)$ allows us to recover $f_\epsilon^{(k)}(0)$. Once we have that, (A2) lets us calculate $f_\theta^{(k)}(0)$ as well, completing the proof of Lemma A1.

With Lemma A1 proved, we now know that if two pairs of distributions $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ both explain the data, we must have $\{f_\theta^{(k)}(0), f_\epsilon^{(k)}(0)\}_{k=0,1,2,\ldots} = \{g_\theta^{(k)}(0), g_\epsilon^{(k)}(0)\}_{k=0,1,2,\ldots}$. By the assumption that all densities are piecewise real analytic, there must be some $\delta > 0$ such that $f_\theta$, $f_\epsilon$, $g_\theta$, and $g_\epsilon$ on $[0, \delta)$ are the restrictions of real analytic functions on an open neighborhood containing $[0, \delta)$. This means that on $[0, \delta')$, all four of these distributions are equal to their Taylor expansions around 0, which are identical, and therefore that $f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for $x \in [0, \delta')$, concluding Part 1 of the proof.

*Part 2: If $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ begin to diverge at $t^* > 0$, they cannot both match the data just above $t^*$.* Let

$$t^* = \min\{\inf\{x : f_\theta(x) \neq g_\theta(x)\},\ \inf\{x : f_\epsilon(x) \neq g_\epsilon(x)\}\}$$

be the point at which $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ begin to diverge from one another. By assumption, $t^*$ exists, because otherwise $f_\theta = g_\theta$ and $f_\epsilon = g_\epsilon$. From Part 1 above, we know that $t^* \geq \delta$ for some $\delta > 0$, or $t^* > 0$. We will show that $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ cannot both match the two observed distributions $f_{T|n}$ and $f_{T|n'}$ in a neighborhood just to the right of $t^*$.

We begin with a rather technical lemma:

*Lemma A3.* If

$$\int_0^t f_\theta(t-s) d\psi_n(F_\epsilon(s)) = \int_0^t g_\theta(t-s) d\psi_n(G_\epsilon(s))$$

and $(f_\theta, f_\epsilon) = (g_\theta, g_\epsilon)$ on $[0, t^*]$, then for any $\gamma \leq t^*$,

$$\int_0^\gamma (f_\theta(t^*+\gamma-s) - g_\theta(t^*+\gamma-s)) d\psi_n(F_\epsilon(s)) = \int_{t^*}^{t^*+\gamma} f_\theta(t^*+\gamma-s)(d\psi_n(G_\epsilon(s)) - d\psi_n(F_\epsilon(s))). \tag{A3}$$

*Proof of Lemma A* 3. The proof is just algebra. Plugging $t = t^* + \gamma$ into the first integral gives

$$\int_0^{t^*+\gamma} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) = \int_0^{t^*+\gamma} g_\theta(t^* + \gamma - s)d\psi_n(G_\epsilon(s)).$$

We rewrite each integral as the sum of three integrals: from 0 to $\gamma$, from $\gamma$ to $t^*$, and from $t^*$ to $t^* + \gamma$. This gives

$$\begin{aligned}&\int_0^{\gamma} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_\gamma^{t^*} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_{t^*}^{t^*+\gamma} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) =\\ &\int_0^{\gamma} g_\theta(t^* + \gamma - s)d\psi_n(G_\epsilon(s)) + \int_\gamma^{t^*} g_\theta(t^* + \gamma - s)d\psi_n(G_\epsilon(s)) + \int_{t^*}^{t^*+\gamma} g_\theta(t^* + \gamma - s)d\psi_n(G_\epsilon(s)).\end{aligned}$$

Noting that $f_\theta(x) = g_\theta(x)$ and $f_\epsilon(x) = g_\epsilon(x)$ for $x \leq t^*$, we can write this as

$$\begin{aligned}&\int_0^{\gamma} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_\gamma^{t^*} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_{t^*}^{t^*+\gamma} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) =\\ &\int_0^{\gamma} g_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_\gamma^{t^*} f_\theta(t^* + \gamma - s)d\psi_n(F_\epsilon(s)) + \int_{t^*}^{t^*+\gamma} f_\theta(t^* + \gamma - s)d\psi_n(G_\epsilon(s)).\end{aligned}$$

Canceling the middle integrals and rearranging, this simplifies to (A3), proving the lemma. □

Next:

*Lemma A4.* If

$$\int_0^{t} f_\theta(t - s)d\psi_n(F_\epsilon(s)) = \int_0^{t} g_\theta(t - s)d\psi_n(G_\epsilon(s))$$

for all $t$ and $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ start to diverge at $t^*$, then there exists $\delta > 0$ such that either

(1) $f_\theta(x) > g_\theta(x)$, $f_\epsilon(x) < g_\epsilon(x)$, and $d\psi_n(F_\epsilon(x)) < d\psi_n(G_\epsilon(x))$ for every $x \in (t^*, t^* + \delta)$, or
(2) $f_\theta(x) < g_\theta(x)$, $f_\epsilon(x) > g_\epsilon(x)$, and $d\psi_n(F_\epsilon(x)) > d\psi_n(G_\epsilon(x))$ for every $x \in (t^*, t^* + \delta)$.

First, note that if two piecewise-analytic functions $f$ and $g$ begin to diverge at $t^*$, then there must be some $\delta > 0$ such that either $f(x) > g(x)$ for $x \in (t^*, t^* + \delta)$ or $f(x) < g(x)$ for $x \in (t^*, t^* + \delta)$. To see this, let $h(x) = f(x) - g(x)$, and note that although $f$ and $g$ need not both be differentiable at $x$, they must both be *right* differentiable at $x$, with $h(z)$ equal to its Taylor expansion (based on the right derivatives of $f - g$) for $z \in [x, x + \varepsilon)$ for some $\varepsilon > 0$. Because by assumption, $h$ is not uniformly 0 on $[x, x + \varepsilon)$, it must have some finite lowest-order right derivative that is not zero at $t^*$, that is, some finite $r$ such that $h_+^{(r)}(t^*) \neq 0$ but $h_+^{(k)}(t^*) = 0$ for all $k < r$. For $t$ close to $t^*$, this term will dominate the Taylor expansion of $h$ at $t^*$, so we will either have $h(t) > 0$ for all $t$ close enough to $t^*$ (if $h_+^{(r)}(t^*) > 0$) or $h(t) < 0$ for all $t$ close enough to $t^*$ (if $h_+^{(r)}(t^*) < 0$).

Next, we show that $f_\epsilon$ and $g_\epsilon$ must start to diverge at $t^*$, that is, it is impossible that $f_\theta$ and $g_\theta$ start to diverge before $f_\epsilon$ and $g_\epsilon$. If they did, then for $\gamma$ sufficiently close to 0, the right-hand side of (A3) would be 0, whereas the left-hand side would not (because $f_\theta(\cdot) - g_\theta(\cdot)$ would be either positive or negative on the whole range of integration). As we are free to switch the labels of $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$, we can therefore assume without loss that there is some $\delta_1 > 0$ such that $f_\epsilon(x) < g_\epsilon(x)$ for $x \in (t^*, t^* + \delta_1)$.

Next, we show that for $x$ sufficiently close to $t^*$, $d\psi_n(F_\epsilon(x)) < d\psi_n(G_\epsilon(x))$. We can write

$$\begin{aligned}d\psi_n(F_\epsilon(x)) - d\psi_n(G_\epsilon(x)) &= n(n-1)\big(F_\epsilon^{n-2}(x)(1 - F_\epsilon(x))f_\epsilon(x) - G_\epsilon^{n-2}(x)(1 - G_\epsilon(x))g_\epsilon(x)\big)dx\\ &\leq n(n-1)G_\epsilon^{n-2}(x)((1 - F_\epsilon(x))f_\epsilon(x) - (1 - G_\epsilon(x))g_\epsilon(x))dx,\end{aligned}$$

so it suffices to show that $(1 - F_\epsilon(x))f_\epsilon(x) < (1 - G_\epsilon(x))g_\epsilon(x)$. Let $e(x) = g_\epsilon(x) - f_\epsilon(x)$. Then

$$\begin{aligned}(1 - F_\epsilon(x))f_\epsilon(x) - (1 - G_\epsilon(x))g_\epsilon(x) &= f_\epsilon(x)(G_\epsilon(x) - F_\epsilon(x)) - (g_\epsilon(x) - f_\epsilon(x))(1 - G_\epsilon(x))\\ &= f_\epsilon(x)\int_0^x e(s)ds - e(x)(1 - G_\epsilon(x))\\ &\propto \frac{\int_0^x e(s)ds}{e(x)} - \frac{1 - G_\epsilon(x)}{f_\epsilon(x)}.\end{aligned}$$

Now, we know that $f_\epsilon(t^*) = g_\epsilon(t^*)$, or $e(t^*) = 0$, and of course $\int_0^{t^*} e(s)ds = 0$ as well. However, we can calculate the limit of their ratio as $x \searrow t^*$ using L'Hopital's rule (using the right derivatives of $e$ at $t^*$), which says it is equal to $\frac{\lim_{x\searrow t^*} e(x)}{\lim_{x\searrow t^*} e'_+(x)}$. The numerator is zero; if the denominator is as well, we can apply L'Hopital again, and find that the limit is $\frac{\lim e'_+(x)}{\lim e''_+(x)}$, and so on. As discussed before, $e$ must have a finite lowest-order nonzero right derivative at $t^*$, at which point we learn that $\lim \frac{\int_0^x e(s)ds}{e(x)} = 0$, and therefore (because $(1 - G_\epsilon(x))/f_\epsilon(x) > 0$) that $\frac{\int_0^x d(s)ds}{e(x)} - \frac{1-G_\epsilon(x)}{f_\epsilon(x)} < 0$ for $x$ close enough to $t^*$; thus, $d\psi_n(F_\epsilon(x)) < d\psi_n(G_\epsilon(x))$ for $x$ close to $t^*$.

Finally, knowing that $f_\epsilon(x) < g_\epsilon(x)$ and $d\psi_n(F_\epsilon(x)) < d\psi_n(G_\epsilon(x))$ for $x \in (t^*, t^* + \delta)$ for some sufficiently small $\delta$, we can return to (A3) and note that for $\gamma$ sufficiently close to $t^*$, the right-hand side is strictly positive. This means the left-hand side must be positive, which means for *some* $x \in (t^*, t^* + \gamma)$, $f_\theta(x) > g_\theta(x)$. As we can make this argument for arbitrarily small $\gamma$, we must have $f_\theta(x) > g_\theta(x)$ for all $x \in (t^*, t^* + \delta')$ for some $\delta'$ sufficiently small, proving Lemma A4. □

So now return to our main objective: proving that $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ cannot both match both $f_{T|n}$ and $f_{T|n'}$ in a neighborhood to the right of $t^*$. Choose a value of $\delta$ smaller than $t^*$, and small enough for Lemma A4 to hold, that is, such that for any $x \in (t^*, t^* + \delta)$, $f_\epsilon(x) < g_\epsilon(x)$, $d\psi_n(F_\epsilon(x)) < d\psi_n(G_\epsilon(x))$, and $f_\theta(x) > g_\theta(x)$. If both pairs of distributions match $f_{T|n}$, then (A3) holds everywhere; plug in $\gamma = \delta$, and define $A_n$ as the left-hand side and $B_n$ as the right-hand side.

If both pairs of distributions were to also match $f_{T|n'}$, then (A3) would also have to hold with $n'$ replacing $n$; let $A_{n'}$ and $B_{n'}$ denote the left- and right-hand sides, respectively, again with $\gamma = \delta$. Now,

$$\begin{aligned}
A_{n'} &= \int_0^\delta (f_\theta(t^* + \delta - s) - g_\theta(t^* + \delta - s)) d\psi_{n'}(F_\epsilon(s)) \\
&= \int_0^\delta (f_\theta(t^* + \delta - s) - g_\theta(t^* + \delta - s)) \frac{d\psi_{n'}(F_\epsilon(s))}{d\psi_n(F_\epsilon(s))} d\psi_n(F_\epsilon(s)) \\
&= \int_0^\delta (f_\theta(t^* + \delta - s) - g_\theta(t^* + \delta - s)) \frac{n'(n'-1)F_\epsilon^{n'-2}(s)(1 - F_\epsilon(s))f_\epsilon(s)ds}{n(n-1)F_\epsilon^{n-2}(s)(1 - F_\epsilon(s))f_\epsilon(s)ds} d\psi_n(F_\epsilon(s)) \\
&= \int_0^\delta (f_\theta(t^* + \delta - s) - g_\theta(t^* + \delta - s)) \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(s) d\psi_n(F_\epsilon(s)) \\
&\leq \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(\delta) A_n,
\end{aligned}$$

because $n' > n$, $F_\epsilon$ is increasing, and the integrand is positive. On the other hand,

$$\begin{aligned}
B_{n'} &= \int_{t^*}^{t^*+\delta} f_\theta(t^* + \delta - s)(d\psi_{n'}(G_\epsilon(s)) - d\psi_{n'}(F_\epsilon(s))) \\
&= \int_{t^*}^{t^*+\delta} f_\theta(t^* + \delta - s) \frac{d\psi_{n'}(F_\epsilon(s))}{d\psi_n(F_\epsilon(s))} (d\psi_n(G_\epsilon(s)) - d\psi_n(F_\epsilon(s))) \\
&\quad + \int_{t^*}^{t^*+\delta} f_\theta(t^* + \delta - s) \frac{d\psi_{n'}(G_\epsilon(s)) - d\psi_{n'}(F_\epsilon(s))}{d\psi_n(F_\epsilon(s))} d\psi_n(G_\epsilon(s)) \\
&\geq \int_{t^*}^{t^*+\delta} f_\theta(t^* + \delta - s) \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(s)(d\psi_n(G_\epsilon(s)) - d\psi_n(F_\epsilon(s))) \\
&\geq \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(t^*) B_n.
\end{aligned}$$

Because $A_n = B_n$, and because $F_\epsilon$ is strictly increasing and $t^* > \delta'$,

$$\begin{aligned}
B_{n'} &\geq \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(t^*) B_n = \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(t^*) A_n \\
&> \frac{n'(n'-1)}{n(n-1)} F_\epsilon^{n'-n}(\delta') A_n \geq A_{n'},
\end{aligned}$$

and therefore if $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ both rationalize $f_{T|n}$ just to the right of $t^*$, they cannot both rationalize $f_{T|n'}$ there, proving identification. □

□ **Proof of Theorem 2.** *Outline.* We follow the same outline as the proof of Theorem 1. First, we show that if there were two pairs of distributions $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ that both rationalized the data, they would need to agree on a neighborhood to the right of 0. And second, we show that they could not both match the data in a neighborhood to the right of $t^*$, the point at which they begin to diverge.

*Part 1.* Recall $\mathbf{p}(n|x) = \Pr(N = n|X = x)$, and let $\mathbf{p}(x)$ be the vector $(\mathbf{p}(n|x))_n$. Let $x$ and $x'$ be two values of $X$. Let $F_{T|X}$ and $f_{T|X}$ denote the distribution and density of transaction prices given $X$. We will show that if $\mathbf{p}(x') \neq \mathbf{p}(x)$ (whether or not they satisfy the MLRP), the derivatives $\{f_\epsilon^{(k)}(0), f_\theta^{(k)}(0)\}_{k=0,1,2,...}$ are uniquely determined by $f_{T|x}$ and $f_{T|x'}$, so both pairs of distributions must have all the same derivatives at 0, and must therefore coincide in a neighborhood of 0.

We prove the result in two steps. First, we consider the simpler case when the support of $N$ when $X = x$ extends below the support of $N$ when $X = x'$; we then extend the result to the case where this does not hold.

*Lemma A5.* Let $\underline{n} = \min\{n : \mathbf{p}(n|x) > 0\}$ and $\underline{n}' = \min\{n : \mathbf{p}(n|x') > 0\}$. If $\underline{n}' > \underline{n}$, then with $\mathbf{p}(x)$ and $\mathbf{p}(x')$ already known,

(1) $f_\theta(0)$ and $f_\epsilon(0)$ can be recovered from the derivatives $f_{T|x}^{(\underline{n}-1)}(0)$ and $f_{T|x'}^{(\underline{n}'-1)}(0)$

(2) For any $k > 0$, the derivatives $f_\theta^{(k)}(0)$ and $f_\epsilon^{(k)}(0)$ can be recovered from the derivatives $f_{T|x}^{(\underline{n}-1+k)}(0)$ and $f_{T|x'}^{(\underline{n}'-1+k)}(0)$ and the lower derivatives $\{f_\theta^{(j)}(0), f_\epsilon^{(j)}(0)\}_{j<k}$

*Part 1*. Let $p_n = \mathbf{p}(n|x)$ and $p'_n = \mathbf{p}(n|x')$. Because $p_n = 0$ for $n < \underline{n}$, when $X = x$, the observed distribution of transaction prices will be

$$f_{T|x}(t) = \sum_{n=\underline{n}}^{\infty} p_n f_{T|n}(t).$$

Applying Lemma A2 part 1 with $k = \underline{n} - 1$,

$$f_{T|x}^{(\underline{n}-1)}(0) = \sum_{n=\underline{n}}^{\infty} p_n \left( \sum_{i=0}^{\underline{n}-2} f_{y|n}^{(i)}(0) f_\theta^{(\underline{n}-2-i)}(0) \right) = \sum_{i=0}^{\underline{n}-2} \left( \sum_{n=\underline{n}}^{\infty} p_n f_{y|n}^{(i)}(0) \right) f_\theta^{(\underline{n}-2-i)}(0).$$

By Lemma A2 part 2, $f_{y|n}^{(i)}(0) = 0$ for $i < n - 2$, or $n > i + 2$; because $n \geq \underline{n}$ and $i \leq \underline{n} - 2$, the only term in the double sum that does not vanish is when $i = \underline{n} - 2$ and $n = \underline{n}$, meaning

$$f_{T|x}^{(\underline{n}-1)}(0) = p_{\underline{n}} f_{y|\underline{n}}^{(\underline{n}-2)}(0) f_\theta(0) = p_{\underline{n}} \underline{n}! (f_\epsilon(0))^{\underline{n}-1} f_\theta(0),$$

by Lemma A2 part 3. By the same logic,

$$f_{T|x'}^{(\underline{n}'-1)}(0) = p'_{\underline{n}'} f_{y|\underline{n}'}^{(\underline{n}'-2)}(0) f_\theta(0) = p'_{\underline{n}'} (\underline{n}')! (f_\epsilon(0))^{\underline{n}'-1} f_\theta(0),$$

and so dividing,

$$\frac{f_{T|x'}^{(\underline{n}'-1)}(0)}{f_{T|x}^{(\underline{n}-1)}(0)} = \frac{p'_{\underline{n}'}}{p_{\underline{n}}} \frac{(\underline{n}')!}{\underline{n}!} (f_\epsilon(0))^{\underline{n}'-\underline{n}}.$$

By assumption, $\underline{n}' > \underline{n}$, $p'_{\underline{n}'} > 0$ and $p_{\underline{n}} > 0$, and everything in this last equation is already known except for $f_\epsilon(0)$, which is therefore identified. Once $f_\epsilon(0)$ is known, we can recover $f_\theta(0)$ as the only remaining unknown in $f_{T|x}^{(\underline{n}-1)}(0) p_{\underline{n}} \underline{n}! (f_\epsilon(0))^{\underline{n}-1} f_\theta(0)$.

*Part 2*. Again from Lemma A2,

$$f_{T|x}^{(\underline{n}-1+k)}(0) = \sum_{n=\underline{n}}^{\infty} p_n f_{T|n}^{(\underline{n}-1+k)}(0) = \sum_{n=\underline{n}}^{\infty} p_n \left( \sum_{i=0}^{\underline{n}-2+k} f_{y|n}^{(i)}(0) f_\theta^{(\underline{n}-2+k-i)}(0) \right).$$

Recalling that $f_{y|n}^{(i)}(0) = 0$ for $i < n - 2$, we can write this as

$$f_{T|x}^{(\underline{n}-1+k)}(0) = \sum_{n=\underline{n}}^{\underline{n}+k} p_n \left( \sum_{i=n-2}^{\underline{n}-2+k} f_{y|n}^{(i)}(0) f_\theta^{(\underline{n}-2+k-i)}(0) \right), \tag{A4}$$

because terms with either $i < n - 2$ (in the inner sum) or $n > \underline{n} + k \geq i + 2$ (in the outer sum) will all vanish.

Next, define

$$B_1 = \sum_{n=\underline{n}+1}^{\underline{n}+k} p_n \left( \sum_{i=n-2}^{\underline{n}-2+k} f_{y|n}^{(i)}(0) f_\theta^{(\underline{n}-2+k-i)}(0) \right)$$

as everything but the $n = \underline{n}$ term of equation (A4). Because $n > \underline{n}$, $i \leq \underline{n} - 2 + k < n - 2 + k$, and therefore for every term in the double sum, $f_{y|n}^{(i)}$ contains no derivative of $f_\epsilon$ higher than $f_\epsilon^{(k-1)}(0)$. Likewise, because $i \geq n - 2 > \underline{n} - 2$, for every term in the double sum, $f_\theta^{(\underline{n}-2+k-i)}(0)$ is at most the $k - 1$st derivative of $f_\theta$. By assumption, then, every term in $B_1$ is already known.

Next, define

$$B_2 = p_{\underline{n}} \left( \sum_{i=\underline{n}-2+1}^{\underline{n}-2+k-1} f_{y|\underline{n}}^{(i)}(0) f_\theta^{(\underline{n}-2+k-i)}(0) \right)$$

as everything but the first and last terms ($i = \underline{n} - 2$ and $i = \underline{n} - 2 + k$) of the inner sum in the $n = \underline{n}$ term of the outer sum of equation (A4), or as 0 in the case $k = 1$ (where the sum is empty). Because $i > \underline{n} - 2$, there are no derivatives of $f_\theta$ higher than the $k - 1$st; and because $i < \underline{n} - 2 + k$, there are no derivatives of $f_\epsilon$ higher than the $k - 1$st; so by assumption, every term in $B_2$ is already known.

Finally, recall from Lemma A2 part 4 (with $n = \underline{n}$ and $m = \underline{n} - 2 + k$) that we can write

$$f_{y|\underline{n}}^{(\underline{n}-2+k)}(0) = \underline{n}! A_{\underline{n}}(k) (f_\epsilon(0))^{(\underline{n}-2)} f_\epsilon^{(k)}(0) + B_3,$$

where $B_3$ contains only terms depending on the first $k - 1$ derivatives of $f_\epsilon$, which are by assumption known.

Putting it all together, then, we have therefore rewritten equation (A4) as

$$f_{T|x}^{(\underline{n}-1+k)}(0) = p_{\underline{n}} f_{y|\underline{n}}^{(\underline{n}-2)}(0) f_\theta^{(k)}(0) + p_{\underline{n}} \big( \underline{n}! A_{\underline{n}}(k) (f_\epsilon(0))^{(\underline{n}-2)} f_\epsilon^{(k)}(0) + B_3 \big) f_\theta(0) + B_2 + B_1,$$

which we can rearrange to give

$$f_{T|x}^{(\underline{n}-1+k)}(0) - p_{\underline{n}}B_3 f_\theta(0) - B_2 - B_1 = p_{\underline{n}} f_{y|\underline{n}}^{(\underline{n}-2)}(0) f_\theta^{(k)}(0) + p_{\underline{n}}\big(\underline{n}! A_{\underline{n}}(k)(f_\epsilon(0))^{(\underline{n}-2)} f_\epsilon^{(k)}(0)\big) f_\theta(0),$$
$$\downarrow$$
$$\frac{f_{T|x}^{(\underline{n}-1+k)}(0) - p_{\underline{n}}B_3 f_\theta(0) - B_2 - B_1}{p_{\underline{n}} f_{y|\underline{n}}^{(\underline{n}-2)}(0)} = f_\theta^{(k)}(0) + \frac{\underline{n}! A_{\underline{n}}(k)(f_\epsilon(0))^{(\underline{n}-2)} f_\epsilon^{(k)}(0)}{p_{\underline{n}} f_{y|\underline{n}}^{(\underline{n}-2)}(0)} f_\theta(0).$$

Plugging in $f_{y|\underline{n}}^{(\underline{n}-2)}(0) = \underline{n}! \cdot (f_\epsilon(0))^{\underline{n}-1}$ (from Lemma A2) and multiplying both sides by $\frac{f_\epsilon(0)}{f_\theta(0)}$, this becomes

$$\frac{f_\epsilon(0)}{f_\theta(0)} \frac{f_{T|x}^{(\underline{n}-1+k)}(0) - p_{\underline{n}}B_3 f_\theta(0) - B_2 - B_1}{p_{\underline{n}}\underline{n}! \cdot (f_\epsilon(0))^{\underline{n}-1}} = \frac{f_\epsilon(0)}{f_\theta(0)} f_\theta^{(k)}(0) + A_{\underline{n}}(k) f_\epsilon^{(k)}(0), \tag{A5}$$

where the left-hand side is a combination of "data" and terms we already know.

Repeating the argument starting with $f_{T|x'}^{(\underline{n}'-1+k)}(0)$, we can similarly calculate the value of

$$\frac{f_\epsilon(0)}{f_\theta(0)} f_\theta^{(k)}(0) + A_{\underline{n}'}(k) f_\epsilon^{(k)}(0)$$

and, subtracting the former from the latter, calculate the value of

$$\big[A_{\underline{n}'}(k) - A_{\underline{n}}(k)\big] f_\epsilon^{(k)}(0).$$

Because by assumption, $\underline{n}' > \underline{n}$, and (from Lemma A2) $A_n(k)$ is strictly increasing in $n$, this allows us to recover $f_\epsilon^{(k)}(0)$; once this is known, we can recover $f_\theta^{(k)}(0)$ as the only remaining unknown in equation (A5), concluding the proof of Lemma A5. □

*Extending to the case where* $\underline{n}' = \underline{n}$. If $\underline{n}' > \underline{n}$, that is, if the distributions of $N|X = x$ and $N|X = x'$ differ in terms of the bottom of their supports, then we have shown the derivatives of $f_\theta$ and $f_\epsilon$ at 0 are uniquely pinned down. The remaining challenge is to extend the result to the case where $\underline{n}' = \underline{n}$. To do this, we will essentially synthesize a new distribution of transaction prices corresponding to a distribution of $N$ which puts no weight on $N \leq \underline{n}$.

Note, crucially, that in the proof of Lemma A5, we never used the fact that $\sum_{n=\underline{n}}^{\infty} p_n = 1$, or even that $p_n \geq 0$ for a particular $n$, only that $\{p_n\}$ was a known collection of weights with $p_n = 0$ for $n < \underline{n}$ and $p_{\underline{n}} > 0$ (so we could divide by it), and likewise for $\{p'_n\}$ with regard to $\underline{n}'$. For the case where $\underline{n}' = \underline{n}$, then, define a new distribution

$$g(t) = p'_{\underline{n}} f_{T|x}(t) - p_{\underline{n}} f_{T|x'}(t).$$

It is easy to show that

$$g(t) = \sum_{n=2}^{\infty} q_n f_{T|n}(t),$$

where $q_n = p'_{\underline{n}} p_n - p_{\underline{n}} p'_n$, and that $q_n = 0$ for $n \leq \underline{n}$.

Next, note that $\{q_n\}$ cannot be uniformly 0 if $\{p_n\} \neq \{p'_n\}$. (If $p'_{\underline{n}} = p_{\underline{n}}$, then $q_n = 0$ requires $p'_n = p_n$, so $\{q_n\} = 0$ would imply $\mathbf{p}(x') = \mathbf{p}(x)$. If $p'_{\underline{n}} > p_{\underline{n}}$, then $q_n = 0$ requires $p'_n > p_n$, and if this holds for every $n \geq \underline{n}$, then $\{p'_n\}$ and $\{p_n\}$ cannot both sum to 1.) Let $\underline{n}'' = \min\{n : q_n > 0\}$. Because by construction $\underline{n}'' > \underline{n}$, Lemma A5 applies, using $(g, \{q_n\}, \underline{n}'')$ in place of $(f_{T|x'}, \{p'_n\}, \underline{n}')$ as the second distribution. That is, we can recover $\{f_\theta^{(k)}(0), f_\epsilon^{(k)}(0)\}_{k=0,1,2,\ldots}$ from the derivatives of $f_{T|x}$ and $g$ at 0, knowing the "distributions" $\{p_n\}$ and $\{q_n\}$.

Once we know (whether $\underline{n}' = \underline{n}$ or not) that the derivatives $\{f_\theta^{(k)}(0), f_\epsilon^{(k)}(0)\}_{k=0,1,2,\ldots}$ can be uniquely determined from the data, the argument is the same as in Part 1 of the proof of Theorem 1 that $(f_\theta(x), f_\epsilon(x)) = (g_\theta(x), g_\epsilon(x))$ for all $x$ in some neighborhood $[0, \delta)$ for any two pairs of distributions that match the data.

*Part 2.* If $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ both rationalize the observed distributions $f_{T|x}$ and $f_{T|x'}$, they must satisfy

$$\int_0^t f_\theta(s) d\psi_p(F_\epsilon(t-s)) = f_{T|x}(t) = \int_0^t g_\theta(s) d\psi_p(G_\epsilon(t-s)),$$

and

$$\int_0^t f_\theta(s) d\psi_{p'}(F_\epsilon(t-s)) = f_{T|x'}(t) = \int_0^t g_\theta(s) d\psi_{p'}(G_\epsilon(t-s)), \tag{A6}$$

where

$$d\psi_p(z) = \sum_{n\geq 2} p_n d\psi_n(z) \quad \text{and} \quad d\psi_{p'}(z) = \sum_{n\geq 2} p'_n d\psi_n(z)$$

and $p_n = \Pr(N = n|X = x)$ and $p'_n = \Pr(N = n|X = x')$. Lemmas A3 and A4 still hold, just with $d\psi_p(\cdot)$ and $d\psi_{p'}(\cdot)$ replacing $d\psi_n(\cdot)$ and $d\psi_{n'}(\cdot)$. Thus, for $\delta$ sufficiently small, we can still conclude that for $(f_\theta, f_\epsilon)$ and $(g_\theta, g_\epsilon)$ to both rationalize the data, we must have $A_p = B_p$ and $A_{p'} = B_{p'}$, where now

$$\begin{aligned}
A_p &= \int_0^\delta (f_\theta(t^*+\delta-s) - g_\theta(t^*+\delta-s))d\psi_p(F_\epsilon(s)),\\
A_{p'} &= \int_0^\delta (f_\theta(t^*+\delta-s) - g_\theta(t^*+\delta-s))\frac{d\psi_{p'}(F_\epsilon(s))}{d\psi_p(F_\epsilon(s))}d\psi_p(F_\epsilon(s)),\\
B_p &= \int_{t^*}^{t^*+\delta} f_\theta(t^*+\delta-s)\big(d\psi_p(G_\epsilon(s)) - d\psi_p(F_\epsilon(s))\big),\\
B_{p'} &= \int_{t^*}^{t^*+\delta} f_\theta(t^*+\delta-s)\frac{d\psi_{p'}(F_\epsilon(s))}{d\psi_p(F_\epsilon(s))}\big(d\psi_p(G_\epsilon(s)) - d\psi_p(F_\epsilon(s))\big)\\
&\quad + \int_{t^*}^{t^*+\delta} f_\theta(t^*+\delta-s)\frac{d\psi_{p'}(G_\epsilon(s)) - d\psi_{p'}(F_\epsilon(s))}{d\psi_p(F_\epsilon(s))}d\psi_p(G_\epsilon(s)).
\end{aligned}$$

Further, we can continue to assume that for $\delta$ sufficiently small, $f_\epsilon(s) < g_\epsilon(s)$, $d\psi_p(F_\epsilon(s)) < d\psi_p(G_\epsilon(s))$, $d\psi_{p'}(F_\epsilon(s)) < d\psi_{p'}(G_\epsilon(s))$, and $f_\theta(s) > g_\theta(s)$ for $s \in (t^*, t^*+\delta)$. This means that it will suffice to show that $d\psi_{p'}(F_\epsilon(s))/d\psi_p(F_\epsilon(s))$ is strictly increasing, as then if $A_p = B_p$ then

$$A_{p'} \le \frac{d\psi_{p'}(F_\epsilon(\delta))}{d\psi_p(F_\epsilon(\delta))}A_p = \frac{d\psi_{p'}(F_\epsilon(\delta))}{d\psi_p(F_\epsilon(\delta))}B_p < \frac{d\psi_{p'}(F_\epsilon(t^*))}{d\psi_p(F_\epsilon(t^*))}B_p \le B_{p'}$$

and so $A_p = B_p$ implies $A_{p'} \neq B_{p'}$, meaning no two distinct pairs of distributions can match both $f_{T|x}$ and $f_{T|x'}$.

$p'$ is said to dominate $p$ via the likelihood ratio ordering (LRO), $p' \ge_{lr} p$, if $\frac{p'_n}{p_n}$ is nondecreasing in $n$. It will suffice to show that if, as Theorem 2 assumes, $p' \ge_{lr} p$ and $p' \neq p$, then

$$\frac{d\psi_{p'}(s)}{d\psi_p(s)} = \frac{\sum_{n\ge2} p'_n n(n-1)s^{n-2}(1-s)}{\sum_{n\ge2} p_n n(n-1)s^{n-2}(1-s)} = \frac{\sum_{n\ge2} p'_n n(n-1)s^{n-2}}{\sum_{n\ge2} p_n n(n-1)s^{n-2}}$$

is strictly increasing on [0,1]. To show this, for $n = 2, 3, \ldots$, define

$$a_{n-2} \equiv \frac{p'_n n(n-1)}{\sum_{m=2}^\infty p'_m m(m-1)} \quad \text{and} \quad b_{n-2} \equiv \frac{p_n n(n-1)}{\sum_{m=2}^\infty p_m m(m-1)}.$$

For $n' > n$, note that

$$\frac{a_{n'}}{a_n} = \frac{p'_{n'+2}(n'+2)(n'+1)/\sum_{m=2}^\infty p'_m m(m-1)}{p'_{n+2}(n+2)(n+1)/\sum_{m=2}^\infty p'_m m(m-1)} = \frac{(n'+2)(n'+1)}{(n+2)(n+1)}\frac{p'_{n'+2}}{p'_{n+2}}$$

and similarly

$$\frac{b_{n'}}{b_n} = \frac{(n'+2)(n'+1)}{(n+2)(n+1)}\frac{p_{n'+2}}{p_{n+2}}$$

and so

$$\frac{a_{n'}}{a_n} \ge \frac{b_{n'}}{b_n} \quad \longleftrightarrow \quad \frac{p'_{n'+2}}{p'_{n+2}} \ge \frac{p_{n'+2}}{p_{n+2}}$$

and therefore $p' \ge_{lr} p$ implies $a \ge_{lr} b$. (Note also that $a$ and $b$ have been normalized such that $\sum_n a_n = \sum_n b_n = 1$, and both have support in $\mathbb{Z}^+$.) Thus, identification follows from the following Lemma:

*Lemma A6.* Let $a$ and $b$ be probability distributions over $\mathbb{Z}^+$, with $a \neq b$ and $a \ge_{lr} b$. Then

$$\frac{a(s)}{b(s)} \equiv \frac{\sum_n a_n s^n}{\sum_n b_n s^n}$$

is strictly increasing in $s$ on [0,1].

To prove Lemma A6, we begin with two observations:

(1) The LRO ranking survives truncation.
Specifically, if we let $\hat{a}$ and $\hat{b}$ be the restrictions of $a$ and $b$ to a subset $Z$ of their domains, then for $n, n' \in Z$ with $n' > n$,

$$\frac{\hat{a}_{n'}}{\hat{a}_n} = \frac{a_{n'}/\sum_{m\in Z} a_m}{a_n/\sum_{m\in Z} a_m} = \frac{a_{n'}}{a_n}$$

and likewise $\frac{\hat{b}_{n'}}{\hat{b}_n} = \frac{b_{n'}}{b_n}$, so $a \ge_{lr} b$ implies $\hat{a} \ge_{lr} \hat{b}$. (Note, however, that $a \neq b$ does not rule out the possibility that $\hat{a} = \hat{b}$.)

(2) If $\bar{n}$ is the top point in the support of $a$ and $b$, then $a \geq_{lr} b$ and $a \neq b$ implies $a_{\bar{n}} > b_{\bar{n}}$.
If not, then $a_{\bar{n}} \leq b_{\bar{n}}$; because $\frac{a_{\bar{n}}}{a_{\bar{n}-1}} \geq \frac{b_{\bar{n}}}{b_{\bar{n}-1}}$, this implies $a_{\bar{n}-1} \leq b_{\bar{n}-1}$; by the same logic, we get $a_n \leq b_n$ for every $n$, and therefore the only way for both $\{a_n\}$ and $\{b_n\}$ to sum to 1 is for $a = b$.

With these two results in hand, we will prove Lemma A6 by induction on the top of the support of $a$ and $b$. First, suppose $a$ and $b$ have support $\{0, 1\}$. As noted above, the LRO condition (combined with $a \neq b$) implies $a_1 > b_1$; so

$$\frac{a(s)}{b(s)} = \frac{a_1 s + 1 - a_1}{b_1 s + 1 - b_1}$$

and so

$$\frac{d}{ds}\left(\frac{a(s)}{b(s)}\right) \propto (b_1 s + 1 - b_1)a_1 - (a_1 s + 1 - a_1)b_1 = a_1 - b_1 > 0$$

so $a(s)/b(s)$ is strictly increasing, proving the base step.

For the inductive step, suppose the result holds for distributions with supports up to $N-1$, and now suppose $a$ and $b$ have support up to $N$, with $a \neq b$ and $a \geq_{lr} b$. Define $\hat{a}$ and $\hat{b}$ as the truncations of $a$ and $b$ to $n \leq N-1$, so that $\hat{a}_n = a_n/(1-a_N)$ and $\hat{b}_n = b_n/(1-b_N)$ for $n < N$, and note that $\hat{a} \geq_{lr} \hat{b}$. Let $\hat{a}(s) = \sum_{n=0}^{N-1} \hat{a}_n s^n$ and $\hat{b}(s) = \sum_{n=0}^{N-1} \hat{b}_n s^n$. Note that

$$\frac{a(s)}{b(s)} = \frac{a_N s^N + (1-a_N)\sum_{n=0}^{N-1} \frac{a_n}{1-a_N} s^n}{b_N s^N + (1-b_N)\sum_{n=0}^{N-1} \frac{b_n}{1-b_N} s^n} = \frac{a_N s^N + (1-a_N)\hat{a}(s)}{b_N s^N + (1-b_N)\hat{b}(s)} = \frac{a_N}{b_N} \frac{s^N + \frac{1-a_N}{a_N}\hat{a}(s)}{s^N + \frac{1-b_N}{b_N}\hat{b}(s)}$$

and, applying the quotient rule and dropping the denominator $b(s)^2 > 0$ and the constant $\frac{a_N}{b_N}$,

$$\begin{aligned}
\frac{d}{ds}\left(\frac{a(s)}{b(s)}\right) &\propto \left(N s^{N-1} + \tfrac{1-a_N}{a_N}\hat{a}'(s)\right)\left(s^N + \tfrac{1-b_N}{b_N}\hat{b}(s)\right) \\
&\quad -\left(s^N + \tfrac{1-a_N}{a_N}\hat{a}(s)\right)\left(N s^{N-1} + \tfrac{1-b_N}{b_N}\hat{b}'(s)\right) \\
&= N s^{2N-1} - s^N \tfrac{1-a_N}{a_N}\hat{a}'(s) + N s^{N-1}\tfrac{1-b_N}{b_N}\hat{b}(s) + \tfrac{1-a_N}{a_N}\tfrac{1-b_N}{b_N}\hat{a}'(s)\hat{b}(s) \\
&\quad - N s^{2N-1} - N s^{N-1}\tfrac{1-a_N}{a_N}\hat{a}(s) - s^N \tfrac{1-b_N}{b_N}\hat{b}'(s) - \tfrac{1-a_N}{a_N}\tfrac{1-b_N}{b_N}\hat{a}(s)\hat{b}'(s) \\
&= \tfrac{1-a_N}{a_N}\tfrac{1-b_N}{b_N}\left(\hat{a}'(s)\hat{b}(s) - \hat{a}(s)\hat{b}'(s)\right) \\
&\quad s^N\left(\tfrac{1-a_N}{a_N}\hat{a}'(s) - \tfrac{1-b_N}{b_N}\hat{b}'(s)\right) - N s^{N-1}\left(\tfrac{1-a_N}{a_N}\hat{a}(s) - \tfrac{1-b_N}{b_N}\hat{b}(s)\right).
\end{aligned}$$

Now, we noted above that $\hat{a} \geq_{lr} \hat{b}$, so by the inductive assumption, either $\hat{a} = \hat{b}$ (in which case the first line is 0), or $\hat{a}(s)/\hat{b}(s)$ is strictly increasing (in which case $\hat{a}'(s)\hat{b}(s) - \hat{a}(s)\hat{b}'(s) > 0$ and the first line is positive). And the second line has the same sign as the derivative of

$$-\frac{1}{s^N}\left(\frac{1-b_N}{b_N}\hat{b}(s) - \frac{1-a_N}{a_N}\hat{a}(s)\right)$$

So to prove $a(s)/b(s)$ is increasing, it suffices to show that $Q(s) \equiv \frac{1}{s^N}(\frac{1-b_N}{b_N}\hat{b}(s) - \frac{1-a_N}{a_N}\hat{a}(s))$ is decreasing. To prove this, we will define

$$A(s) = \frac{1}{s^N}\frac{1-b_N}{b_N}\hat{b}(s) \quad \text{and} \quad B(s) = 1 - \frac{\frac{1-a_N}{a_N}\hat{a}(s)}{\frac{1-b_N}{b_N}\hat{b}(s)}$$

and note that because $Q(s) = A(s)B(s)$, if $A(s)B(s)$ is strictly decreasing, then $a(s)/b(s)$ is strictly increasing and the lemma is proved.

To show that $A(s)B(s)$ is strictly decreasing, we first note that $B(s)$ is positive. By the inductive assumption, $\frac{\hat{a}(s)}{\hat{b}(s)}$ is either 1 (because $\hat{a} = \hat{b}$) or increasing in $s$; and $\hat{a}(1) = \hat{b}(1) = 1$, so $\frac{\hat{a}(s)}{\hat{b}(s)} \leq 1$ for all $s$. Further, because $a \neq b$ and $a \geq_{lr} b$, we noted earlier that $a_N > b_N$, and therefore $\frac{1-a_N}{a_N} / \frac{1-b_N}{b_N} < 1$; so $B(s) > 0$.

Next, note that

$$A(s) = \frac{1-b_N}{b_N}\sum_{n=0}^{N-1}\hat{b}_n s^{n-N}$$

and is therefore strictly decreasing in $s$. Finally, as noted above, $\hat{a}(s)/\hat{b}(s)$ is weakly increasing in $s$, and therefore $B(s)$ is weakly decreasing. Putting it together, then, $A(s)B(s)$ is strictly decreasing, and therefore $\frac{a(s)}{b(s)}$ is strictly increasing, proving Lemma A6, which completes the proof of Theorem 2. □

□ **Proof of Theorem 3.** Let $\bar{N}$ denote the upper bound on the support of $N$. For each $n \leq \bar{N}$, an entry model uniquely defines the probabilities $\Pr(N_{\text{obs}} = k|N = n)$ for each $k \leq n$; this probability is zero for $k > n$. Define the matrix

$$Z = \begin{bmatrix} \Pr(N_{\text{obs}} = 2|N = 2) & \Pr(N_{\text{obs}} = 2|N = 3) & \Pr(N_{\text{obs}} = 2|N = 4) & \cdots & \Pr(N_{\text{obs}} = 2|N = \bar{N}) \\ 0 & \Pr(N_{\text{obs}} = 3|N = 3) & \Pr(N_{\text{obs}} = 3|N = 4) & \cdots & \Pr(N_{\text{obs}} = 3|N = \bar{N}) \\ 0 & 0 & \Pr(N_{\text{obs}} = 4|N = 4) & \cdots & \Pr(N_{\text{obs}} = 4|N = \bar{N}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Pr(N_{\text{obs}} = \bar{N}|N = \bar{N}) \end{bmatrix}.$$

Because $\Pr(N_{\text{obs}} = k) = \sum_{n \geq k} \Pr(N = n) \Pr(N_{\text{obs}} = k|N = n)$, straightforward matrix algebra establishes

$$\begin{bmatrix} \Pr(N_{\text{obs}} = 2) \\ \Pr(N_{\text{obs}} = 3) \\ \vdots \\ \Pr(N_{\text{obs}} = \bar{N}) \end{bmatrix} = Z \begin{bmatrix} \Pr(N = 2) \\ \Pr(N = 3) \\ \vdots \\ \Pr(N = \bar{N}) \end{bmatrix}.$$

$Z$ is upper triangular, and by assumption, has strictly positive diagonal elements, which means it is invertible, so

$$\begin{bmatrix} \Pr(N = 2) \\ \Pr(N = 3) \\ \vdots \\ \Pr(N = \bar{N}) \end{bmatrix} = Z^{-1} \begin{bmatrix} \Pr(N_{\text{obs}} = 2) \\ \Pr(N_{\text{obs}} = 3) \\ \vdots \\ \Pr(N_{\text{obs}} = \bar{N}) \end{bmatrix}$$

so the distribution of $N$ is unqiuely determined by the entry model ($Z$) and observables (the distribution of $N_{\text{obs}}$). □

□ **Counterexample to Theorem 3 with unbounded support.** As noted in the text, the distribution of $N$ is not necessarily identified from $N_{\text{obs}}$ if it has unbounded support. For a simple counterexample, suppose that

$$\Pr(N_{\text{obs}} = k|N = n) = \begin{cases} 1 & \text{if} k = n = 2 \\ \frac{1}{3} & \text{if} n > 2 \text{and} k = n \\ \frac{2}{3} & \text{if} n > 2 \text{and} k = n - 1 \\ 0 & \text{otherwise} \end{cases}$$

and the distribution of $N_{\text{obs}}$ is

$$\Pr(N_{\text{obs}} = n) = \frac{1}{2^{n-1}}$$

for $n \geq 2$. It is straightforward to verify that this distribution of $N_{\text{obs}}$ is consistent with either of the following two distributions of $N$:

$$\Pr(N = n) = \begin{cases} \frac{1}{2} & \text{if} n = 2 \\ \frac{3}{2^{n-1}} & \text{if} n > 2 \text{and} n \text{even} \\ 0 & \text{otherwise} \end{cases} \quad \text{or} \quad \begin{cases} \frac{3}{4} & \text{if} n = 3 \\ \frac{3}{2^{n-1}} & \text{if} n > 3 \text{and} n \text{odd} \\ 0 & \text{otherwise} \end{cases}$$

(or any mixture of the two), so the distribution of $N$ is not identified from $N_{\text{obs}}$.

□ **Proof of Theorem 4.** Given bounded support and knowledge of the entry model, Theorem 3 says that the distributions of $N|X = x$ and $N|X = x'$ are identified from the distributions of $N_{\text{obs}}|X = x$ and $N_{\text{obs}}|X = x'$, respectively, which are identified from $(N_{\text{obs}}, X)$. Once the distributions of $N|X = x$ and $N|X = x'$ are known, Theorem 2 says the rest of the model is identified given $(T, X)$. □

□ **Simplifying the likelihood function when $N$ has known distribution.** We would like to use the density function of $T$ (for each realization of $X$) as the likelihood function for empirical estimation of $f_\epsilon$ and $f_\theta$. Given a probability distribution $\mathbf{p}(\cdot|x)$ for $n$, the density function is

$$f_{T|x}(t) = \sum_{n=2}^{\infty} \mathbf{p}(n|x) f_{T|n}(t) = \sum_{n=2}^{\infty} \mathbf{p}(n|x) \int_0^t f_\theta(t - s) f_{y|n}(s) ds,$$

where $f_{y|n}$ is the distribution of the second highest of $n$ independent draws from $F_\epsilon$, which is $f_{y|n}(s) = n(n - 1)F_\epsilon^{n-2}(s)(1 - F_\epsilon(s))f_\epsilon(s)$. Moving the sum inside the integral,

$$f_{T|x}(t) = \int_0^t f_\theta(t - s) \sum_{n=2}^{\infty} \mathbf{p}(n|x) n(n - 1) F_\epsilon^{n-2}(s)(1 - F_\epsilon(s)) f_\epsilon(s) ds$$

but this still involves both an integral and an infinite sum (if the distribution of $N$ is parameterized with a distribution with unbounded support). However, for two common parameterizations of the distribution of $N$, we can explicitly evaluate the sum, greatly simplifying the expression of the likelihood function.

For the following results, we assume $N$ is drawn from the truncation of a well-known distribution over $\mathbb{Z}^+$ to $N \geq 2$, as auctions with zero or one bidder would not generate positive prices and would therefore not appear in the data. (Platt (2017) gives an analogous simplification for the CDF of $T$ in the Poisson case without unobserved heterogeneity.)

*Theorem 5.* If conditional on $X = x$, $N$ is drawn from a truncated Poisson distribution with parameter $\lambda > 0$, then the density function simplifies to

$$f_{T|x}(t) = \frac{\lambda^2 e^{-\lambda}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} \int_0^t f_\theta(t-s)(1 - F_\epsilon(s)) f_\epsilon(s) e^{\lambda F_\epsilon(s)} ds.$$

If conditional on $X = x$, $N$ is drawn from a truncated (generalized) negative binomial distribution with parameters $r > 0$ and $p \in (0, 1)$, then the density function simplifies to

$$f_{T|x}(t) = \frac{(r+1)rp^2(1-p)^r}{1-(1-p)^r - rp(1-p)^r} \int_0^t f_\theta(t-s)(1 - F_\epsilon(s))(1 - pF_\epsilon(s))^{-(r+2)} f_\epsilon(s) ds.$$

*Proof of Theorem* 5. For the truncated Poisson case, the distribution of $y$ is

$$\begin{aligned}
f_{y|x}(s) &= \sum_{n=2}^{\infty} \frac{\frac{\lambda^n e^{-\lambda}}{n!}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} \left( n(n-1) F_\epsilon^{n-2}(s)(1 - F_\epsilon(s)) f_\epsilon(s) \right) \\
&= \frac{e^{-\lambda}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} \left( \lambda^2 (1 - F_\epsilon(s)) f_\epsilon(s) \right) \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} F_\epsilon^{n-2}(s) \\
&= \frac{e^{-\lambda}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} \frac{\lambda^2 (1 - F_\epsilon(s)) f_\epsilon(s)}{e^{-\lambda F_\epsilon(s)}} \sum_{n=2}^{\infty} \frac{(\lambda F_\epsilon(s))^{n-2} e^{-\lambda F_\epsilon(s)}}{(n-2)!} \\
&= \frac{\lambda^2 e^{-\lambda}}{1 - e^{-\lambda} - \lambda e^{-\lambda}} (1 - F_\epsilon(s)) f_\epsilon(s) e^{\lambda F_\epsilon(s)} \sum_{n'=0}^{\infty} \frac{(\lambda')^{n'} e^{-\lambda'}}{(n')!},
\end{aligned}$$

where $n' = n - 2$ and $\lambda' = \lambda F_\epsilon(s)$. The sum is now simply the sum of probabilities of a (different) untruncated Poisson distribution, and is therefore 1; plugging $f_{y|x}(s)$ back into the convolution expression for $f_{T|x}(t)$ gives the result.

For the negative binomial case (with $r$ any positive real number),

$$\begin{aligned}
f_{y|x}(s) &= \sum_{n=2}^{\infty} \frac{\frac{\Gamma(r+n)}{n!\Gamma(r)} p^n (1-p)^r}{1 - (1-p)^r - rp(1-p)^r} n(n-1) F_\epsilon^{n-2}(s)(1 - F_\epsilon(s)) f_\epsilon(s) \\
&= \frac{(1-p)^r (1 - F_\epsilon(s)) f_\epsilon(s)}{1 - (1-p)^r - rp(1-p)^r} \sum_{n=2}^{\infty} \frac{\Gamma(r+n)}{n!\Gamma(r)} p^n n(n-1) F_\epsilon^{n-2}(s) \\
&= \frac{(1-p)^r (1 - F_\epsilon(s)) f_\epsilon(s)}{1 - (1-p)^r - rp(1-p)^r} \frac{\frac{\Gamma(r+2)}{\Gamma(r)} p^2}{(1 - pF_\epsilon(s))^{r+2}} \sum_{n=2}^{\infty} \frac{\Gamma(r+n)}{(n-2)!\Gamma(r+2)} p^{n-2} F_\epsilon^{n-2}(s)(1 - pF_\epsilon(s))^{r+2} \\
&= \frac{(1-p)^r (1 - F_\epsilon(s)) f_\epsilon(s)}{1 - (1-p)^r - rp(1-p)^r} \frac{r(r+1)p^2}{(1 - pF_\epsilon(s))^{r+2}} \sum_{n'=0}^{\infty} \frac{\Gamma(r'+n')}{n'!\Gamma(r')} (p')^{n'} (1-p')^{r'},
\end{aligned}$$

where $n' = n - 2$, $r' = r + 2$, and $p' = pF_\epsilon(s)$. Once again, the sum is now the sum of probabilities of a different untruncated negative binomial distribution, which therefore sum to 1; plugging the remaining expression into the convolution equation for $f_{T|x}(t)$ gives the result. □

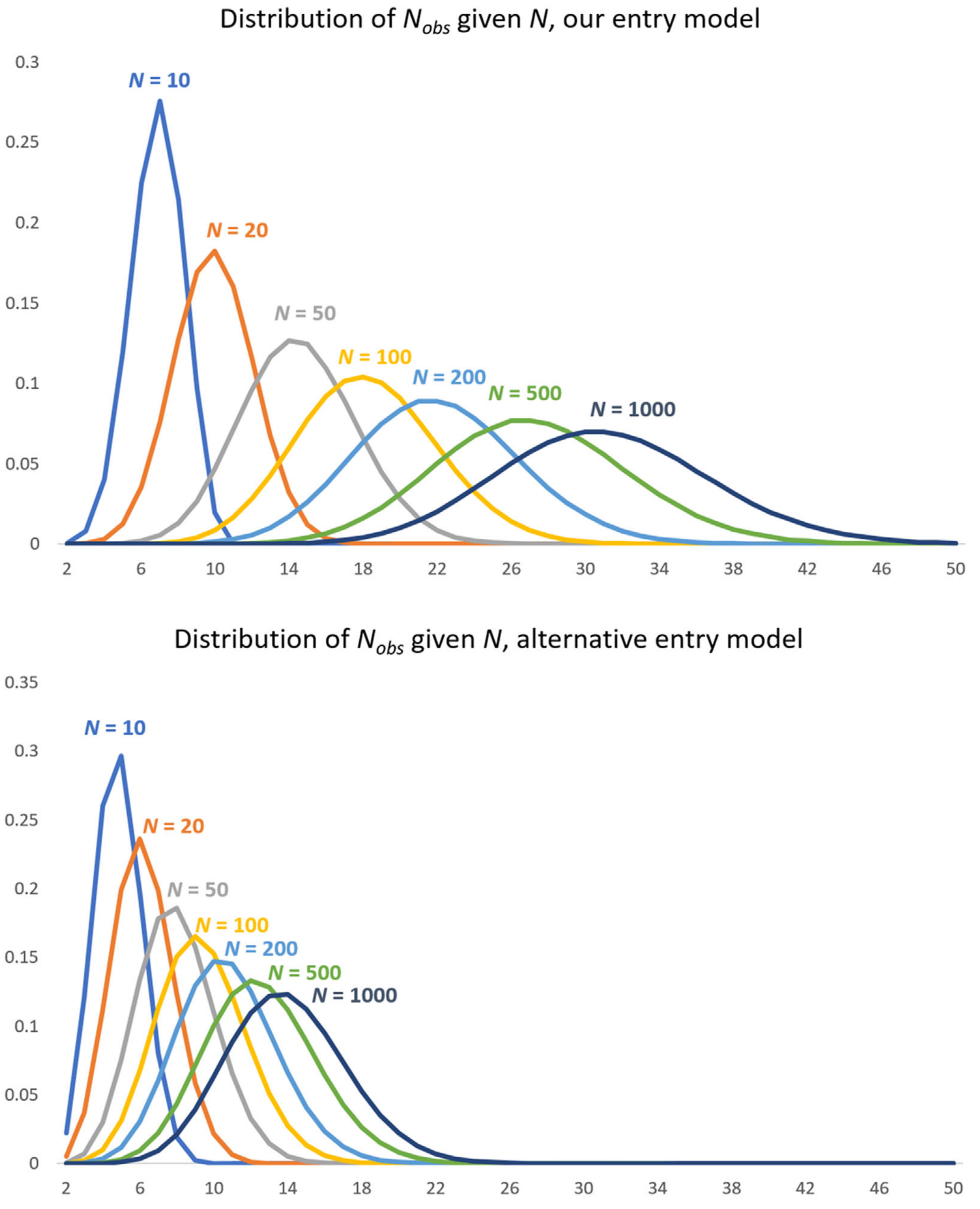
□ **A note on a different entry model.** As noted in the text (footnote 15), before settling on the entry model used in the article, we first tried the following one, equivalent to the one used in Hickman, Hubbard, and Paarsch (2017), Platt (2017), and Bodoh-Creed, Boehnke, and Hickman (2018b). Suppose the $N$ potential bidders in an auction arrive in random order, and each one, when she arrives, submits a proxy bid equal to her valuation (if it is above the current standing high bid). Under this model, the standing high bid at any time is the second-highest valuation among the bidders who have already arrived, and so the number of observed bids $N_{\text{obs}}$ is the number of bidders who, at the time they arrived, had either the highest or second-highest valuation so far. Conveniently, this allows one to iteratively calculate each value $\Pr(N_{\text{obs}} = k | N = j)$, as follows. (Recall that we assume $N \geq 2$, as auctions with zero or one bidder would have been dropped from the data.)

*Lemma A7.* For the entry model described above,

(1) $\Pr(N_{\text{obs}} = k | N = j) = 0$ for any $k > j \geq 2$.
(2) $\Pr(N_{\text{obs}} < 2 | N = j) = 0$ for any $j \geq 2$.
(3) $\Pr(N_{\text{obs}} = 2 | N = 2) = 1$.

FIGURE 6

COMPARISON OF TWO CANDIDATE ENTRY MODELS [Color figure can be viewed at wileyonlinelibrary.com]




(4) For any $j \geq k \geq 2$,

$$\Pr(N_{\text{obs}} = k | N = j) = \frac{j-2}{j} \Pr(N_{\text{obs}} = k | N = j-1) + \frac{2}{j} \Pr(N_{\text{obs}} = k-1 | N = j-1).$$

The first three points are trivial given the setup of the entry model. For the fourth, note that if there are $N = j$ potential bidders, the probability the last bidder to arrive will submit a bid is equal to $\frac{2}{j}$, the probability she has one of the highest two valuations. In this event, $N_{\text{obs}}$ will be equal to $k$ if $k-1$ of the first $j-1$ bidders to enter cast bids, which occurs with probability $\Pr(N_{\text{obs}} = k-1|N = j-1)$. On the other hand, with probability $\frac{j-2}{j}$, the last bidder will not bid, which means $N_{\text{obs}} = k$ if $k$ of the first $j-1$ bidders to enter cast bids, which occurs with probability $\Pr(N_{\text{obs}} = k|N = j-1)$. Lemma A7 allows us to fully populate the matrix $[\Pr(N_{\text{obs}} = k|N = j)]$ iteratively, rather than relying on simulation to calculate it.

The ability to calculate closed-form expressions for each term $\Pr(N_{\text{obs}} = k|N = j)$ is appealing, as is the fact that these do not depend on the other details of the environment ($f_\theta$ and $f_\epsilon$). In the end, however, we chose not to use this entry model, as it requires a very high number of actual bidders $N$ to generate levels of $N_{\text{obs}}$ seen in our data. As Figure 2 illustrates, our data contains a number of auctions with 18 to 22 observed bidders. Figure 6 illustrates the probability distribution of $N_{\text{obs}}$ for various values of $N$, for both the entry model in our article (top pane) and the model just described (bottom pane). To "explain" auctions with 18 to 22 observed bidders, the latter model would require values of $N$ in the several hundreds, which seemed intuitively unrealistic to us; under the model we settled on, observations like this could occur with $N$ on the order of 50. The model we chose also has the nice feature of predicting bidding both throughout the auction (as bidders arrive) and near the end of the auction (updated bids from "serious" contenders), and multiple bids from the same bidder, all of which occur routinely in our data (see footnote 23).

## References


Ackerberg, D., Chen, X., and Hahn, J. "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators." *Review of Economics and Statistics*, Vol. 94 (2012), pp. 481–498.

Aradillas-López, A., Gandhi, A, and Quint, D. "Identification and Inference in Ascending Auctions with Correlated Private Values." *Econometrica*, Vol. 81 (2013), pp. 489–534.

Aradillas-López, A., Gandhi, A, and Quint, D. "A Simple Test for Moment Inequality Models with an Application to English Auctions." *Journal of Econometrics*, Vol. 194 (2016), pp. 96–115.

Aryal, G., Grundl, S., kim, D., and Zhu, Y. "Empirical Relevance of Ambiguity in First-Price Auctions." *Journal of Econometrics*, Vol. 204 (2018), pp. 189–206.

Athey, S. and Haile, P. "Identification of Standard Auction Models." *Econometrica*, Vol. 70 (2002), pp. 2107–2140.

Athey, S. and Haile, P. "Empirical Models of Auctions." Technical Report, National Bureau of Economic Research (2006).

Athey, S., Levin, J., and Seira, E. "Comparing Open and Sealed Bid Auctions: Evidence From Timber Auctions." *Quarterly Journal of Economics*, Vol. 126 (2011), pp. 207–257.

Bajari, P. and Hortaçsu, A. "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data." *Journal of Political Economy*, Vol. 113 (2005), pp. 703–741.

Bodoh-Creed, A., Boehnke, J., and Hickman, B. "Using Machine Learning to Explain Price Dispersion." Working paper, 2018a.

Bodoh-Creed, A., Boehnke, J., and Hickman, B. "How Efficient Are Decentralized Auction Platforms?" Working paper, 2018b.

Chen, X. and Liao, Z. "Sieve *M* Inference on Irregular Parameters." *Journal of Econometrics*, Vol. 182 (2014), pp. 70–86.

Coey, D., Larsen, B., Sweeney, K., and Waisman, C. "Ascending Auctions with Bidder Asymmetries." *Quantitative Economics*, Vol. 8 (2017), pp. 181–200.

Compiani, G., Haile, P., and Sant'Anna, M. "Common Values, Unobserved Heterogeneity, and Endogenous Entry in U.S. Offshore Oil Lease Auctions." Working paper, 2019.

Decarolis, F. "Comparing Public Procurement Auctions." *International Economic Review*, Vol. 59 (2018), pp. 391–419.

Dyer, D., Kagel, J., and Levin, D. "Resolving Uncertainty about the Number of Bidders in Independent Private-Value Auctions: An Experimental Analysis." *RAND Journal of Economics*, Vol. 20 (1989), pp. 268–279.

Fox, J., Kim, K., Ryan, S., and Bajari, P. "The Random Coefficients Logit Model is Identified." *Journal of Econometrics*, Vol. 166 (2012), pp. 204–212.

Freyberger, J. and Larsen, B. "Identification in Ascending Auctions, with an Application to Digital Rights Management." Working paper, 2017.

Gallant, R. and Nychka, D. "Semi-Nonparametric Maximum Likelihood Estimation." *Econometrica*, Vol. 55 (1987), pp. 363–390.

Gillen, B. "Identification and Estimation of Level-*k* Auctions." Working paper, 2010.

Guerre, E., Perrigne, I., and Vuong, Q. "Optimal Nonparametric Estimation of First-Price Auctions." *Econometrica*, Vol. 68 (2000), pp. 525–574.

Guerre, E., Perrigne, I., and Vuong, Q. "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions." *Econometrica*, Vol. 77 (2009), pp. 1193–1227.

Haile, P., Hong, H., and Shum, M. "Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions." Working paper, 2003.

Haile, P. and Tamer, E. "Inference With an Incomplete Model of English Auctions." *Journal of Political Economy*, Vol. 111 (2003), pp. 1–51.

Harter, L. "Expected Values of Normal Order Statistics." *Biometrika*, Vol. 48 (1961), pp. 151–165.

Hickman, B., Hubbard, T., and Paarsch, H. "Identification and Estimation of a Bidding Model for Electronic Auctions." *Quantitative Economics*, Vol. 8 (2017), pp. 505–551.

Hu, Y., McAdams, D., and Shum, M. "Identification of First-Price Auctions with Non-Separable Unobserved Heterogeneity." *Journal of Econometrics*, Vol. 174 (2013), pp. 186–193.

Jank, W. and Shmueli, G. *Modeling Online Auctions*. Hoboken, NJ: Wiley, 2010.

Komarova, T. "Partial Identification in Asymmetric Auctions in the Absence of Independence." *Econometrics Journal*, Vol. 16 (2013), pp. S60–S92.

Konstantopoulos, T. and Yuan, L. "Does the Ratio of Laplace Transforms of Powers of a Function Identify the Function?" Working paper, 2019.

Krantz, S. and Parks, H. *A Primer of Real Analytic Functions*, second edition. Basel: Berkhaüser, 2002.

Krasnokutskaya, E. "Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity." *Review of Economic Studies*, Vol. 78 (2011), pp. 293–327.

Levin, D. and Smith, J. "Equilibrium in Auctions with Entry." *American Economic Review*, Vol. 84 (1994), pp. 585–599.

Lewis, G. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors." *American Economic Review*, Vol. 101 (2011), pp. 1535–1546.

Li, T., Perrigne, I., and Vuong, Q. "Conditionally Independent Private Information in OCS Wildcat Auctions." *Journal of Econometrics*, Vol. 98 (2000), pp. 129–161.

Li, T., Perrigne, I., and Vuong, Q. "Structural Estimation of the Affiliated Private Value Auction Model." *RAND Journal of Economics*, Vol. 33 (2002), pp. 171–193.

Li, T. and Vuong, Q. "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators." *Journal of Multivariate Analysis*, Vol. 65 (1998), pp. 139–165.

Liu, N. and Luo, Y. "A Nonparametric Test for Comparing Valuation Distributions in First-Price Auctions." *International Economic Review*, Vol. 58 (2017), pp. 857–888.

Lu, J. "Entry Coordination and Auction Design with Private Costs of Information Acquisition." *Economic Inquiry*, Vol. 48 (2010), pp. 274–289.

Lu, J. and Ye, L. "Efficient and Optimal Mechanisms with Private Information Acquisition Costs." *Journal of Economic Theory*, Vol. 148 (2013), pp. 393–408.

Luo, Y. and Xiao, R. "Identification of Auction Models Using Order Statistics." Working paper, 2019.

Mbakop, E. "Identification of Auctions with Incomplete Bid Data in the Presence of Unobserved Heterogeneity." Working paper, 2017.

Moreno, D. and Wooders, J. "Auctions with Heterogeneous Entry Costs." *RAND Journal of Economics*, Vol. 42 (2011), pp. 313–336.

Platt, B. "Inferring Ascending Auction Participation from Observed Bidders." *International Journal of Industrial Organization*, Vol. 54 (2017), pp. 65–88.

Quint, D. "Identification in Symmetric English Auctions with Additively Separable Unobserved Heterogeneity." Working paper, 2015 (supplanted by this article).

Roberts, J. "Unobserved Heterogeneity and Reserve Prices in Auctions." *RAND Journal of Economics*, Vol. 44 (2013), pp. 712–732.

Song, U. "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders." Working paper, 2004.