Electronic Journal of Statistics
Vol. 18 (2024) 2477–2530
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2258

# Order statistics approaches to unobserved heterogeneity in auctions

**Yao Luo**[1], **Peijun Sang**[2] **and Ruli Xiao**[3]

[1] *Department of Economics, University of Toronto,*
*e-mail:* yao.luo@utoronto.ca

[2] *Department of Statistics and Actuarial Science, University of Waterloo,*
*e-mail:* psang@uwaterloo.ca

[3] *Department of Economics, Indiana University,*
*e-mail:* rulixiao@iu.edu

**Abstract:** We establish nonparametric identification of auction models with continuous and nonseparable unobserved heterogeneity using three consecutive order statistics of bids. We then propose sieve maximum likelihood estimators for the joint distribution of the unobserved heterogeneity and the private value, as well as their conditional and marginal distributions. Lastly, we apply our methodology to a novel dataset from judicial auctions in China. Our estimates suggest substantial gains from accounting for unobserved heterogeneity when setting reserve prices. We propose a simple scheme that achieves nearly optimal revenue by using the appraisal value as the reserve price.

**MSC2020 subject classifications:** 91B26, 62G30, 62P20.
**Keywords and phrases:** Sieve estimation, nonseparable, measurement error, consecutive order statistics, judicial auctions.

Received January 2024.

## 1. Introduction

Empirical auction literature often estimates the underlying bidder value distributions using the recorded bids to answer some counterfactual questions to inform policy recommendations about auction design such as optimal reserve prices [32], limiting competition to soften the winner's curse [13], the choice of auction format [39], bid discounts or entry subsidies [33], and quantifying damages and inefficiency of collusion [5]. See also, e.g., [26].

A prevalent challenge to identifying and estimating the underlying bidder value distribution is that bidders often have more information about the item for sale than the researcher, resulting in auction-level unobserved heterogeneity (UH). Ignoring this UH wrongly attributes auction-level variations to bidder value distribution dispersion, leading to overestimating the variation in bidder values and, hence, information rents to bidders and poor policy recommendations about auction design. For instance, [25] finds that UH accounts for two-thirds of price variation after controlling for information provided in the eBay Motors auctions and that ignoring this feature would dramatically mis-estimate the welfare measures. See also discussion in, e.g., [28], [44] and [38].

The existing literature adapts measurement error approaches to tackle such an issue. Suppose the analyst observes all bids. The analyst could then identify the value distribution using observed bids as measurements for the unobserved characteristics since these bids are independent conditional on such unobserved characteristics.

However, this conditional independence condition fails when the analyst only observes incomplete bid data. This could occur for various reasons. First, in English or ascending outcry auctions, the bidder with the highest value only needs to outbid the bidder with the second-highest value to win, which means the recorded bids do not contain the highest value. Moreover, even in first-price sealed-bid auctions, where all bids are supposed to be submitted to the auctioneer, the auctioneer may still not record all the bids in practice: sometimes the auctioneer only records the most competitive bids, such as the top three bids in regular auctions or apparent low bids in procurement auctions. Thus, the econometrician can only observe a few order statistics of the bids, i.e., incomplete bid information. For instance, the U.S. Forest Service timber auctions only record at most the top 12 bids regardless of the number of bidders. The Washington State Department of Transportation provides an online archive of bid opening results that are six months or older, but only for the top three apparent low bids. Even if the auctioneer records all bids, the most competitive bids are often more accessible to the public. For instance, The Federal Deposit Insurance Corporation resolves insolvent banks using first-price auctions but only publishes the top two bids and bidders' identities [1]. The three apparent low bids are one-click downloadable on the website of the California Department of Transportation. These order statistics are naturally dependent, invalidating conventional identification strategies.

We provide identification results for auction models using order statistics of bids. We make three contributions in this paper. First, our paper is the first to study identification of auction models with *continuous and nonseparable* UH using incomplete bid data. Our specification allows for flexibility in how UH affects both bidder value and the equilibrium bidding strategy − namely the mapping from a bidder's private value to his/her bid.[1]

Our identification strategy adapts [29] for nonclassical measurement error models to the auction setting. This extension is nontrivial in that we only observe order statistics of UH-contaminated bids. As a result, we cannot achieve a parsimonious conditional independence structure as in their work.[2] Instead, we follow [40] and consider the most common case of incomplete bid data: *consecutive* order statistics of bids. Their main insight is that consecutive order statistics have a semi-multiplicatively separable joint distribution with a simple indicator function capturing the correlation. Unlike both papers using two measurements with an instrument, we use *three* consecutive order statistics of bids. Given a partition on the range of the measurements, we again obtain a separa-

[1]Even if one assumes separable UH in the value, separability passing to the bid often requires additional institutional features or assumptions. See, e.g., [3].

[2]They assume that the outcome variable is independent of the observed independent variable and an instrument conditional on the unobserved true regressor.

ble structure traditionally achieved under conditional independence. This turns the identification problem into an operator diagonalization problem, allowing constructive identification arguments using linear operator tools. Moreover, we use these tools differently by considering bounded linear operators defined on a Hilbert space and taking values in another Hilbert space. This space differs from the $\mathcal{L}^1$ space adopted in [29], which focuses on a Banach space. While we could also work with Banach space, using Hilbert space simplifies the analysis of relevant operators and thus our proofs thanks to many existing theoretical results.[3]

Second, we propose sieve maximum likelihood estimators (MLE) of the model primitives and provide conditions that guarantee their consistency. The estimation of auction models allows for counterfactual policy analysis, such as computing the optimal reserve price. If UH is common knowledge among agents in the auction, it is a critical control in policy analysis. Therefore, optimal policy recommendation requires estimating the joint distribution of UH and bidder private value.[4] In particular, we approximate the joint density of bids and UH using the tensor product of two univariate sieve bases. We then represent the marginal density of the UH and the conditional distribution of the value using the sieve-approximated joint distribution. Therefore, these distributions are all estimated nonparametrically.[5] [29] proposes sieve approximations to the conditional distribution and marginal distribution. Our sieve approximation to the joint distribution is more convenient as we just need to impose the normalization assumption on the joint distribution approximation once.

The consistency of our estimator relies on the condition that the sieve space approximates well the joint distribution of bids and UH. To formalize this intuition, we quantify the complexity of this space using bracket entropy and prove the consistency of the sieve MLEs for the joint, conditional, and marginal densities. We establish a concentration inequality based on the bracketing number, a similar notation to covering numbers used in [29]. In Appendix B.2 we further investigate the properties of B-splines and Bernstein polynomials, both of which are popular in empirical applications.

Lastly, we apply our identification and estimation method to a novel dataset from judicial auctions conducted by a municipal court in China. By default, this court uses 70% of the appraisal value as the starting price, which also serves as a reserve price. Our estimation results suggest substantial gains from accounting for UH when designing reserve prices. The court can gain 5.81% more revenue using an optimal reserve price for each item. However, this scheme is complex; the seller would need to know UH and recover the conditional density of bidder values. Instead, we propose a simple scheme that achieves nearly optimal revenue by using the appraisal value as the reserve price. Specifically, using the estimated

[3]For instance, it is straightforward to define the adjoint operator by using the concept of inner product in Hilbert spaces.

[4]Since there is a known mapping between the bid distribution and the value distribution, we will use the two terms interchangeably. See [21] and [6] for this mapping.

[5]In contrast, previous research only focuses on the estimation of the joint distribution using a semiparametric structure [8] and [29] or a nonparametric structure [49].

model, we find that using the appraisal value as the reserve price achieves 98.85% of the potential gains from the optimal reserve prices.

### *Literature review*

The auction literature has widely applied techniques developed in the measurement error literature for identifying auction models with UH. If the UH is continuous and has a separable structure on bidder valuations, identification relies on the deconvolution approach and requires two random bids for each auction. See [35], [34], and [32], among others. If the UH is finite and discrete, which by nature is nonseparable, identification relies on the condition that the bids are independent conditional on the UH and requires three random bids for each auction. See [27], [28], and [38].

Moreover, the literature has seen rapid growth in identifying and estimating auction models using order statistics of bids. [6] shows that symmetric independent private value (IPV) auctions are identifiable by the transaction price and the number of bidders using the one-to-one mapping between the distribution of an order statistic and its parent distribution; [30] identifies asymmetric second-price auctions using the winner's identity and the transaction price; [20] shows that IPV first-price auctions without observable competition is identifiable using the transaction price; [42] studies large sample properties for nonparametric estimators using order statistics of bids.

A growing literature tackles the identification of auction models with UH and incomplete bid information. Assuming the UH is finite and discrete, [41] provides identification results from (any) five order statistics to restore the conditional independence condition by the Markov property of order statistics. [40] provides an alternative identification strategy using two *consecutive* order statistics of bids and an instrument. Finiteness simplifies their identification arguments because model restrictions can be written in matrix algebra. In contrast, we use linear operators, which is not a trivial extension of the matrix operations. Moreover, we extend our identification results to allow for binding reserve prices and apply them in our empirical application.

In the framework of additively separable continuous UH, [25] achieves point identification using English auction models, assuming piecewise real analytic density functions and using variations in the number of bidders across auctions; [17] provides identification results for ascending auctions, relying on reserve prices and two order statistics of bids. [10] studies deconvolution using two order statistics. Our paper is the first to show point identification of auction models with continuous and nonseparable UH using incomplete bid data.

The remainder of this paper is organized as follows. Section 2 presents our main identification results. Section 3 proposes sieve maximum likelihood estimators. We showcase the finite-sample performance of the proposed estimator in Section 4. Section 5 presents an application to judicial auctions in China. Section 6 concludes. The Appendix contains detailed proofs of the identification results and the asymptotic properties.

## 2. Main identification results

### *2.1. Identification assumptions and steps*

For simplicity, we abstract from observable (to the analyst) characteristics. Suppose $n \geq 2$ symmetric bidders participate in an auction with zero reserve price.[6] All bidders observecharacteristic $\mathtt{T}$ before they submit bids.[7] We assume that the unobserved characteristic $\mathtt{T}$ is continuous.[8] Our identification strategy applies regardless of whether the seller observes $\mathtt{T}$ or not. Among $n$ potential bidders, bidder $i$, where $i = 1, ..., n$, draws his/her value $V_i$ from the conditional value distribution $f^{V|\mathtt{T}}(v|\tau)$ and submits a bid $X_i$. We consider the situation wherein the latent auction characteristic and bids/values are continuous. We denote the marginal distribution of the latent characteristic $\mathtt{T}$ as $f^{\mathtt{T}}(\tau)$ and the optimal conditional bid distribution as $f^{X|\mathtt{T}}(x|\tau)$, where $x$ is the optimal bid.

We first introduce the standard assumption regarding the value distribution.

**Assumption 1.** *(Conditional Independence) Bidder values, $V_1$,..., $V_n$, are i.i.d. conditional on the auction-level heterogeneity* $\mathtt{T}$.[9]

In a first-price auction, the bidder with the highest bid wins and pays his own bid price. [21] provides a one-to-one mapping between the conditional value distribution $f^{V|\mathtt{T}}(v|\tau)$ and the conditional bid distribution $f^{X|\mathtt{T}}(x|\tau)$ given that the competition $n$ is known. Thus, the identification of the conditional value distribution boils down to recovering the conditional bid distribution $f^{X|\mathtt{T}}(x|\tau)$ from the bid data. If the data record all bids in each auction, the conditional independence property passes from values to bids. Consequently, the joint distribution of three independent bids, e.g., $X_1$, $X_2$, and $X_3$, denoted as $f(x, y, z)$, has the following multiplicatively separable structure:

$$f(x, y, z) = \int_{\mathcal{T}} \underbrace{f^{X|\mathtt{T}}(x|\tau) f^{X|\mathtt{T}}(y|\tau) f^{X|\mathtt{T}}(z|\tau)}_{\text{repeated measurements}} f^{\mathtt{T}}(\tau) d\tau, \tag{1}$$

based on which the conditional densities $f^{X|\mathtt{T}}(x|\tau)$ can be identified via eigenfunction decomposition [29]. The main idea is to exploit the property that the

[6] We assume the number of potential bidders is known. Otherwise, we can treat it as an additional dimension of UH, as in [40], or construct it through alternative data sources. In procurement auctions, we can construct it using the number of qualified firms in the local market via public information, such as the list of qualified firms and their contact information.

[7] While we focus on regular auctions here, our results extend trivially to procurement auctions.

[8] UH may arise for many reasons. Discrete examples include an unknown number of bidders ([40]), implicit reserve prices or bidding costs ([28]), unknown bidder types or bounded rationality ([2]), and multiple equilibria ([50], [38]). Continuous ones include traffic intensity and quality of the existing surface in highway procurement auctions ([32]), unobserved quality of timber ([23]), and rust, dents, and tire quality of second-hand automobiles ([44]).

[9] In some settings, heterogeneity across bidders might be more interesting. It is worth noting that common unobserved auction-level factors with homogeneous bidders are also prevalent in the existing literature. See [26]. Moreover, one can test whether the bidders' value distributions are symmetric or asymmetric following [36].

recorded bids are repeated measurements of UH. Under Assumption 1, their correlation reveals how UH affects the bids. Specifically, the observed joint distribution on the left-hand side of (1) identifies the conditional and marginal distributions on the right-hand side.

Unfortunately, the auctioneer often does not record all bid information, and instead only records the most competitive bids. That is, the data essentially record a few order statistics of all bids, under which the conditional independence condition fails to hold. This is because order statistics are ordered by definition.

In an ascending auction, the bidder with the highest bid wins and pays the second-highest submitted price, so a weakly dominant strategy is to continue bidding until the standing bid reaches one's value. Therefore, all bidders bid their values except the one with the highest value, who can simply outbid the second-highest value by a small amount. That is, the highest bid and the second highest bid reveal essentially the same information regarding the second highest value, indicating that the highest bid is redundant. Because of this particular auction format, it is impossible to observe the highest value from the bids. Equivalently, we can view the auction as everyone bids her/his value, but the auction fails to observe the highest bid/value. Consequently, one cannot follow the aforementioned identification results to recover the conditional value distribution $f^{V|\mathtt{T}}(v|\tau)$, because the conditional independence condition fails.

Facing the data limitation of incomplete bids, this paper focuses on identifying the conditional bid distribution $f^{X|\mathtt{T}}(x|\tau)$ for both first-price and ascending auctions from any three consecutive order statistics of all bids, i.e., $\{X_{r-2:n}, X_{r-1:n}, X_{r:n}\}$, where $X_{r-2:n} \leq X_{r-1:n} \leq X_{r:n}$. Once the conditional bid distribution is identified, the conditional value distribution can be identified using the one-to-one mapping between the bid and the value.

Let $\mathcal{V}$, $\mathcal{X}$, and $\mathcal{T}$ denote the supports of the distributions of the random variables $V$, $X$, and $\mathtt{T}$, respectively. We first introduce the following regularity assumption.

**Assumption 2.** *(Bound and Continuity) The joint density of $X$ and $\mathtt{T}$ admits a bounded and continuous density with respect to the product measure of some dominating measure $\mu$ (defined on $\mathcal{X}$) and the Lebesgue measure on $\mathcal{T}$. All marginal and conditional densities are also bounded and positive.*

We use $f_{r-2,r-1,r:n}(\cdot)$ and $f_{r-2,r-1,r:n}(\cdot|\tau)$ $(r \geq 3)$ to represent the unconditional and conditional joint probability density functions (PDF) of the three order statistics, respectively, and $f^X_{r:s}(\cdot)$ and $f^{X|\mathtt{T}}_{r:s}(\cdot|\tau)$ represent the unconditional and conditional PDF of the $r$th order statistic of measurements $X$ out of a sample of size $s$ $(r \leq s)$.

The identification exploits the fact that the conditional joint distribution of three consecutive order statistics has a multiplicative separable structure. Specifically, the unconditional joint distribution, which can be estimated from the data, can be expressed as

$$
\begin{aligned}
f_{r-2,r-1,r:n}(x,y,z) &= \int_{\mathcal{T}} f_{r-2,r-1,r:n}(x,y,z|\tau) f^{\mathtt{T}}(\tau) d\tau \\
= c_{r,n} \underbrace{\mathbb{1}(x \le y \le z)}_{\text{correlation}} &\int_{\mathcal{T}} \underbrace{f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) f^{X|\mathtt{T}}(y|\tau) f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)}_{\text{multiplicatively separable}} f^{\mathtt{T}}(\tau) d\tau, \qquad (2)
\end{aligned}
$$

where $c_{r,n} = \frac{n!}{(r-2)!\cdot(n-r+1)!}$, and $\mathbb{1}(\cdot)$ is the indicator function. The first equality holds by the law of total probability, and the second extends [40]'s Lemma 1 to three consecutive order statistics.[10] This joint distribution of the consecutive order statistics has a semi-separable structure in the sense that we can separate the observed joint density function into the integration of three density functions, which is similar to (1) in the measurement error literature, but it has an extra restriction by the nature of order statistics, i.e., $\mathbb{1}(x \le y \le z)$, which cannot be separated. This semi-separable structure precludes us from readily borrowing the same identification procedure in the existing literature to identify the conditional latent distributions directly.

Fortunately, the restriction by the indicator function can be safely circumvented if we divide the original support by two cutoff points $c_1$ and $c_2$, where $c_1 < c_2$, to separate the support into three parts, referred to as “low”, “middle”, and “high”. We denote these segments as $\mathcal{X}_l \equiv \{x : x \le c_1\}, \mathcal{X}_m \equiv [c_1, c_2]$, and $\mathcal{X}_h \equiv \{x : x \ge c_2\}$, respectively. Our context of three order statistics calls for a three-part discretization, which extends [40]'s two-part discretization using two order statistics and an IV. That is, the separable structure of the joint distribution $f_{r-2,r-1,r:n}(x,y,z)$ reappears if we always restrict $x \in \mathcal{X}_l$, $y \in \mathcal{X}_m$, and $z \in \mathcal{X}_h$. Specifically, if $x \in \mathcal{X}_l$, $y \in \mathcal{X}_m$, and $z \in \mathcal{X}_h$, the joint distribution can be expressed as

$$
f_{r-2,r-1,r:n}(x,y,z) = c_{r,n} \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) f^{X|\mathtt{T}}(y|\tau) f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau) f^{\mathtt{T}}(\tau) d\tau, \quad (3)
$$

which has the same structure as the measurement error models but a different conceptual interpretation for each component. Figure 1 provides a visualization of the discretization.

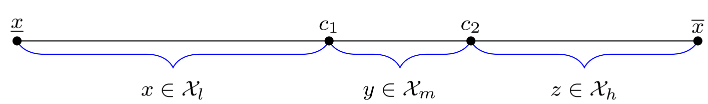



FIG 1. *Discretization.*

Following the identification strategy developed in [29], we introduce the following integral operator that associates a function of two variables.

[10]The joint distribution of any three order statistics does not have such a multiplicatively separable structure, i.e., $f_{r,s,t:n}(x,y,z) \sim f(x)f(y)f(z)[F(x)]^{r-1}[F(y)-F(x)]^{s-r-1}[F(z)-F(y)]^{t-s-1}[1-F(z)]^{n-t}$, where $r < s < t$, see [15]. We derive Equation (2) in Appendix A.1.

**Definition 2.1.** Let $L_{x|\tau}$ denote an operator that maps function $g$, where $g \in \mathcal{G}(\mathcal{T})$, to $L_{x|\tau}g \in \mathcal{G}(\mathcal{X}_l)$; and $H_{x|\tau}$ maps function $g$, where $g \in \mathcal{G}(\mathcal{X}_h)$, to $H_{x|\tau}g \in \mathcal{G}(\mathcal{T})$. Specifically, the two operators are defined as

$$[L_{x|\tau}g](x) \equiv \int_{\mathcal{T}} f^{X|\mathtt{T}}(x|\tau)g(\tau)d\tau \quad \text{and} \quad [H_{x|\tau}g](\tau) \equiv \int_{\mathcal{X}_h} f^{X|\mathtt{T}}(x|\tau)g(x)dx.$$

Note that both operators involve one segment of bid support $\mathcal{X}$. We further introduce another linear operator based on the joint distribution and the diagonal operator defined as follows. In particular, for a given $y \in \mathcal{X}_m$, let $J_y$ denote an operator mapping $g \in \mathcal{G}(\mathcal{X}_h)$ to $J_y g \in \mathcal{G}(\mathcal{X}_l)$:

$$[J_y g](x) \equiv \int_{\mathcal{X}_h} f_{r-2,r-1,r:n}(x,y,z)g(z)dz.$$

Given a particular partition $\{\mathcal{X}_l, \mathcal{X}_m, \mathcal{X}_h\}$, $J_y$ is defined for every given $y$ in $\mathcal{X}_m$. Let $\Delta_{X=y,\mathtt{T}}$ denote the diagonal operator mapping $g \in \mathcal{G}(\mathcal{T})$ to $\Delta_{X=y,\mathtt{T}}g \in \mathcal{G}(\mathcal{T})$:

$$[\Delta_{X=y,\mathtt{T}}g](\tau) \equiv c_{r,n} f^{X|\mathtt{T}}(y|\tau) f^{\mathtt{T}}(\tau) g(\tau).$$

We derive the equivalence of operators in Appendix A.2 as follows:

$$J_y = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}, \tag{4}$$

based on Equation (3) and by exploiting the following features: (i) an interchange of the order of integrations (justified by Fubini's theorem), (ii) the definition of $H_{X_{1:n-r+1}|\mathtt{T}}$, (iii) the definition of $\Delta_{X=y,\mathtt{T}}$ operating on $H_{X_{1:n-r+1}|\mathtt{T}}g$, and (iv) the definition of $L_{X_{r-2:r-2}|\mathtt{T}}$ operating on $[\Delta_{X=y,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}g]$. Note that such equivalence between the operators holds for any value of $y \in \mathcal{X}_m$.

For identification, we impose the following injective assumption.

**Assumption 3.** *(Injective) There exists one division of the domain such that the operators $L_{\mathtt{T}|X_{r-2:r-2}}$ and $H_{X_{1:n-r+1}|\mathtt{T}}$ are injective for $\mathcal{G} = \mathcal{L}^2$, where $\mathcal{L}^2(\mathcal{X})$ denotes the set of all square-integrable functions with domain $\mathcal{T}$ and $\mathcal{X}_h$, respectively.*

It is worth noting that our identification is agnostic about the specific division of the domain that Assumption 3 is satisfied, i.e., specific values of $c_1$ and $c_2$. The identification holds as long as there exists one pair of those constants and the associated rank conditional holds. One does not need to determine those constants for identification. Moreover, there might exist multiple pairs of $c_1$ and $c_2$ that the identification results apply, which could be used to conduct an over-identification test.

An operator $A$ is injective if $Af = Ag$ implies $f = g$ for any $f, g$ in the domain of $A$. A linear operator being injective is equivalent to the family of kernel functions used to define the operator being complete; see [29]. In our context, if the family of distributions $\{f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) : x \in \mathcal{X}_l\}$ is complete over $\mathcal{L}^2(\mathcal{T})$, that is, the unique solution $\tilde{g}$ to the equation $\int_{\mathcal{T}} g(\tau) f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)d\tau = 0$ for all $x \in \mathcal{X}_l$ is $\tilde{g}(\cdot) = 0$, then $L_{\mathtt{T}|X_{r-2:r-2}}$ is injective under Assumption 2. We further provide conditions on the parental distributions under which the family of the order statistics' distributions is complete in Appendix A.3. However, the equivalence between the injectiveness of operator $H_{X_{1:n-r+1}|\mathtt{T}}$ and the completeness of the kernel function family $\{f^{X|\mathtt{T}}_{1:n-r+1}(x|\tau) : \tau \in \mathcal{T}\}$ over $\mathcal{L}^2(\mathcal{X}_h)$ is not straightforward, because the operator is defined only in a segment of the support. We prove that as long as the original distribution family is complete, i.e., $\{f^{X|\mathtt{T}}_{1:n-r+1}(x|\tau) : \tau \in \mathcal{T}\}$ over $\mathcal{L}^2(\mathcal{X})$ is complete, there exists at least one division of the support such that operator $H_{X_{1:n-r+1}|\mathtt{T}}$ is injective. See Appendix A.3.

Completeness of the relevant family of distributions provides one way to characterize the injectivity of an operator. Intuitively, the family of distributions $\{f^{X|\mathtt{T}}(x|\tau) : x \in \mathcal{X}\}$ being complete implies there is *sufficient variation* in the conditional density of $X$ across different values of $\mathtt{T}$. An example of such a complete distribution is a normal distribution with mean $\tau$ and variance 1. On the other hand, if the conditional density of $X$ does not vary sufficiently across $\tau$, such as the standard normal distribution, the distribution family is not complete. Obviously, in such a scenario, $X$ is independent of $\mathtt{T}$, and hence we can easily find $g \neq 0$ such that $\int g(\tau) f^{X|\mathtt{T}}(x|\tau) d\tau = 0$ for any $x$.

Assumption 3 also specifies that we consider the identification with $\mathcal{G} = \mathcal{L}^2$. Such consideration is due to the following two reasons. First, this space is sufficiently large such that the density can be sampled everywhere, which ensures a one-to-one mapping between a density function and its corresponding operator. Thus, the density function can be uniquely determined by the associated operator with such a choice of $\mathcal{G}$.[11] Second, it is a Hilbert space if equipped with the norm $\|g\|_{\mathcal{L}^2} = \left(\int_{\mathcal{X}} g^2(x) dx\right)^{1/2}$ for any $g \in \mathcal{G}(\mathcal{X})$. One advantage of considering Hilbert spaces is that it is easier to use properties of the operators such as $L_{X_{r-2:r-2}|\mathtt{T}}$ and $H_{X_{1:n-r+1}|\mathtt{T}}$ later, because there are many existing theoretical results developed for operators defined in Hilbert spaces. For instance, it is straightforward to define the adjoint operator by using the concept of inner product in Hilbert spaces. It is also worth noting that this space differs from the $\mathcal{L}^1$ space adopted in [29], which is a Banach space.

If an operator is injective, its inverse is well-defined but may be defined over a restricted domain. We further prove that $L_{X_{r-2:r-2}|\mathtt{T}}$ is surjective in addition to being injective so that the domain of its inverse is the whole space $\mathcal{L}^2(\mathcal{X})$. This is important for proving the equivalence of operators defined in the data and in the distributions to be identified. We summarize this result in the following lemma and relegate the proof to Appendix A.4.

**Lemma 2.2.** *If Assumptions 1-3 hold, then $L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ exists and is densely defined over $\mathcal{L}^2(\mathcal{X}_l)$.*

Lemma 2.2 essentially indicates that operator $L_{X_{r-2:r-2}|\mathtt{T}}$ is surjective if it is injective. We use the following simple example to facilitate understanding the

[11] Because $f^X_{r-2:r-2}(x|\tau_0) = \lim_{n\to\infty} [L_{X_{r-2:r-2}|\mathtt{T}} g_{n,\tau_0}](x)$, where $g_{n,\tau_0}(\tau) = n\mathbb{1}(|\tau - \tau_0| \leq n^{-1})$, a sequence of bounded and square-integrable functions,the space $\mathcal{G} = \mathcal{L}^2$ is sufficiently rich.

necessity of the surjective property and the difference between linear operators and matrices. Suppose that $\mathcal{D}^1$ and $\mathcal{D}^2$ are two linear spaces, and $L$ is a linear transformation from $\mathcal{D}^1$ to $\mathcal{D}^2$. If both $\mathcal{D}^1$ and $\mathcal{D}^2$ are finite-dimensional, $L$ is injective if and only if it is surjective. In particular, if $\dim(\mathcal{D}^1) = \dim(\mathcal{D}^2)$ and $L$ is associated with a square matrix $A$, then $L$ is both injective and surjective if and only if $A$ has full rank. But this relationship does not trivially hold in infinite-dimensional cases. For example, let $\{e_i\}_{i=1}^{\infty}$ be the basis of $\mathcal{D}^1$ as well as $\mathcal{D}^2$. We assume that $Le_i = e_{i+1}$ for every $i \geq 1$. Such an operator $L$ is obviously injective but not surjective, because the base $e_1$ is missing in its range.

Since $H_{X_{1:n-r+1}|\mathtt{T}}$ is injective under Assumption 3, we can eliminate the common operator $H_{X_{1:n-r+1}|\mathtt{T}}$ by equivalence of operators specified in Equation (4) for any two different values of $y$, i.e., $y_1$ and $y_2$, leading to the following main equation for identification:

$$J_{y_1} J_{y_2}^{-1} = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y_1,\mathtt{T}} \Delta_{X=y_2,\mathtt{T}}^{-1} L_{X_{r-2:r-2}|\mathtt{T}}^{-1}. \tag{5}$$

By Lemma 2.2, the relation (5) is established over a dense subset of $\mathcal{L}^2(\mathcal{X}_l)$. In fact, it can be further extended to the full space $\mathcal{L}^2(\mathcal{X}_l)$ by leveraging the extension procedure of linear operators. This equation ensures that operator $J_{y_1} J_{y_2}^{-1}$ can be represented as an eigenvalue-eigenfunction decomposition with the two unknown operators $L_{X_{r-2:r-2}|\mathtt{T}}$ and $\Delta_{X=y_1,\mathtt{T}} \Delta_{X=y_2,\mathtt{T}}^{-1}$ being the eigenfunctions and eigenvalues, respectively. Consequently, diagonalizing operator $J_{y_1} J_{y_2}^{-1}$, which can be computed from the data directly since it is defined using observable densities, provides the eigenfunctions $L_{X_{r-2:r-2}|\mathtt{T}}$, indexed by the latent UH, and further provides the unobserved densities of order statistic $X_{r-2:r-2}|\mathtt{T}$.

Note that there are three features prevalent in identification using decomposition: The identification may not be unique; the identification is up to scales; the identification is up to ordering and location. We tackle the three issues one at a time below.

### *Unique decomposition*

To guarantee unique decomposition, we impose restrictions on the relationship between observed measurement $X$ and UH $\mathtt{T}$ in segment $\mathcal{X}_m$.

**Assumption 4.** *(Distinct) there exists one division of the domain such that, for all $\tau_1, \tau_2 \in \mathcal{T}$, the set $\{(y_1, y_2) : \frac{f^{X|\mathtt{T}}(y_1|\tau_1)}{f^{X|\mathtt{T}}(y_2|\tau_1)} \neq \frac{f^{X|\mathtt{T}}(y_1|\tau_2)}{f^{X|\mathtt{T}}(y_2|\tau_2)}$, where $(y_1, y_2) \in \mathcal{X}_m \times \mathcal{X}_m\}$ has positive probability whenever $\tau_1 \neq \tau_2$.*

This assumption is weaker than assuming that the associated operator is injective in segment $\mathcal{X}_m$. Note that we just need one division where such an assumption holds. This assumption fails only if the distribution of the measurement conditional on the latent factor is the same at some pair of two distinct values $\tau_1$ and $\tau_2$. Assumption 4 guarantees unique eigenvalues, so that conducting the decomposition to operator $J_{y_1} J_{y_2}^{-1}$ identifies operator $L_{X_{r-2:r-2}|\mathtt{T}}$, and thus identifies the conditional density $f_{r-2:r-2}^{X|\mathtt{T}}(x|\tau)$, for $x \in \mathcal{X}_l$.

Our identification is agnostic about the specific values of $y_1$ and $y_2$, and only requires to have one pair of $y_1$ and $y_2$ that Assumption 4 holds. If our model is correctly specified, different pairs of $y_1$ and $y_2$ should lead to the same identified components. Therefore, we could in theory use different pairs of $y_1$ and $y_2$ to conduct an over-identification test.

However, even if the decomposition is unique, such identification is up to scales and locations. That is, the conditional density $f^{X|\mathrm{T}}_{r-2:r-2}(x|\tau)$ is identified as the true density multiplied by an unknown constant, which could differ for each UH. The existing literature relies on the property that the total probability is equal to 1 for each conditional distribution to pin down the scales. Such an approach is not feasible in our framework because, from the decomposition, we only identify the conditional distribution in one segment of the full support, i.e., $\mathcal{X}_l$. Mover, one can neither pin down the ordering or the actual values of UH, which calls for extra restrictions.

To proceed, we propose to leave the ordering of the UH and the scales in the low segment undetermined and proceed to identify the conditional distributions in the other two segments first. In this procedure, we mainly use Equation (5). One main feature worth noting during this process is that we keep the value of the UH consistently matched across the three segments. Furthermore, the undetermined scales are the same for the same UH in the same segment but may vary across UH or segments. Given these, we can then pin down the scales and order in what follows.

### *Unique scale*

Note that we can identify the conditional distributions in all three segments up to different scales. That is, each segment of the conditional distribution is associated with one scale parameter, so together there are three scale parameters to pin down for each conditional distribution. These scales can then be pinned down by invoking the continuity of the component PDFs and the total probability argument. First, the PDFs identified separately in the three segments should be the same at the cutoff points due to the continuity of the true conditional distributions. Second, the fact that each conditional distribution should integrate to 1 provides the third restriction on the scales. These restrictions uniquely pin down the scales.

### *Unique ordering and location*

Given that the conditional distributions are identified in the full support, we provide a condition using the auction setting to pin down the exact location of the UH. Specifically, letting UH be the unobserved quality of the auctioned item, we would expect that bidders' values/bids are, on average, higher and of better quality. For instance, in second-hand automobile auctions, omitted details from the car description, such as dents and scratches, are revealed upon pre-auction inspection and enter bidder values.

**Assumption 5.** *(Monotonicity and Location) The expected value/bid is strictly monotone with UH; that is, $E(X|\mathtt{T} = \tau)$ is strictly monotone with $\tau$ for all $\tau \in \mathcal{T}$. Moreover, we assume that the support of UH is [0, 1].*

The monotonicity assumption is useful to pin down UH's relative ordering. However, its exact location/value is still unidentified. That is, one could always apply an affine transformation to the UH and obtain an observationally equivalent model that satisfies all assumptions. To pin down UH's exact location, we normalize its support to be $[0, 1]$, which is without loss of generality. Such a normalization is similar to the mean zero normalization.

**Theorem 2.3.** *If Assumptions 1-5 are satisfied, conditional bid distribution $f^{X|\mathtt{T}}(x|\tau)$ for $x \in \mathcal{X}$ and $\tau \in \mathcal{T}$ and UH's distribution $f^{\mathtt{T}}(\tau)$ for any $\tau \in \mathcal{T}$ are identified using any three consecutive order statistics of bids.*

We summarize the main steps of the proofs below and leave the details to Appendix A.6.[12] The identification proceeds sequentially. First, we identify operator $L_{X_{r-2:r-2}|\mathtt{T}}$ from the decomposition of Equation (5). Such identification is unique by Assumption 4, but up to scales and location. Second, we identify the operator $H_{X_{1:n-r+1}|\mathtt{T}}$ up to different scales, similar to the identification of $L_{X_{r-2:r-2}|\mathtt{T}}$. Third, for any value $y \in \mathcal{X}_m$, we can identify operator $\Delta_{X=y,\mathtt{T}}$ up to the same scales for all $y$ once we plug the identified operators $L_{X_{r-2:r-2}|\mathtt{T}}$ and $H_{X_{1:n-r+1}|\mathtt{T}}$ into Equation (4). Using the one-to-one mapping between operators and the associated densities, we then identify the unobserved densities $f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)$ for $x \in \mathcal{X}_l$, $f^{X|\mathtt{T}}(y|\tau)f^{\mathtt{T}}(\tau)$ for $y \in \mathcal{X}_m$, and $f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)$ for $z \in \mathcal{X}_h$ up to scales. The scales are the same in the same segment but may vary across different segments. Furthermore, we show that the one-to-one mapping between the distribution of an order statistic and its parent distribution can be extended from the full support to a segment. Thus, we identify the conditional distribution up to different scales in all three segments. Lastly, the scales are then pinned down using three restrictions.

Once the conditional bid distributions are identified as in Theorem 2.3, we can exploit the one-to-one mapping between the conditional value and bid distributions to recover the conditional value distributions, which are the target of interest. Specifically, for ascending auctions, where bidders' weakly dominant strategy is to bid their values, the conditional value distribution is the same as the conditional bid distribution;[13] for first-price auctions, we can identify the conditional value distribution by exploiting the one-to-one mapping established in [21]. We summarize this result in the following Corollary.

**Corollary 2.4.** *If Assumptions 1-5 are satisfied, the conditional value distribution $f^{v|\mathtt{T}}(v|\tau)$ for $v \in \mathcal{V}$ and $\tau \in \mathcal{T}$ and the latent variable's distribution $f^{\mathtt{T}}(\tau)$ for $\tau \in \mathcal{T}$ are identified using any three consecutive order statistics of bids.*

[12] We thank Yingyao Hu and Ji-Liang Shiu for valuable insights about proving the theorem.

[13] Many empirical studies adopt the same assumption in ascending auctions; see, e.g., [37], [4], and [26]. We exclude other possible bidding strategies such as jump bidding allowed in [24]. Such abstraction is a good approximation for online auctions and button auctions. For instance, eBay allows bidders to set up a proxy bid.

The identification results in Theorem 2.3 are achieved under the assumption that the reserve price is not binding. However, in practice, the reserve price appears to be binding in many cases, leading to a truncation in the observed bid distribution. We show in the following corollary that we can still identify the bid/value distribution with a truncation. We can also identify the conditional probability of the truncation when the number of potential bidders is observed.

**Remark 1.** *Our results are useful for many auction-based applications such as online advertising auctions and other settings including beauty contests, war of attrition models, where many players compete for multiple prizes, wage offers, where only the top offers are recorded, and repeated experiments such as in reliability testing, where consecutive low-order failure times are recorded.*

### *2.2. Reserve price for ascending auctions*

If the reserve price is binding, the optimal bidding strategy for any bidder is to submit the optimal bid computed without reserve prices when such an optimal bid is above the reserve price, and to not bid otherwise. Therefore, the presence of a binding reserve price $R$ creates a truncation in the observed bid distribution, i.e., $\tilde{F}^{X|\mathtt{T}}(x|\tau) \equiv \frac{F^{X|\mathtt{T}}(x|\tau) - F^{X|\mathtt{T}}(R|\tau)}{1 - F^{X|\mathtt{T}}(R|\tau)}$, where $x \in [R, \overline{x}]$. Let $n$ denote the number of actual bidders and $N$ denote the number of potential bidders. In first-price auctions, even if entry is exogenous, the observed bid distribution depends on both $N$ and $n$, while in ascending auctions, it only depends on $n$. Therefore, to illustrate the intuition, we focus on ascending auctions.

Under such a situation, even with a truncation caused by a binding reserve price, we can still follow the identification strategy in Theorem 2.3 to identify the truncated CDF $\tilde{F}^{X|\mathtt{T}}(x|\tau)$, PDF $\tilde{f}^{X|\mathtt{T}}(x|\tau)$, and the marginal distribution of the UH without information on $N$ as long as $n$ is known. Specifically, the joint distribution of three consecutive active bids with a bidding reserve price can be expressed as

$$\begin{aligned}
&\tilde{f}_{r-2,r-1,r:n}(x,y,z) \\
&= c_{r,n} \cdot \mathbb{1}(x \le y \le z) \cdot \int_{\mathcal{T}} \tilde{f}^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) \tilde{f}^{X|\mathtt{T}}(y|\tau) \tilde{f}^{X|\mathtt{T}}_{1:n-r+1}(z|\tau) f^{\mathtt{T}}(\tau) d\tau.
\end{aligned}$$

A few features are worth noticing. First, identification using eigen-decomposition applies regardless of whether $N$ is observed, as the bidding strategy does not vary with $N$ under exogenous entry. Second, without observing bids below the reserve price, there is no information to identify the bid/value distribution for this segment. Lastly, we establish that we can identify the conditional probability of the truncation $F^{X|\mathtt{T}}(R|\tau)$.

**Corollary 2.5.** *In ascending auctions, when $N$ is observed and has large support, the conditional probability of truncation $F^{X|\mathtt{T}}(R|\tau)$ is identified using the distribution of the number of actual bidders conditional on the potential bidders.*

*Therefore, for all $x \geq R$, $F^{X|\mathrm{T}}(x|\tau)$ is identified from*

$$\tilde{F}^{X|\mathrm{T}}(x|\tau) \equiv \frac{F^{X|\mathrm{T}}(x|\tau) - F^{X|\mathrm{T}}(R|\tau)}{1 - F^{X|\mathrm{T}}(R|\tau)}.$$

Intuitively, the distribution of $n$ conditional on $N$ is a mixture of binomial distributions with the success probability being the conditional truncated probability. That is,

$$\Pr(n|N) = \int_{\tau \in \mathcal{T}} C_{N,n}[1 - F^{X|\mathrm{T}}(R|\tau)]^n [F^{X|\mathrm{T}}(R|\tau)]^{N-n} dF^{\mathrm{T}}(\tau), \tag{6}$$

where $\Pr(n|N)$ is estimable from the data, $C_{N,n}$ is a constant, $F^{\mathrm{T}}(\tau)$ can be treated as known, and conditional truncation probability $F^{X|\mathrm{T}}(R|\tau)$ is the object of interest. This is similar in structure but differs conceptually from the identification in the mixture literature [22], where the goal is to identify the mixture distribution with the success probability taking any value in $[0, 1]$. We show that our identification problem can be viewed as a dual problem by changing variables in the integral. The detailed proof for Corollary 2.5 can be found in Appendix A.7.

### *2.3. The number of order statistics*

Our discussion so far assumes that three consecutive order statistics of bids are available. There are various ways to extend this main identification result. First, the required number of consecutive order statistics reduces to two if there exists an instrument that is independent of the bids conditional on UH; see [40].[14] Second, while consecutiveness barely restricts the data with incomplete bids, exploiting the Markov property of order statistics relaxes this requirement. In Appendix C, we show that any four order statistics identify the model.[15]

## 3. Sieve maximum likelihood estimation

### *3.1. Consistency of sieve estimation*

Note that conducting counterfactual policy analysis requires one to estimate the joint distribution of UH and bidder private values. In principle, the conditional bid distribution and UH's marginal distribution could be estimated fully nonparametrically by following the constructive identification argument step-by-step. Specifically, one could do a partition in the full support and conduct eigenfunction decomposition to estimate the distribution of the order statistics

[14] Measurement error approaches are inapplicable when only one order statistic, such as the winning bid, is observed. This calls for alternative strategies, such as density discontinuity approaches first proposed by [20].

[15] The idea of using Markov property for dealing with UH and incomplete bid data simultaneously is first explored in [41], who uses five order statistics in finite UH framework.

in the three segments, then use the one-to-one mapping between the distribution of an order statistic and its parent distribution to estimate the parent distribution. Such a fully nonparametric estimator not only poses a high demand on the data but is also of low efficiency, as it depends critically on the partition of the support and involves sequential estimation.

Considering the fact that, in applications, the analyst oftentimes can only access modest-sized data, we propose to estimate these two densities using the method of sieves [19, 46, 9, 7] to fully exploit variations in the data instead of relying on a particular partition. We establish consistency and convergence rates for such estimators.

Our strategy is to first provide some regularity assumptions on the sieve approximation for consistency, which usually depends on the smoothness of the function to be approximated and the complexity of the sieve space. Such complexity is characterized by its upper bound and bracketing numbers.[16]

We represent the log-likelihood function of the joint distribution of the three consecutive order statistics, i.e., data $\equiv \{X_{r-2:n} = x^i, X_{r-1:n} = y^i, X_{r:n} = z^i\}_{i=1}^m$, as follows:

$$\begin{aligned}\log L(\text{data}; f^{X|\mathtt{T}}, \quad f^{\mathtt{T}}) = \tfrac{1}{m}\tfrac{n!}{(r-3)!(n-r)!}\textstyle\sum_{i=1}^m \log \int_\tau [F^{X|\mathtt{T}}(x^i|\tau)]^{r-3} f^{X|\mathtt{T}}(x^i|\tau)\\ f^{X|\mathtt{T}}(y^i|\tau)[1-F^{X|\mathtt{T}}(z^i|\tau)]^{n-r} f^{X|\mathtt{T}}(z^i|\tau) f^{\mathtt{T}}(\tau)d\tau.\end{aligned}$$

As both the conditional density and the marginal density can be derived from a joint density, we propose to approximate joint distribution $f^{X,\mathtt{T}}(x,\tau)$ by using tensor product bases of univariate series. Specifically, let $\mathcal{B}_m$ be the finite-dimensional sieve space and $\xi_1, \ldots, \xi_{p_m}$ be its basis, where $p_m$ is the number of basis functions in the sieve space.

With slight abuse of notation, we denote the sieve representation of this joint distribution as $\mathfrak{f}$. We then represent the marginal distribution, the conditional distribution, and the CDF of such a conditional distribution as follows:

$$\begin{aligned} f^{\mathtt{T}}(\tau) &= \int_{\mathcal{X}} f^{X,\mathtt{T}}(x,\tau)dx \simeq \int_{\mathcal{X}} \mathfrak{f}(x,\tau)dx, && (7)\\ f^{X|\mathtt{T}}(x|\tau) &= \frac{f^{X,\mathtt{T}}(x,\tau)}{f^{\mathtt{T}}(\tau)} \simeq \frac{\mathfrak{f}(x,\tau)}{\int_{\mathcal{X}} \mathfrak{f}(x,\tau)dx}, && (8)\\ F^{X|\mathtt{T}}(x|\tau) &= \int_{-\infty}^{x} f^{X|\mathtt{T}}(t|\tau)dt \simeq \frac{\int_{-\infty}^{x} \mathfrak{f}(t,\tau)dt}{\int_{\mathcal{X}} \mathfrak{f}(x,\tau)dx}. \end{aligned}$$

Consequently, the sieve estimator for the joint distribution of the three observed consecutive bids can be represented as

$$\hat{\mathfrak{f}} = \operatorname*{arg\,max}_{\mathfrak{f}\in\mathcal{B}_m} \log L\left(\text{data}; \frac{\mathfrak{f}(x,\tau)}{\int_{\mathcal{X}} \mathfrak{f}(x,\tau)dx}, \int_{\mathcal{X}} \mathfrak{f}(x,\tau)dx\right). \tag{9}$$

Next, we show that under some regularity conditions the proposed sieve estimator for the joint distribution in Equation (9) is consistent. Once the joint distribution is consistently estimated, the conditional and marginal distributions,

[16] In contrast, [29] uses a covering number to characterize complexity.

specified in Equations (7) and (8) respectively, are also consistently estimated. Let $f_0^{X,\mathtt{T}}(x,\tau)$ denote the true joint density, and let $f_0^{X|\mathtt{T}}(x|\tau)$ and $f_0^{\mathtt{T}}(\tau)$ denote the true conditional density of $X$ given $\mathtt{T}=\tau$ and the marginal density of the latent variable, respectively. We introduce some regularity conditions.

**Assumption 6.** *(Compactness) $X$ has a compact support. Without loss of generality, we assume that its support is [0, 1].*

This compact support assumption is standard in the auction literature. Moreover, we can linearly transform random variables with compact support to ones that have support on $[0,1]$. Note that such a transformation has to be linear, rather than an arbitrary monotone transformation. The linear transformation is for the convenience of using the observed data in estimation. The support of the two random variables, $X$ and $\mathtt{T}$, plays an important role in choosing an appropriate sieve space to perform maximum likelihood estimation. For example, the trigonometric sieve is inapplicable when the support is $\mathbb{R}$. In this case, Hermite polynomials and B-splines are preferable. A B-spline approximation is also useful when the support is compact. It is worth emphasizing that our identification results hold regardless of this normalization.

**Assumption 7.** *(Sieve approximation) There exists $f_m^{X,\mathtt{T}}(x,\tau)$, which is represented in terms of the bases $\xi_1,\ldots,\xi_{p_m}$ in the sieve space, for some $\beta>0$, such that*

$$\|f_m^{X,\mathtt{T}}(x,\tau)-f_0^{X,\mathtt{T}}(x,\tau)\|_{L_\infty([0,1]^2)}=O(p_m^{-\beta}).$$

Assumption 7 ensures that the joint density can be approximated sufficiently well in the sieve space. It is worth noting that the sieve space constructed by either B-spline or Bernstein basis functions, which are popular sieve spaces in auctions, satisfies Assumption 7. With this assumption satisfied, by Equations (7) and (8), both the conditional density and the marginal density can be approximately sufficiently well by functions in the sieve space. That is, there exist $f_m^{X|\mathtt{T}}(x|\tau)$ and $f_m^{\mathtt{T}}(\tau)$, both represented in terms of $\xi_1,\ldots,\xi_{p_m}$ in the sieve space, such that

$$\begin{aligned}\|f_m^{X|\mathtt{T}}(x|\tau)-f_0^{X|\mathtt{T}}(x|\tau)\|_{L_\infty([0,1]^2)}&=O(p_m^{-\beta}),\ \text{and}\\ \|f_m^{\mathtt{T}}(\tau)-f_0^{\mathtt{T}}(\tau)\|_{L_\infty([0,1]^2)}&=O(p_m^{-\beta}).\end{aligned}$$

To study the asymptomatic properties of the proposed estimator, we first establish the relationship among the sieve estimator, the sieve representation, and the underlying true densities. Let $G(x,y,z;f^{X|\mathtt{T}},f^{\mathtt{T}})$ be the log-likelihood function from one single observation that depends on the conditional density of $X$ given $\mathtt{T}=\tau$ and the marginal density of $\mathtt{T}$.

**Lemma 3.1.** *Let $\hat f_m^{X|\mathtt{T}}(x|\tau)$ and $\hat f_m^{\mathtt{T}}(\tau)$ denote the estimated conditional density of $X$ and the marginal density of the latent variable $\mathtt{T}$, respectively. We have*

$$\frac{1}{\sqrt{m}}\boldsymbol{G}_m\left[\log\frac{G(x,y,z;\hat{f}_m^{X|\mathtt{T}},\hat{f}_m^{\mathtt{T}})}{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})}\right]\geq\boldsymbol{P}\left[\log\frac{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})}{G(x,y,z;f_0^{X|\mathtt{T}},f_0^{\mathtt{T}})}\right]$$
$$+\boldsymbol{P}\left[\log\frac{G(x,y,z;f_0^{X|\mathtt{T}},f_0^{\mathtt{T}})}{G(x,y,z;\hat{f}_m^{X|\mathtt{T}},\hat{f}_m^{\mathtt{T}})}\right], \qquad (10)$$

*where* $\boldsymbol{G}_m = \sqrt{m}(\boldsymbol{P}_m - \boldsymbol{P})$*,* $\boldsymbol{P}_m$ *denotes the empirical measure of data* $(x_i, y_i, z_i)_{i=1}^m$*, and* $\boldsymbol{P}$ *denotes the true distribution.*

Lemma 3.1 holds by definition of $f_0$ and $\hat{f}$. The detailed proof can be found in Appendix B.1. This inequality follows from that our sieve estimator maximizes the empirical measure of the log-likelihood function over the sieve space. The left-hand side measures the difference between the empirical and true measures for the log-likelihood ratio. The first term on the right-hand side measures how close the sieve approximation is to the true functions; the second term measures how close our sieve estimator is to the true functions. To find an upper bound on the second term in order to establish consistency and the convergence rate of sieve estimators, we first bound the left-hand side from above using empirical process techniques and then bound from below the first term using the Lipschitz continuous property of $G$.

To accomplish (11), we resort to empirical process theories and impose restrictions on the complexity of the sieve space. We first introduce the following two assumptions to characterize its complexity.

**Assumption 8** (Bound of sieve space)**.** *The logarithm of the upper bound over* $\mathcal{B}_m$*, denoted by* $Q_m$*, satisfies* $\log\{\sup_{\mathfrak{f}\in\mathcal{B}_m}\|\mathfrak{f}\|_{L_\infty([0,1]^2)}\}\leq Q_m = O(\log\log m)$*.*

**Assumption 9** (Bracketing number)**.** *The* $\epsilon$ *bracketing number of the sieve space* $\mathcal{B}_m$ *is of order* $O\left((e^{2Q_m}/\epsilon)^{p_m+2}\right)$ *for some constant* $p_m = O(m^\alpha)$ *with* $0<\alpha<1/2$*.*

Intuitively, $Q_m$ would be larger for a larger space. We define the bracketing number following [48]. Specifically, given two functions $l$ and $u$, the bracket $[l,u]$ is the set of all functions $f$ with $l\leq f\leq u$. An $\epsilon$-bracket is a bracket $[l,u]$ with $\|u-l\|\leq\epsilon$ under a certain norm $\|\cdot\|$. The $\epsilon$ bracketing number $N_{[]}(\epsilon,\mathcal{B},\|\cdot\|)$ is the minimum number of $\epsilon$-brackets needed to cover $\mathcal{B}$. A larger $\epsilon$ bracketing number corresponds to a more complex sieve space.

To guarantee consistency, we consider the function class $\mathcal{F}_m$, defined by

$$\left\{\log\frac{G(x,y,z;\tilde{f}_m^{X|\mathtt{T}},\tilde{f}_m^{\mathtt{T}})}{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})}:\tilde{f}_m^{X|\mathtt{T}}=\frac{\mathfrak{f}_m(x,\tau)}{\int_{\mathcal{X}}\mathfrak{f}_m(x,\tau)dx},\tilde{f}_m^{\mathtt{T}}=\int_{\mathcal{X}}\mathfrak{f}_m(x,\tau)dx\right\},$$

where $\mathfrak{f}_m$ is represented in terms of $\xi_1,\dots,\xi_{p_m}$ in sieve space $\mathcal{B}_m$. If the complexity of sieve space $\mathcal{B}_m$ satisfies Assumptions 8-9, we are able to quantify the upper bound on $\mathcal{F}_m$, which is the upper bound on the left-hand side of (10).

We now establish consistency of the proposed sieve estimator.

**Theorem 3.2.** *Under Assumptions 6-9, the proposed sieve MLE for the joint distribution is consistent. Moreover, both the conditional and marginal distributions are consistently estimated. That is,*

$$\|\hat{f}_m^{X|\mathtt{T}}(x|\tau) - f_0^{X|\mathtt{T}}(x|\tau)\|_{L_2} \xrightarrow{p} 0, \text{ and } \|\hat{f}_m^{\mathtt{T}}(\tau) - f_0^{\mathtt{T}}(\tau)\|_{L_2} \xrightarrow{p} 0.$$

*The convergence rate for these estimators is derived to be $B(m, p_m, Q_m)^{1/2}$, where $B(m, p_m, Q_m) = e^{c_2 Q_m} p_m \log p_m/\sqrt{m} + e^{c_2 Q_m}/p_m^{\beta}$, with $c_2$ being a constant.*

The detailed proof is given in Appendix B.1. Note that we consider $L_2$ convergence of our proposed estimator. Establishing the (uniform) convergence rate is beyond the scope of this paper and thus left for future research. As pointed out in [42], the uniform rate depends on $r$, and the nonparametric MLE of the parent distribution obtained using order statistics may have a slower convergence rate near the tail of the parent distribution. The primary reason is that the mapping from the distribution of order statistics to the corresponding parent distribution may not be Lipschitz continuous. The derivative of this mapping may diverge near the tail. In this context, we found similar issues with respect to the proposed sieve MLE using Berstein polynomials from simulation studies.

### *3.2. The conditional value distributions*

Theorem 3.2 concerns the distribution of UH and the conditional bid distributions. While the bid equals the value in ascending auctions, recovering the value distributions in first-price auctions requires several additional steps. First, we estimate the conditional bid quantile functions $\widehat{b}(\alpha|\tau)$ by inverting the estimated conditional bid distribution. That is, $\widehat{b}(\alpha|\tau) = \widehat{F}^{-1}(\alpha|\tau)$, where we have omitted supscript $X|\mathtt{T}$ for simplicity. Second, following [21], we can recover the conditional value quantile function

$$\widehat{v}(\alpha|\tau) = \widehat{b}(\alpha|\tau) + \frac{1}{n-1}\alpha\widehat{b}'(\alpha|\tau), \tag{11}$$

which allows construction of the conditional value density and distribution. By the continuous mapping theorem [11], the estimated conditional value quantile function, density, and distribution are also consistent. Moreover, if we impose higher-order smoothness assumptions on the value distribution, we may achieve a faster convergence rate, which is similar to the results in [42].

## 4. Simulation studies

In this section, we first conduct Monte Carlo experiments to demonstrate the proposed sieve estimator's finite sample performance using Bernstein polynomials as bases. The results show that the sieve estimators perform well even with a modest sample size. Second, we compare the performance of our proposed

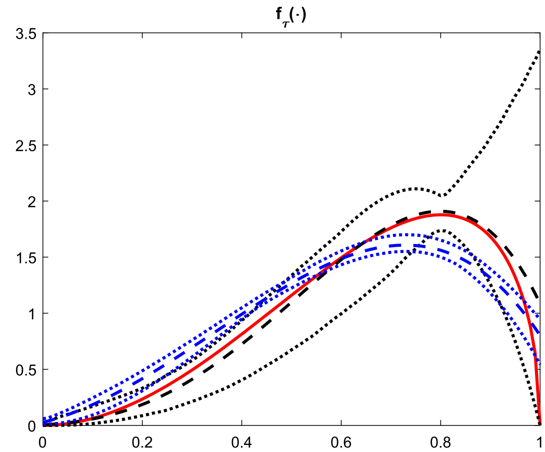



FIG 2. *Estimated $f^{\mathtt{T}}(\tau)$ with Bernstein Polynomials. Note: the red solid line depicts the true density. The black dashed line represents the average of the estimated densities based on our proposed method, with black dotted lines indicating the 5% and 95% pointwise quantiles. In contrast, the blue dashed line shows the average of the estimated density based on the naive imputation accompanied by blue dotted lines for the respective 5% and 95% pointwise quantiles.*

estimator with a naive imputation method to further provide evidence of the effectiveness of our method.

We focus on ascending auctions. The data-generating process is as follows. We first generate the latent variable $\mathtt{T}$ using beta distribution Beta$(\alpha^*, \beta^*)$, where $\alpha^* = 3$ and $\beta^* = 1.5$. We then generate the measurement/value/bid $X$ using the conditional distribution $f^{X|\mathtt{T}}(x|\tau)$, which is specified again as beta distributions, with $\alpha(\tau) = 1.5$ and $\beta(\tau) = 1.5(1+\tau)$. To generate a set of order statistics, we generate $n = 4$ measurements for each $\tau_t$ and record the lowest three, i.e., $X_{1:n} = x^t, X_{2:n} = y^t, X_{3:n} = z^t$. Repeating this process $1,000$ times produces a sample of $1,000$ consisting of the lowest three order statistics.

In estimation, we choose the number of basis functions to be $p = 5$. To guarantee that the sieve approximation functions are valid density functions, we impose restrictions on the sieve parameters $\theta$. Specifically, $\theta_{ij}$'s satisfy $\theta_{ij} \geq 0$ and $\sum_{ij} \theta_{ij} = 1$. For ease of computation, we re-parametrize those parameters as $\theta_{ij} = \exp(\gamma_{ij})/\{\sum_{ij} \exp(\gamma_{ij})\}$, where $\gamma_{11}$ is normalized to be 0, so that the restrictions are trivially satisfied and parameters are identifiable. We consider 100 independent simulation runs. Figures 2 and 3 display the pointwise 5%, mean, and 95% quantile of the estimated marginal distribution $f^{\mathtt{T}}(\tau)$ and the conditional distributions $f^{X|\mathtt{T}}(x|\tau)$ for $\tau = 0.25, 0.50, 0.75$, respectively. We observe that even with a sample size of 1000, our density estimator performs quite well with the mean very close to the true density and the 5% and 95% quantiles tightly cover the true density.

The identification problem we tackle can also be treated as a missing data problem. What is unique is that auction-level heterogeneity $\mathtt{T}$ is *completely* miss-

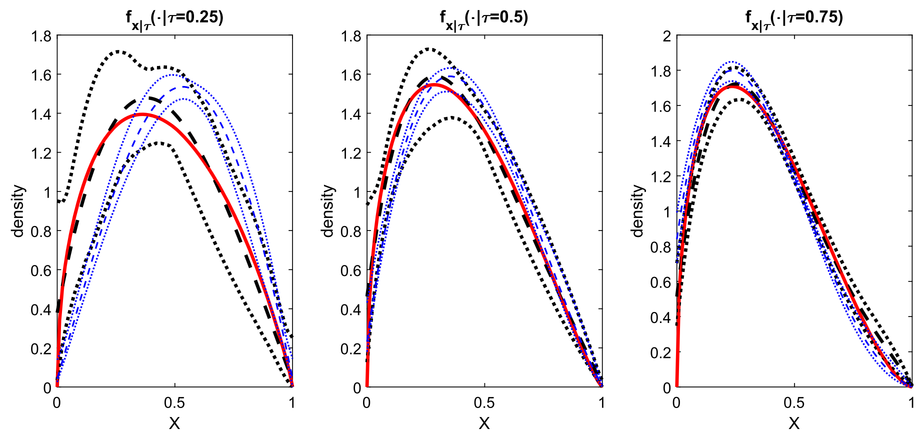



Fig 3. *Estimated $f^{X|\mathtt{T}}(x|\tau)$ for $\tau = 0.25, 0.50, 0.75$ with Bernstein Polynomials. Note: in each panel, the red solid line depicts the true density. The black dashed line represents the average of the estimated densities based on our proposed method, with black dotted lines indicating the 5% and 95% pointwise quantiles. In contrast, the blue dashed line shows the average of the estimated density based on the naive imputation accompanied by blue dotted lines for the respective 5% and 95% pointwise quantiles.*

ing. Our approach to this problem tackles the missing data directly and recovers the primitive for each unobserved level of UH. A natural question is how a naive imputation method performs. For example, we can fill in the missing auction-level heterogeneity $\mathtt{T}$ by the average of the observed order statistics

$$\widetilde{\mathtt{T}} = \frac{X_{r-2:n} + X_{r-1:n} + X_{r:n}}{3}.$$

Note that the larger the auction-level heterogeneity value, the more left-skewed the density is in our data-generating process. Therefore, the higher the imputed heterogeneity, the more likely the auction is from a less left-skewed density. To make the imputed heterogeneity comparable with the one from our proposed approach, we bring all imputed values into the range $[0, 1]$ and reverse their order. Specifically, we normalize $\widetilde{\mathtt{T}}$ by taking the additive inverse, subtracting the minimum, and dividing by the range in the simulation results. Intuitively such a naive imputation method leads to biased estimates.

We also conduct a simulation exercise to compare the performance of this naive method with that of our method. Specifically, after imputing the missing heterogeneity, we represent the log-likelihood function of the joint distribution of the three consecutive order statistics and the imputed heterogeneity, i.e., data $\equiv \{X_{r-2:n} = x^i, X_{r-1:n} = y^i, X_{r:n} = z^i, \widetilde{\mathtt{T}} = \widetilde{\tau}^i\}_{i=1}^m$, as follows:

$$\widetilde{\log L}(\text{data}; f^{X|\mathtt{T}}, \quad f^{\mathtt{T}}) = \tfrac{1}{m} \tfrac{n!}{(r-3)!(n-r)!} \textstyle\sum_{i=1}^m \log \Big\{ [F^{X|\mathtt{T}}(x^i|\tau)]^{r-3} f^{X|\mathtt{T}}(x^i|\tau)$$
$$f^{X|\mathtt{T}}(y^i|\tau)[1 - F^{X|\mathtt{T}}(z^i|\tau)]^{n-r} f^{X|\mathtt{T}}(z^i|\tau) f^{\mathtt{T}}(\tau) \Big\}.$$

We present the estimation results in Figures 2 and 3, from which we can see that the naive imputation leads to biased estimates of the marginal distribution $f^{\mathtt{T}}(\tau)$ and the conditional densities $f^{X|\mathtt{T}}(x|\tau=0.25)$ and $f^{X|\mathtt{T}}(x|\tau=0.50)$, but estimation of $f^{X|\mathtt{T}}(x|\tau=0.75)$ is less affected. This result validates the effectiveness of our proposed estimator. That is, our proposed estimator generates unbiased and consistent estimation of the underlying distributions, while the naive imputation estimation results in biased and inconsistent estimation due to the unsatisfactory treatment of the unobserved heterogeneity. However, the confidence band is narrower in the naive imputation method than that of our estimator because the unobserved heterogeneity is imputed.

After demonstrating that our estimator produces more unbiased and consistent estimations than the naive imputation, we conduct two additional sets of experiments to assess the performance of our estimator compared with the naive imputation method under various simulation settings. First, we investigate the performance of the proposed estimator using different conditional distribution functions. Specifically, we modify the $\beta$ parameter in the conditional distribution $f^{X|\mathtt{T}}(x|\tau)$ as follows: $\beta(\tau)=1.5+1.5\tau\gamma$, with $\gamma$ ranging from 0 to 1.8 in increments of 0.2. Intuitively, as $\gamma$ increases, the conditional distribution becomes more sensitive to the unobserved heterogeneity, indicating that the unobserved factor plays a bigger role in determining the overall value.

To compare the performance of our proposed estimator with that of the naive imputation method, we compute the average of the IMSE of $\hat{f}^{\mathtt{T}}(\tau)$[17] with varying values of $\gamma$ and the corresponding standard error, and plot these values in Figure 4. Several distinct patterns emerge from this comparison. First, our method performs better than the naive imputation when $\gamma$ is below 0.5, while the naive imputation outperforms our method as $\gamma$ increases. Secondly, the IMSE of the proposed estimator rises with increasing $\gamma$, indicating worsening performance with higher heterogeneity. Conversely, the naive imputation improves as $\gamma$ increases, since a simple average better approximates the unobserved heterogeneity when it plays a significant role in determining the overall value.

We conduct the same analysis for the conditional distributions and plot the average and standard error of IMSEs across 100 repetitions for both our proposed estimator and the naive imputation method in Figure 5. Similar patterns to those previously described emerge for the conditional distributions. Specifically, our proposed estimator performs better for $\tau=0.25$ and $\tau=0.5$ when $\gamma$ is low. However, for $\tau=0.75$, the pattern reverses, with our proposed estimator performing better as $\gamma$ increases

Secondly, we study how our proposed estimator performs with different numbers of measurements. Specifically, we vary the number of measurements, $n$, from 4 to 22, examining the impact of tail order statistics. Intuitively, an increase in the number of bidders ($n$) intensifies bid selectivity, reducing the informational value for estimation. Figure 6 displays the average IMSE of $\hat{f}^{\mathtt{T}}(\tau)$

[17] For an arbitrary function $g$ defined on [0, 1] and its estimate $\hat{g}$, we define integrated mean square error (IMSE) of $\hat{g}$ as $\int_0^1 \hat{g}(t)-g(t)^2dt$.

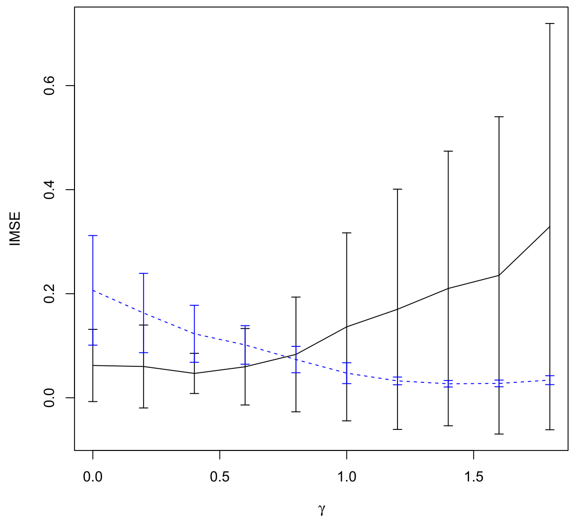
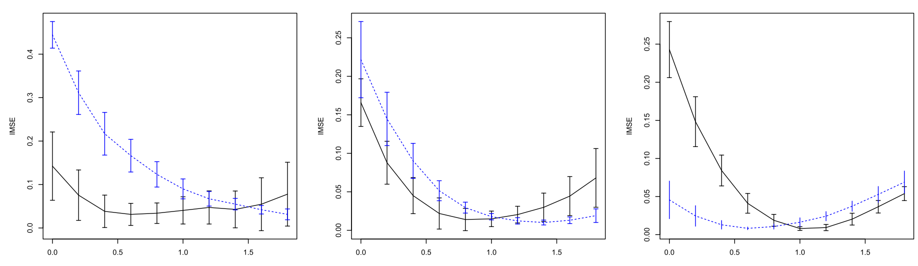



FIG 4. *The average and standard error of IMSEs of estimated $f^{\mathtt{T}}(\tau)$ for varying values of $\gamma$. The black solid and blue dashed lines represent the average of the IMSEs across 100 simulation runs of our proposed method and the imputation method, respectively.*



(a) $\tau = 0.25$ (b) $\tau = 0.5$ (c) $\tau = 0.75$

FIG 5. *The average and standard error of IMSEs of estimated $f^{X|\mathtt{T}}(x|\tau)$ for varying values of $\gamma$. The black solid and blue dashed lines represent the average of the IMSEs across 100 simulation runs of our proposed method and the imputation method, respectively.*

for both our method and the naive imputation method, plotted against varying values of $n$ with corresponding standard errors. As $n$ increases, the estimation of the marginal density of $\mathtt{T}$ by our proposed estimator becomes less accurate. This decline in accuracy also applies to our estimated conditional density of $X$ given $\mathtt{T} = \tau$. These findings underscore the conclusions drawn at the end of Section 3.1. In contrast, the naive imputation method performs more stably as $n$ varies.

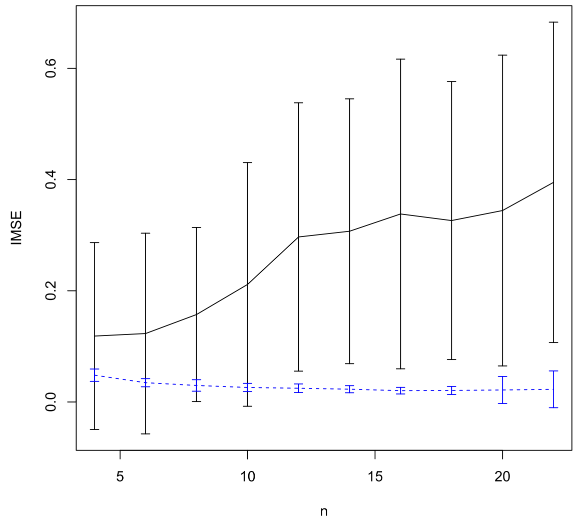



Fig 6. *The average and standard error of IMSEs of estimated $f^{\mathrm{T}}(\tau)$ for varying values of $n$. The black solid and blue dashed lines represent the average of the IMSEs across 100 simulation runs of our proposed method and the imputation method, respectively.*

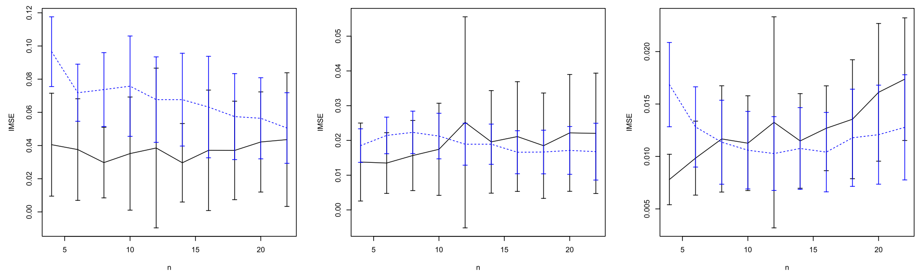
## 5. Empirical application

In this section, we apply our methodology to an empirical analysis of judicial auctions in China. Chinese courts began holding online auctions in 2012 through taobao.com, the shopping site of Chinese e-commerce giant Alibaba. As of 2022, almost all of China's courts have registered on this judicial sales platform, auctioning assets ranging from cars, diamonds, property, land use rights, and Boeing 747s to company shares. As of December 2019, over 500,000 items have been sold, with turnover reaching about 1.3 trillion yuan on the Taobao judicial sales platform.[18]

The court first posts the property-related information on taobao.com, including the appraisal value, obtained through a third-party appraisal company, and a starting price. Potential buyers can view the information page online and visit the property physically before the auction starts. Interested bidders can register to participate in the bidding by paying a security deposit and then bid in an ascending fashion. They can also set up automatic bidding.[19] The highest bidder wins the object and pays his/her bid.

[18] Source: China Daily.

[19] On average, a sold item receives 55 bids from 3 bidders, suggesting that jump bidding may not be a big concern.




(a) $\tau = 0.25$ (b) $\tau = 0.5$ (c) $\tau = 0.75$

Fig 7. *The average and standard error of IMSEs of estimated $f^{X|T}(x|\tau)$ for varying values of $n$. The black solid and blue dashed lines represent the average of the IMSEs across 100 simulation runs of our proposed method and the imputation method, respectively.*

### 5.1. Data

We collected a sample of residential property auctions from taobao.com, which contains all sales by the court in Jiangmen city of Guangdong Province between January 2018 and June 2020. We drop a few sales below ten thousand RMB or above five million RMB. In total, we have 477 auctions with 329 successful sales. By default, this court uses 70% of the appraisal value as the starting price, which also serves as a reserve price.

These auctions are subject to UH for many reasons. A third party provides appraisal based on available information at hand but may miss important details that become revealed upon careful study of the listing and a physical visit. For example, any unpaid electricity bills or property management fees of a sold property are the responsibility of the winning bidder. Some condos may have defects that are unknown to the appraisal firm. These unobserved factors constitute a significant portion of potential bidders' values. But how they enter bidder value is unknown. Therefore, it is preferable to retain flexibility when specifying how bidder value depends on UH and private information.

Following the literature, we homogenize the bids by dividing them by the appraisal value.[20] We further rescale the homogenized bids by dividing them by the maximum value in estimation but report the results in homogenized terms for convenience. As usual, the highest and second-highest bids are close to each other, both revealing information about the second-highest value among all bidders. To avoid redundant information, we use the highest bid as the second-highest value among all the bidders and exclude the second-highest bid from the data.[21]

[20]For the homogenization to be valid, we need either 1) the appraisal value to be realized before the realization of UH or 2) the seller or the third-party appraisal company to have the same access to UH but choose to ignore the additional knowledge.

[21]We obtained almost identical estimation and counterfactual results using the second highest bid as the second highest valuation.

Table 1 provides some summary statistics of our data consisting of 477 auctions of residential property. The data include the appraisal value in millions RMB, the number of bidders registered for participating in the auction (# of potential bidders), the number of bidders who submitted a bid (# of bidders), and whether the auction was successful (sold). For the auction that was successful, we also compute the ratio of the winning bid and the appraisal value. On average, each property is worth one million RMB, which is approximately \$140,000. Only about 70% of listings are sold successfully, at a transaction price close to the appraisal value on average.

TABLE 1
*Summary statistics.*

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| appraisal value (million RMB) | 477 | 1.02 | 0.91 | 0.108 | 4.97 |
| # of potential bidders | 477 | 5.285 | 5.867 | 0 | 31 |
| # of bidders | 477 | 2.182 | 1.946 | 0 | 10 |
| sold | 477 | 0.690 | 0.463 | 0 | 1 |
| $\frac{\text{winning bid}}{\text{appraisal value}}$ | 329 | 0.995 | 0.267 | 0.700 | 2.359 |

### *5.2. Empirical model with a binding reserve price*

Our empirical model accounts for the binding reserve price. Upon arrival, $N$ potential bidders observe the realization of UH $\tau$ and draw i.i.d. private values from $F^{X|\mathtt{T}}(\cdot|\tau)$. Those with a valuation higher than reserve price $R$ submit a bid equal to their value. As a result, the amount of truncation for a given UH is $F^{X|\mathtt{T}}(R|\tau)$, where $R = 0.7$.

Conditional on the number of potential bidders $N$, the probability of observing the bid vector $\boldsymbol{b}_n \equiv \{b_{1:n}, ...., b_{n-1:n-1}\}$ is

$$\int f^{\mathtt{T}}(\tau)p(n|N,\tau)g(\boldsymbol{b}_n|n,\tau)d\tau,$$

where $p(n|N,\tau) = C_{N,n}\left[1 - F^{X|\mathtt{T}}(R|\tau)\right]^n \left[F^{X|\mathtt{T}}(R|\tau)\right]^{N-n}$ is the probability of observing $n$ active bidders given the number of potential bidders $N$ and UH $\tau$, and $g(\boldsymbol{b}_n|n,\tau)$ represents the joint PDF of the bid vector including all active bids.[22] If $n = 0$, $g(\mathbf{0}|n,\tau) = 1$ because there is no bid. If $n = 1$, $g(R|1,\tau) = 1$ the bid will be $R$, as there is no reason to bid higher than the reserve price when there is only one bidder. If $2 \leq n \leq N$, the joint PDF simply becomes

$$g(\boldsymbol{b}_n|n,\tau) = n!\left[1 - \widetilde{F}^{X|\mathtt{T}}(b_{n-1:n})\right]\Pi_{j=1}^{n-1}\widetilde{f}^{X|\mathtt{T}}(b_{j:n}). \tag{12}$$

To estimate the model, we ignore the fact that we cannot identify $F^{X|\mathtt{T}}(\cdot|\tau)$

[22]Note that the identification requires the number of active bidders to be at least four. We pool bids from all auctions, including those with fewer than four active bidders, to improve estimation efficiency but rely on the auctions with $n \geq 4$ for identification.

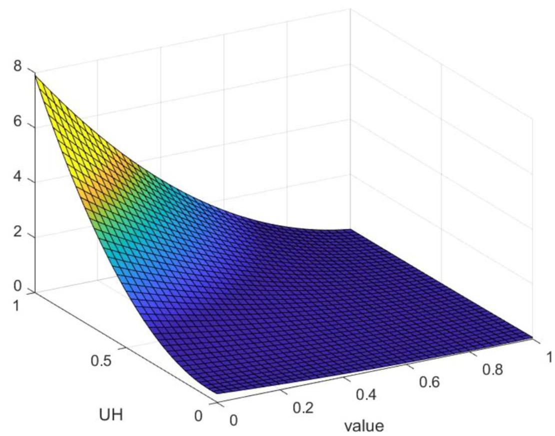



Fig 8. *Estimated Joint Density of UH and Bidder Value.*

below the reserve price[23] and approximate the joint density function using Berstein polynomials, $f(x,\tau;\theta) \approx \sum_{i,j} \theta_{ij}\beta_i(x)\beta_j(\tau)$, and solve the following optimization problem:[24]

$$\max_{\theta} \quad \sum_{\ell=1}^{L} \log \left[ \sum_j \left( \sum_i \theta_{ij} \right) \left\{ \int \beta_j(\tau) p(n_\ell | N_\ell, \tau; \theta) g(\boldsymbol{b}_\ell | n_\ell, \tau; \theta) d\tau \right\} \right]. \quad (13)$$

### 5.3. Empirical findings

We let the number of sieve bases $J = 3$. Figure 8 shows the estimated joint density function of bidder value $X$ and UH $\mathtt{T}$ in homogenized and rescaled terms. Two important features are worth noting. First, the conditional densities are skewed to the left. This suggests an abundance of low willingness-to-pay amongst the potential bidders in the market, consistent with the observation that the number of registered bidders exceeds the number of actual bidders. Second, UH has important effects on bidder value. The higher $\mathtt{T}$ is, the more skewed (to the left) the density becomes.

To demonstrate the practical use of our estimation results, we use the distribution estimated allowing for UH to calculate the optimal reserve price for

[23] Fortunately, this abstraction is barely binding for calculating the optimal reserve prices. In fact, [24] shows that as long as the existing reserve price is below the optimal, we obtain the same optimal $p^*$ by replacing $F_0$ and $f_0$ with the truncated version $F$ and $f$, respectively.

[24] We approximate the integration by Monte Carlo simulations

$$\int \beta_j(\tau) p(n_\ell | N_\ell, \tau) g(\boldsymbol{b}_\ell | n_\ell, \tau) d\tau \approx \frac{1}{S_j} \sum_{i=1}^{S_j} p(n_\ell | N_\ell, \tau_{ij}) g(\boldsymbol{b}_\ell | n_\ell, \tau_{ij}),$$

where $\tau_{ij}$ represent i.i.d. random draws from the beta density function $\beta_j(\cdot)$. By fixing the random draws, we make the maximization smooth in the sieve parameters $\theta$.

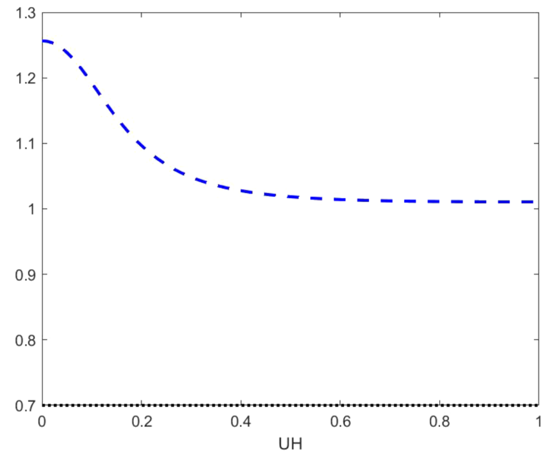



FIG 9. *UH-Specific Optimal Reserve Prices.*

each UH. We should emphasize that we assume UH is only unobserved or latent to researchers, but observed to bidders and auctioneers. So, it is practical for auctioneers to use such information to compute the optimal UH-specific reserve price. Given the number of potential bidders, the optimal reserve price maximizes

$$\pi(r,N) = N\,[1-F(r)]\,F(r)^{N-1}(r-v_0) + N(N-1)\int_r^{\overline{v}}(v-v_0)f(v)\,[1-F(v)]\,F(v)^{N-2}dv,$$

where $v_0$ is the seller's reserve value for keeping the item. The first term represents the seller's expected gain due to selling at the reserve price when only one value is higher than $r$, and the second term represents the gain due to selling at the second highest value when two values are higher than $r$. Its FOC leads to the following optimal reserve price

$$r^* = v_0 + \frac{1 - F^{X|\mathrm{T}}(r^*|\tau)}{f^{X|\mathrm{T}}(r^*|\tau)}, \tag{14}$$

which is strictly increasing in the reserve value. We can infer the auctioneer's reserve value from the series of judicial rules for judicial auctions issued by the Supreme Court. Specifically, one important rule says that the reserve price cannot be lower than 50% of the appraisal value. This seems a reasonable proxy for $v_0$, i.e., $v_0 = 0.5$.

Figure 9 shows the optimal reserve price for different levels of UH. The reserve price is strictly monotone in UH, which is consistent with the monotonicity assumption, i.e., Assumption 5, and the estimated joint density in Figure 8. It is also reassuring that the optimal reserve prices are well above the current reserve price, which means that underidentification below the reserve does not

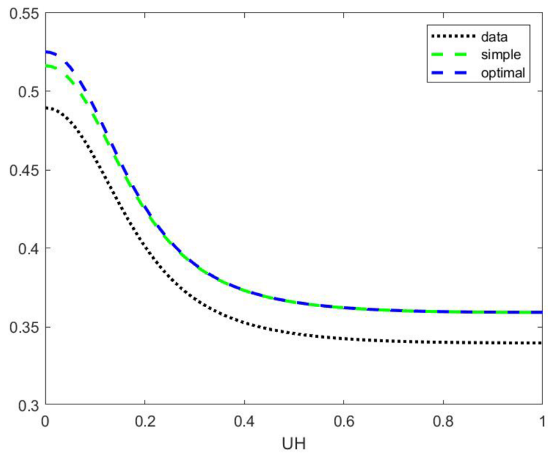



FIG 10. *Simple v.s. Optimal Reserve Price Schemes.*

prevent us from calculating the optimal reserve price.[25] In Figure 10, the blue dashed line shows the optimal expected seller gain as a function of UH. The unconditional optimal gain $\sum_N p_N \pi(r^*, N)$ is 36.61% of the appraisal value, which is 5.81% higher than the current one (34.60% of the appraisal value).

To achieve the optimal gain, the seller would need to know the UH and recover the conditional density of bidder values. However, the auctioneer may not have perfect information on UH or it is difficult to imagine that the seller adopts such a complex strategy. Simpler strategies that require less knowledge of the value distributions are often preferable.[26] In our context, we observe that the optimal reserve price is almost constant and close to one when UH is above 0.4. Moreover, the density of UH is heavily skewed to the right (near 1). Therefore, a simple alternative to a complex UH-specific reserve price is to use the appraisal value as the reserve price. We calculate the expected revenue in this simple scheme. In this case, the unconditional expected gain is 36.59% of the appraisal value, which achieves 98.85% of the potential gains from the optimal reserve prices.[27]

## 6. Conclusion

Auction data often contain incomplete bids and miss some payoff-relevant covariates. The conventional measurement error approaches to UH are inapplicable.

[25] The optimal reserve prices are still above 0.7 with more conservative values as low as $v_0 = 0.1$.

[26] [12] makes a similar point. They provide an approach to calculate optimal reserve prices without fully recovering value distributions. Our approach can be applied to other datasets, and similar simple strategies can be estimated as practical policy recommendations.

[27] The appraisal value as the reserve price is nearly optimal; this finding is robust to "large" auctions, different seller reserve values, and alternative tuning parameters.

In this paper, we extend the analysis of [29] to auctions with continuous UH while accounting for incomplete bid data. Specifically, we provide point identification results for auctions with nonseparable continuous UH using consecutive order statistics of bids. We then propose sieve maximum likelihood estimators jointly for the value distribution conditional on UH and its marginal distribution. We illustrate our methodology using a novel dataset from judicial auctions conducted by a municipal court in China. After recovering the model primitives, we propose a simple scheme that achieves nearly optimal revenue by using the appraisal value as the reserve price.

## Appendix A: Identification proofs

In this section, we provide the proof details omitted in the identification section.

### A.1. Derivation of Equation (2)

Ignoring UH for the moment, the joint distribution of any three order statistics does not have a multiplicatively separable structure, i.e., $f_{r,s,t:n}(x,y,z) \sim f(x)f(y)f(z)[F(x)]^{r-1}[F(y)-F(x)]^{s-r-1}[F(z)-F(y)]^{t-s-1}[1-F(z)]^{n-t}$, where $r<s<t$; see [15]. Considering consecutive order statistics, we have

$$\begin{aligned}
&f_{r-2,r-1,r:n}(x,y,z)\\
&= \frac{n!}{(r-3)!(n-r)!}[F(x)]^{r-3}f(x)f(y)[1-F(z)]^{n-r}f(z)\mathbb{1}(x\le y\le z)\\
&\equiv c_{r,n}f_{r-2:r-2}(x)\cdot f(y)\cdot f_{1:n-r+1}(z)\cdot \mathbb{1}(x\le y\le z),
\end{aligned}$$

where $c_{r,n}\equiv \frac{n!}{(r-2)!(n-r+1)!}$, $f_{r-2:r-2}(x)$ is the PDF of the top-order statistic of a sample of size $r-2$, and $f_{1:n-r+1}(z)$ is the PDF of the bottom-order statistic of a sample of size $n-r+1$, and $\mathbb{1}(\cdot)$ is an indicator function. Bringing back UH gives Equation (2) by the law of total probability.

### A.2. Derivation of the equivalence of operators (Equation (4))

We derive the equivalence of the operators as follows. Specifically, for any given $x\in\mathcal{X}_l$ and $y\in\mathcal{X}_m$, we have

$$\begin{aligned}
&[J_y g](x)\\
&\equiv \int_{\mathcal{X}_h} f_{r-2,r-1,r:n}(x,y,z)g(z)dz\\
&= \int_{\mathcal{X}_h} c_{r,n}\cdot \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)f^{X|\mathtt{T}}(y|\tau)f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)f^{\mathtt{T}}(\tau)d\tau g(z)dz\\
&= \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)c_{r,n}f^{X|\mathtt{T}}(y|\tau)f^{\mathtt{T}}(\tau)\left\{\int_{\mathcal{X}_h} f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)g(z)dz\right\}d\tau
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) c_{r,n} f^{X|\mathtt{T}}(y|\tau) f^{\mathtt{T}}(\tau) [H_{X_{1:n-r+1}|\mathtt{T}} g](\tau) d\tau \\
&\equiv \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) [\Delta_{X=y,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}} g](\tau) d\tau \\
&= [L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}} g](x),
\end{aligned} \tag{A.1}
$$

which implies that the operators from both sides are equivalent.

### *A.3. Injectivity condition*

Here, we prove that (i) the family of distributions $\{f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) : x \in \mathcal{X}_l\}$ is complete when the following condition is satisfied: if

$$
\int_{\mathcal{T}} g(\tau) F^{X|\mathtt{T}}(x|\tau)^{r-2} d\tau = 0
$$

holds for all $x \in \mathcal{X}_l$, we have $g(\tau) = 0$, $\forall \tau \in \mathcal{T}$; (ii) if the original distribution family is complete, i.e., $\{f^{X|\mathtt{T}}_{1:n-r+1}(x|\tau) : \tau \in \mathcal{T}\}$ over $\mathcal{L}^2(\mathcal{X})$ is complete, there exists at least one division of the support such that operator $H_{X_{1:n-r+1}|\mathtt{T}}$ is injective.

*Proof.* (i) The proof is achieved by contraction. Specifically, if the family distribution $\{f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) : x \in \mathcal{X}_l\}$ is not complete over $\mathcal{L}^2(\mathcal{T})$, there exists $\tilde{g}(\tau) \neq 0$, such that

$$
\int_{\mathcal{T}} (r-2) \tilde{g}(\tau) f^{X|\mathtt{T}}(x|\tau) F^{X|\mathtt{T}}(x|\tau)^{r-3} d\tau = 0.
$$

We can then integrate the component on the left-hand side of the above equation with respect to the support of $x$, leading to

$$
\int_{\mathcal{T}} \int_{x} (r-2) \tilde{g}(\tau) f^{X|\mathtt{T}}(x|\tau) F^{X|\mathtt{T}}(x|\tau)^{r-3} dx d\tau = \int_{\mathcal{T}} \tilde{g}(\tau) F^{X|\mathtt{T}}(x|\tau)^{r-2} d\tau = 0,
$$

which contradicts the provided condition. Therefore, we have proved that the family of distributions $\{f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) : x \in \mathcal{X}_l\}$ is complete given the condition in $(i)$ holds.

(ii) We only need to show that there exists at least a division, $\{\mathcal{X}_l, \mathcal{X}_m, \mathcal{X}_h\}$, such that the unique solution to $\int_{\mathcal{X}_h} g(x) f^{X|\mathtt{T}}_{1:n-r+1}(x|\tau) dx = 0$ for all $\tau \in \mathcal{T}$ is $g = 0$. We show this by contradiction. Suppose, for any division $\{\mathcal{X}_l, \mathcal{X}_m, \mathcal{X}_h\}$, there must exist a nonzero $g$ such that

$$
\int_{\mathcal{X}_h} g(x) f^{X|\mathtt{T}}_{1:n-r+1}(x|\tau) dx = 0,
$$

for any $\tau \in \mathcal{T}$.

We construct a new function $\tilde{g}$, which equals to $g$ on $\mathcal{X}_h$ and 0 otherwise. Obviously, $\tilde{g} \neq 0$, but we have

$$\int_{\mathcal{X}} \tilde{g}(x) f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau)dx = \int_{\mathcal{X}_h} g(x) f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau)dx = 0,$$

which holds for every $\tau \in \mathcal{T}$. This is contradictory to the assumption that the family of distribution $\{f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau) : \tau \in \mathcal{T}\}$ is complete on $\mathcal{L}^2(\mathcal{X})$. $\square$

### *A.4. Proof of Lemma 2.2*

*Proof.* We first show that $L^*_{X_{r-2:r-2}|\mathtt{T}}$, the adjoint operator of $L_{X_{r-2:r-2}|\mathtt{T}}$, is injective under Assumptions 1-3. To prove this, for any $g_1 \in \mathcal{L}^2(\mathcal{T})$ and $g_2 \in \mathcal{L}^2(\mathcal{X}_l)$, we have, from the definition of the adjoint operator,

$$\langle L_{X_{r-2:r-2}|\mathtt{T}} g_1, g_2 \rangle_{\mathcal{L}^2(\mathcal{X}_l)} = \langle g_1, L^*_{X_{r-2:r-2}|\mathtt{T}} g_2 \rangle_{\mathcal{L}^2(\mathcal{T})}, \tag{A.2}$$

where $\langle \cdot, \cdot \rangle$ is the inner product defined on the $\mathcal{L}^2$ space. The left-hand side of the equation above can be further expressed as

$$\int_{\mathcal{X}_l} \int_{\mathcal{T}} f_{r-2:r-2}^{X|\mathtt{T}}(x|\tau) g_1(\tau) d\tau g_2(x) dx.$$

Obviously, Equation (A.2) holds if and only if

$$[L^*_{X_{r-2:r-2}|\mathtt{T}} g_2](\tau) = \int_{\mathcal{X}_l} f_{r-2:r-2}^{X|\mathtt{T}}(x|\tau) g_2(x) dx.$$

The right-hand side of the equation above can be rewritten as

$$\int_{\mathcal{X}_l} f^{\mathtt{T}|X_{r-2:r-2}}(\tau|x) \cdot \frac{f^{X_{r-2:r-2}}(x)}{f^{\mathtt{T}}(\tau)} g_2(x) dx.$$

Thus, the adjoint operator $L^*_{X_{r-2:r-2}|\mathtt{T}}$ is injective, according to the equivalent of a family of distributions and given that $L_{\mathtt{T}|X_{r-2:r-2}}$ is injective.

Given that $L_{X_{r-2:r-2}|\mathtt{T}}$ is an operator from one Hilbert space to another Hilbert space, its null space is the complement of the closure of the range of $L^*_{X_{r-2:r-2}|\mathtt{T}}$, denoted as $\overline{\mathcal{R}(L^*_{X_{r-2:r-2}|\mathtt{T}})}$. Therefore, $L_{X_{r-2:r-2}|\mathtt{T}}$ is injective when it is viewed as a mapping of $\overline{\mathcal{R}(L^*_{X_{r-2:r-2}|\mathtt{T}})}$ to $\mathcal{L}^2(\mathcal{X}_l)$. It follows that $L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ exists.

Moreover, the closure of the range of $L_{X_{r-2:r-2}|\mathtt{T}}$, $\overline{\mathcal{R}(L_{X_{r-2:r-2}|\mathtt{T}})}$, is the orthogonal complement of the null space of $L^*_{X_{r-2:r-2}|\mathtt{T}}$. This null space is $\{0\}$, because $L^*_{X_{r-2:r-2}|\mathtt{T}}$ is injective. Consequently, $\overline{\mathcal{R}(L_{X_{r-2:r-2}|\mathtt{T}})} = \mathcal{L}^2(\mathcal{X}_l)$. Therefore, $L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ is defined over a dense subset of $\mathcal{L}^2(\mathcal{X}_l)$. $\square$

### A.5. Derivation of the main identification equation (Equation (5) )

We first derive the main equation for identification in the following. Specifically, we have the following equations for any two values of $y$:

$$J_{y_1} = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y_1,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}, \tag{A.3}$$
$$J_{y_2} = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y_2,\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}. \tag{A.4}$$

From Equation (A.4), we obtain the following equivalence of the operator

$$\Delta^{-1}_{X=y_2,\mathtt{T}} L^{-1}_{X_{r-2:r-2}|\mathtt{T}} J_{y_2} = H_{X_{1:n-r+1}|\mathtt{T}}, \tag{A.5}$$

which holds for the same domain $\mathcal{G}(\mathcal{X}_h)$ as in Equation (A.4) because the inverse operators ($L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ and $\Delta^{-1}_{X=y_2,\mathtt{T}}$) were applied from the left side of Equation (A.4) in the correct order.

We plug this equation back into Equation (A.3), leading to the following equation:

$$J_{y_1} = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y_1,\mathtt{T}} \Delta^{-1}_{X=y_2,\mathtt{T}} L^{-1}_{X_{r-2:r-2}|\mathtt{T}} J_{y_2}. \tag{A.6}$$

Note that the operator $J_{y_2}$ is injective, guaranteed by the injection of operators $L_{X_{r-2:r-2}|\mathtt{T}}$ and $H_{X_{1:n-r+1}|\mathtt{T}}$. Thus, we obtain the main equation for identification by right multiplying the inverse of the operator $J_{y_2}$:

$$J_{y_1} J^{-1}_{y_2} = L_{X_{r-2:r-2}|\mathtt{T}} \Delta_{X=y_1,\mathtt{T}} \Delta^{-1}_{X=y_2,\mathtt{T}} L^{-1}_{X_{r-2:r-2}|\mathtt{T}}.$$

### A.6. Proof of Theorem 2.3

*Proof.* This proof consists of the following steps: 1) we identify the conditional marginal distribution $f^{X|\mathtt{T}}(x|\tau)$ in segment $\mathcal{X}_l$ up to scales, ordering, and location, 2) we identify the conditional marginal distribution $f^{X|\mathtt{T}}(x|\tau)$ in segment $x \in \mathcal{X}_h$ up to scales, ordering, and location, 3) we identify the conditional marginal distribution $f^{X|\mathtt{T}}(x|\tau)$ in segment $x \in \mathcal{X}_m$ up to scales, ordering, and location. Moreover, the ordering and location of UH are consistently matched across all three segments, but scales may vary; 4) we then proceed to pin down the scales and exact location of UH, and 5) we identify the marginal distribution $f^{\mathtt{T}}(\tau)$.

**Step 1** The identification of step 1) mainly consists of the following three steps. First, we show that Equation (5) admits a unique representation. Second, we provide sufficient conditions under which the eigen-decomposition of the component on the left-hand side of Equation (5) is unique. Thus, the main equation for identification generates unique eigenfunctions $f^{X|\mathtt{T}}_{r-2:r-2}(x|\mathtt{T})$ for $x \in \mathcal{X}_l$. Lastly, we show that the conditional marginal distribution $f^{X|\mathtt{T}}(x|\tau)$ is identified in the same domain using the one-to-one mapping between the distribution of an order statistic and its parent distribution in a given segment. The detailed proofs for each step are as follows.

First, we show that Equation (5) admits a unique representation. The operator on the left-hand side of Equation (5), $J_{y_1}J_{y_2}^{-1}$, is determined by the densities of the observed three consecutive order statistics and so can be viewed as known. This equation implies that $J_{y_1}J_{y_2}^{-1}$ admits a spectral decomposition. More specifically, the "diagonal elements" of operator $\Delta_{X=y_1,\mathtt{T}}\Delta_{X=y_2,\mathtt{T}}^{-1}$, i.e., $\left\{\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}\right\}$ for given $y_1, y_2$ and for all $\tau$, and the kernel of the integral operator $L_{X_{r-2:r-2}|\mathtt{T}}$, i.e., $\{f_{r-2:r-2}^{X|\mathtt{T}}(.|\tau)\}$ for all $\tau$, are the eigenvalues and eigenfunctions of operator $J_{y_1}J_{y_2}^{-1}$, respectively.

We follow the sufficient and necessary conditions provided in Theorem XV.4.5 in [16] for the existence of a unique representation via spectral decomposition of a linear operator. Specifically, if a bounded and linear operator $A$ can be written as $A = U + V$, and $U$ is an operator represented as

$$U = \int_\sigma \lambda Q(d\lambda), \tag{A.7}$$

where $Q$ is a projection-valued measure with the support being spectrum $\sigma$, a subset of the complex field, and $V$ is a "quasi-nilpotent" operator computing with $U$, then this representation is unique. We apply this general result to our problem where $A = J_{y_1}J_{y_2}^{-1}$, $\sigma \subset \mathbb{R}$, and $V = 0$. Thus, we need to prove that linear operator $A = J_{y_1}J_{y_2}^{-1}$ is bounded, which can be accomplished by showing that the spectrum of operator $L_{X_{r-2:r-2}|\mathtt{T}}\Delta_{X=y_1,\mathtt{T}}\Delta_{X=y_2,\mathtt{T}}^{-1}L_{X_{r-2:r-2}|\mathtt{T}}^{-1}$ is a compact set, so that this operator is bounded because it is positive and its spectrum is a compact set.

First, we prove that the spectrum of $L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L_{X_{r-2:r-2}|\mathtt{T}}^{-1}$ is a compact set. Because $\frac{f^{X|\mathtt{T}}(y_1|\cdot)}{f^{X|\mathtt{T}}(y_2|\cdot)}$ is continuous and bounded by Assumption 2, denote the range of $\left\{\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)} : \tau \in \mathcal{T}\right\}$ as $\mathcal{I} = [\lambda_1(y_1, y_2), \lambda_2(y_1, y_2)]$. Denote the spectrum of $L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L_{X_{r-2:r-2}|\mathtt{T}}^{-1}$ as $\sigma$. We will show that $\mathcal{I} = \sigma$, which consists of the following two parts.

(i) $\sigma \subset \mathcal{I}$. Define $\mathcal{D}_{y_1,y_2}$ as an operator from $\mathcal{L}^2(\mathcal{T})$ to $\mathcal{L}^2(\mathcal{T})$ such that

$$(\mathcal{D}_{y_1,y_2}g)(\tau) = \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}g(\tau)$$

for any $g \in \mathcal{L}^2(\mathcal{T})$. Let $\sigma_\mathcal{D}$ denote the spectrum of $\mathcal{D}_{y_1,y_2}$. By definition,

$$[(\mathcal{D}_{y_1,y_2} - \lambda \cdot id)g](\tau) = \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}g(\tau) - \lambda g(\tau) = \left\{\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)} - \lambda\right\}g(\tau).$$

For any $\lambda \notin \mathcal{I}$, $f^{X|\mathtt{T}}(y_1|\tau)f^{X|\mathtt{T}}(y_2|\tau)^{-1} - \lambda \neq 0$ for any $\tau \in \mathcal{T}$. Following the last displayed equation, $[(\mathcal{D}_{y_1,y_2} - \lambda \cdot id)g](\tau) \equiv 0$ implies that $g \equiv 0$. Hence $\lambda \notin \sigma_\mathcal{D}$.

Next, we show that $L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L_{X_{r-2:r-2}|\mathtt{T}}^{-1} - \lambda \cdot id$ is invertible if $\lambda \notin \mathcal{I}$. Actually, given that

$$L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L_{X_{r-2:r-2}|\mathtt{T}}^{-1} - \lambda \cdot id = L_{X_{r-2:r-2}|\mathtt{T}}(\mathcal{D}_{y_1,y_2} - \lambda \cdot id)L_{X_{r-2:r-2}|\mathtt{T}}^{-1},$$

which is invertible because $(\mathcal{D}_{y_1,y_2} - \lambda \cdot id)$ is invertible from $\lambda \notin \sigma_{\mathcal{D}}$, it is thus bijective if $\lambda \notin \mathcal{I}$; namely $\lambda \notin \sigma$ if $\lambda \notin \mathcal{I}$. It follows that $\sigma \subset \mathcal{I}$.

(ii) $\mathcal{I} \subset \sigma$. It is straightforward to verify that, for any $\lambda \in \mathcal{I}$, we can find a non-zero $g \in \mathcal{L}(\mathcal{X}_l)$ such that $(L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}} - \lambda \cdot id)g \equiv 0$. Actually, we can take $g$ as a nonzero constant function to satisfy the equation above.

Combining (i) and (ii), we have $\sigma = \mathcal{I}$.

Second, we show that $L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ is a bounded operator. By definition, we have

$$\begin{aligned}
&\|L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}\| \\
=& \sup_{\|u\|_{\mathcal{L}^2}=1} \langle L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}u, L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}u\rangle_{\mathcal{L}^2} \\
=& \sup_{\|u\|_{\mathcal{L}^2}=1} \int_{\mathcal{I}} \lambda^2 d\mu_{u,u}(\lambda) \\
\leq& \sup_{\lambda\in\mathcal{I}} |\lambda|^2 \sup_{\|u\|_{\mathcal{L}^2}=1} \|\mu_{u,u}\| \leq \sup_{\lambda\in\mathcal{I}} |\lambda|^2 < \infty.
\end{aligned}$$

The last inequality follows from $\mathcal{I}$ and is compact. Therefore,

$L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$ is a bounded operator.

As a result, we prove that $J_{y_1}J^{-1}_{y_2}$ is bounded because

$$J_{y_1}J^{-1}_{y_2} = L_{X_{r-2:r-2}|\mathtt{T}}\mathcal{D}_{y_1,y_2}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}.$$

We define the projection-valued measure $Q$ in the following way: For any $\Lambda \subset \mathbb{R}$,

$$Q(\Lambda) = L_{X_{r-2:r-2}|\mathtt{T}}I_\Lambda L^{-1}_{X_{r-2:r-2}|\mathtt{T}},$$

where operator $I_\Lambda$ is defined as

$$[I_\Lambda g](\tau) = \mathbb{1}\left(\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)} \in \Lambda\right) g(\tau),$$

we want to show that $\int_\sigma \lambda Q(d\lambda) = L_{X_{r-2:r-2}|\mathtt{T}}\Delta_{X=y_1,\mathtt{T}}\Delta^{-1}_{X=y_2,\mathtt{T}}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$.

Based on the definition of $Q$, we have

$$\begin{aligned}
\int_\sigma \lambda Q(d\lambda) &= \int_\sigma \lambda \left(\frac{d}{d\lambda}Q((-\infty,\lambda])\right) d\lambda \\
&= L_{X_{r-2:r-2}|\mathtt{T}}\left(\int_\sigma \lambda \frac{d\mathbb{1}_{(-\infty,\lambda]}}{d\lambda} d\lambda\right) L^{-1}_{X_{r-2:r-2}|\mathtt{T}}.
\end{aligned}$$

To find the operator $\int_\sigma \lambda \frac{d\mathbb{1}_{(-\infty,\lambda]}}{d\lambda} d\lambda$, we investigate its evaluation when operating on a function $g$. That is,

$$\left[\int_\sigma \lambda \frac{d\mathbb{1}_{(-\infty,\lambda]}}{d\lambda} d\lambda g\right](\tau) = \int_\sigma \lambda \frac{d}{d\lambda}\mathbb{1}\left(\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)} \in (-\infty,\lambda]\right) g(\tau) d\lambda$$

$$= \int_\sigma \lambda\delta\left(\lambda - \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}\right) g(\tau)d\lambda$$
$$= \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)} g(\tau) = [\Delta_{X=y_1,\mathtt{T}}\Delta^{-1}_{X=y_2,\mathtt{T}} g](\tau),$$

where we have used the Dirac delta function $\delta$ satisfying the property that $\int \delta(x-x_0)h(x)dx = h(x_0)$ for any function $h$ continuous at $x = x_0$. It follows that $\int_\sigma \lambda Q(d\lambda) = L_{X_{r-2:r-2}|\mathtt{T}}\Delta_{X=y_1,\mathtt{T}}\Delta^{-1}_{X=y_2,\mathtt{T}}L^{-1}_{X_{r-2:r-2}|\mathtt{T}}$.

Second, we provide sufficient conditions under which the eigen-decomposition given in Equation (5) is unique. Note that the uniqueness of the representation in Equation (A.7) does not necessarily indicate that spectral decomposition of $J_{y_1}J^{-1}_{y_2}$ in Equation (5) is unique. This uniqueness problem is similar in spirit to a unique eigen-decomposition of a square matrix in the following two respects:

1. There is a unique eigen-space $S_\lambda$ spanning eigenfunctions corresponding to each eigenvalue $\lambda$. However, there are many different ways to select a basis for this space.

1a. Each basis can be multiplied by a constant. Unlike in [29], where the scaling problem of the eigenfunction $\{f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau)\}$ could be addressed using the fact that total probability is 1, we cannot apply the same logic in our context because the eigenfunction in our context is density functions in a segment of the full support.

1b. If the dimension of one eigen-space is larger than 1, then a new eigenfunction can be constructed through a linear combination of original basis functions.

2. Our economic model indexes the eigenvalues by $\tau$ and establishes the one-to-one mapping between eigenvalues and eigen-space. However, other methods can be used to index eigenvalues. In other words, if we use $\lambda(\tau)$ to denote the mapping between $\tau$ and $\lambda$ (and hence $S_{\lambda(\tau)}$), the choice of $\lambda(\tau)$ is not unique. The supplementary material of [29] shows non-uniqueness of indexing eigenvalues in some scenarios.

We harness Assumption 4 to address issue (1b). Note that the integral operator $L_{X_{r-2:r-2}|\mathtt{T}}$, with the kernel being the eigenfunction, depends on neither $y_1$ nor $y_2$, but the eigenvalues $\frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}$ do. If there exist two different values of $y$, say $y_1$ and $y_2$, such that there are two eigenfunctions $f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau_1)$ and $f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau_2)$ corresponding to the same eigenvalue, we can address this issue by simply finding another pair of $y$ that does not lead to this problem. In particular, for a given eigenfunction $f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau)$, let $D(y_1, y_2, \tau) = \left\{\tilde\tau : \frac{f^{X|\mathtt{T}}(y_1|\tilde\tau)}{f^{X|\mathtt{T}}(y_2|\tilde\tau)} = \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}\right\}$, the set of $\tau$ values that define eigenfunctions with the same eigenvalue. Then, any linear combination of eigenfunctions indexed by $\tilde\tau$ for $\tilde\tau \in D(y_1, y_2, \tau)$ is a potential candidate for the eigenfunctions of $J_{y_1}J^{-1}_{y_2}$. Then, define $v(\tau) \equiv \cap_{(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}}\text{span}(\{f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tilde\tau) : \tilde\tau \in D(y_1, y_2, \tau)\})$. If $v(\tau)$ is one-dimensional, this set will uniquely determine the eigenfunction

$f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau)$, even though such identification is up to scales. Next, we will show that if the set $v(\tau)$ has more than one dimension, then Assumption 4 would be violated. When the dimension of $v(\tau)$ is greater than one, we can at least find two eigenfunctions, say $f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tau)$ and $f^{X|\mathtt{T}}_{r-2:r-2}(\cdot|\tilde{\tau})$. Therefore, $\cap_{(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}} D(y_1,y_2,\tau)$ must contain at least two points $\tau$ and $\tilde{\tau}$. It follows that $\frac{f^{X|\mathtt{T}}(y_1|\tilde{\tau})}{f^{X|\mathtt{T}}(y_2|\tilde{\tau})} = \frac{f^{X|\mathtt{T}}(y_1|\tau)}{f^{X|\mathtt{T}}(y_2|\tau)}$ for any $(y_1,y_2)\in\mathcal{Y}\times\mathcal{Y}$ by definition of $D(y_1,y_2,\tau)$. Thus, Assumption 4 is violated.

Lastly, we exploit the one-to-one mapping between $f^{X|\mathtt{T}}_{r-2:r-2}(x|\mathtt{T})$ and the parent distribution $f^{X|\mathtt{T}}(x|\mathtt{T})$ to identify $f^{X|\mathtt{T}}(x|\mathtt{T})$ for $x\in\mathcal{X}_l$. Specifically,

$$\begin{aligned}
f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau) &= (r-2)[F^{X|\mathtt{T}}(x|\tau)]^{r-3}f^{X|\mathtt{T}}(x|\tau)\\
\leftrightarrow \quad \int_{\underline{x}}^{x} f^{X|\mathtt{T}}_{r-2:r-2}(v|\tau)dv &= (r-2)\int_{\underline{x}}^{x}[F^{X|\mathtt{T}}(v)]^{r-3}f^{X|\mathtt{T}}(v|\tau)dv\\
\leftrightarrow \quad \int_{\underline{x}}^{x} f^{X|\mathtt{T}}_{r-2:r-2}(v|\tau)dv &= [F^{X|\mathtt{T}}(x|\tau)]^{r-2}\\
\leftrightarrow F^{X|\mathtt{T}}(x|\tau) &= \Big[\int_{\underline{x}}^{x} f^{X|\mathtt{T}}_{r-2:r-2}(v|\tau)dv\Big]^{\frac{1}{r-2}},
\end{aligned}\tag{A.8}$$

where the first equality holds by definition, the second by taking the integral over $\bar{x}$ to any value in the low segment. Therefore, we can identify the conditional CDF of the parent distribution at the low segment. Additionally, we can also derive the conditional marginal distribution for $x\in\mathcal{X}_l$ as follows:

$$f^{X|\mathtt{T}}(x|\tau) = \frac{1}{r-2}\Big[\int_{\underline{x}}^{x} f^{X|\mathtt{T}}_{r-2:r-2}(v|\tau)dv\Big]^{\frac{1}{r-2}-1} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau). \tag{A.9}$$

Note that the conditional marginal distribution in the "low" portion is identified up to scales, ordering, and location.

**Step 2** We now proceed to identify $f^{X|\mathtt{T}}(x|\mathtt{T})$ for $x\in\mathcal{X}_h$, which is achieved given that operator $L_{X_{1:n-r+1}|\mathtt{T}}$ is identified up to scales, ordering, and location.

Specifically, we redefine the operator $J_y$ by abuse of notation:

$$[J_y g](z) \equiv \int_{\mathcal{X}_l} f_{r-2,r-1,r:n}(x,y,z)g(x)dx,$$

for any $y\in\mathcal{X}_m$. Then, $J_y$ is a map from $\mathcal{L}^2(\mathcal{X}_l)$ to $\mathcal{L}^2(\mathcal{X}_h)$ and satisfies

$$\begin{aligned}
[J_y g](z) &\equiv \int_{\mathcal{X}_l} f_{r-2,r-1,r:n}(x,y,z)g(x)dx\\
&= \int_{\mathcal{X}_l} c_{r,n}\cdot\int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)f^{X|\mathtt{T}}(y|\tau)f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)f^{\mathtt{T}}(\tau)d\tau g(x)dx\\
&= \int_{\mathcal{T}} f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)c_{r,n}f^{X|\mathtt{T}}(y|\tau)f^{\mathtt{T}}(\tau)\left(\int_{\mathcal{X}_l} f^{X|\mathtt{T}}_{r-2:r-2}(x|\tau)g(x)dx\right)d\tau
\end{aligned}$$

$$\begin{aligned}
&= \int_{\mathcal{T}} f_{1:n-r+1}^{X|\mathtt{T}}(z|\tau) c_{r,n} f^{X|\mathtt{T}}(y|\tau) f^{\mathtt{T}}(\tau) [H_{X_{r-2:r-2}|\mathtt{T}} g](\tau) d\tau \\
&\equiv \int_{\mathcal{T}} f_{1:n-r+1}^{X|\mathtt{T}}(z|\tau) [\Delta_{X=y,\mathtt{T}} H_{X_{r-2:r-2}|\mathtt{T}} g](\tau) d\tau \\
&= [L_{X_{1:n-r+1}|\mathtt{T}} \Delta_{X=y,\mathtt{T}} H_{X_{r-2:r-2}|\mathtt{T}} g](z). \qquad \text{(A.10)}
\end{aligned}$$

Using the same approach as Step 1 for two points $y_1$ and $y_2$ in $\mathcal{X}_m$, we are able to uniquely recover the conditional distribution $f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau)$ for $x \in \mathcal{X}_h$, up to scales, ordering, and location. However, the ordering of UH can be matched consistently with that in Step 1, because the eigenvalues are the same in both decomposition.

Given that the conditional distribution for order statistic $X_{1:n-r+1}$ in segment $\mathcal{X}_h$ is identified up to scales, ordering, and location, we now exploit the one-to-one mapping between the parent distribution and the distribution of its order statistic to identify the parent distribution in the high segment. Specifically,

$$\begin{aligned}
f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau) &= (n-r+1)[1-F^{X|\mathtt{T}}(x|\tau)]^{n-r} f^{X|\mathtt{T}}(x|\tau) \\
\leftrightarrow \quad \int_x^{\bar{x}} f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau) dx &= (n-r+1) \int_x^{\bar{x}} [1-F^{X|\mathtt{T}}(x|\tau)]^{n-r} f^{X|\mathtt{T}}(x|\tau) dx \\
\leftrightarrow \quad \int_x^{\bar{x}} f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau) dx &= [1-F^{X|\mathtt{T}}(x|\tau)]^{n-r+1} \\
\leftrightarrow F^{X|\mathtt{T}}(x|\tau) &= 1 - \Big[ \int_x^{\bar{x}} f_{1:n-r+1}^{X|\mathtt{T}}(x|\tau) dx \Big]^{\frac{1}{n-r+1}}. \qquad \text{(A.11)}
\end{aligned}$$

We can then identify the conditional marginal distribution $f^{X|\mathtt{T}}(x|\tau)$ for any $x \in \mathcal{X}_h$. Note that such an identification is up to scales but has the same ordering of UH as in $\mathcal{X}_l$.

**Step 3** The identification of $f^{X|\mathtt{T}}(y|\tau) f^{\mathtt{T}}(\tau)$ for $y \in \mathcal{X}_m$ is achieved using Equation (4). Specifically, since we have already identified $f^{X|\mathtt{T}}(x|\tau)$ for $x \in \mathcal{X}_l \cup \mathcal{X}_h$, the numerator of $\int_{y_1 \in \mathcal{X}_m} f^{X|\mathtt{T}}(y_1|\mathtt{T}) dy_1 / f^{X|\mathtt{T}}(y_2|\mathtt{T})$ is known. As a result, the denominator $f^{X|\mathtt{T}}(y_2|\tau)$ is uniquely specified for any $y_2 \in \mathcal{X}_m$. Note that such identification is up to scales, ordering, and location of UH.

**Step 4** In summary, we can identify the conditional distribution $f^{X|\mathtt{T}}(x|\tau)$ up to scales and ordering for the three segments $\mathcal{X}_l$, $\mathcal{X}_m$, and $\mathcal{X}_l$. Note that the scales may vary across the three segments. We invoke the continuity of the component PDFs and the total probability argument. The scales can be pinned down using the following three restrictions. First, the PDFs identified separately in the three segments should be the same at the cutoff points due to the continuity of the true conditional distributions. Second, the fact that each conditional distribution should integrate to 1 provides the third restriction on the scales. These restrictions uniquely identify the scales.

Once the scales are pinned down, we exploit Assumption 5 to resolve the indexing problem in issue (2). Given monotonicity, we can fix the ordering of

$\tau$ by using the conditional mean. And the further restriction on the support of UH pins down the exact location of UH.

**Step 5** The identification of the marginal distribution of UH is achieved by using the unconditional joint distribution of order statistics $X_{r-1:n}$ and $X_{r:n}$, which can be represented as

$$f_{r-1,r:n}(x,z) = \int_{\mathcal{T}} c^1_{r,n} f^{X|\mathtt{T}}_{r-1:r-1}(x|\tau) f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau) f^{\mathtt{T}}(\tau) d\tau$$

for any $x \leq z$. Let $K$ denote an operator mapping $g \in \mathcal{G}(\mathcal{X}_h)$ to $Kg \in \mathcal{G}(\mathcal{X}_l \cup \mathcal{X}_m)$ with the definition:

$$[Kg](x) \equiv \int_{\mathcal{X}_h} f_{r-1,r:n}(x,z) g(z) dz.$$

Then, based on the above equation, we have for any $x \in \mathcal{X}_l \cup \mathcal{X}_m$,

$$\begin{aligned}
[Kg](x) &\equiv \int_{\mathcal{X}_h} f_{r-1,r:n}(x,z) g(z) dz \\
&= \int_{\mathcal{X}_h} c^1_{r,n} \cdot \left( \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-1:r-1}(x|\tau) f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau) f^{\mathtt{T}}(\tau) d\tau \right) g(z) dz \\
&= \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-1:r-1}(x|\tau) c^1_{r,n} f^{\mathtt{T}}(\tau) \left( \int_{\mathcal{X}_h} f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau) g(z) dz \right) d\tau \\
&= \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-1:r-1}(x|\tau) c^1_{r,n} f^{\mathtt{T}}(\tau) [H_{X_{1:n-r+1}|\mathtt{T}} g](\tau) d\tau \\
&\equiv \int_{\mathcal{T}} f^{X|\mathtt{T}}_{r-1:r-1}(x|\tau) [\Delta_{\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}} g](\tau) d\tau \\
&= [L_{X_{r-1:r-1}|\mathtt{T}} \Delta_{\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}} g](x),
\end{aligned}$$

where the diagonal operator $[\Delta_{\mathtt{T}} g](\tau) \equiv c^1_{r,n} f^{\mathtt{T}}(\tau) g(\tau)$ for any $\tau \in \mathcal{T}$. That is to say, we obtain the following operator equivalence:

$$K = L_{X_{r-1:r-1}|\mathtt{T}} \Delta_{\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}.$$

Note that operator $H_{X_{1:n-r+1}|\mathtt{T}}$ is injective and identified; operator $L_{X_{r-1:r-1}|\mathtt{T}}$ is also known and injective since we have identified the conditional density of $f^{\mathtt{T}}(y|\tau)$. The injection of $L_{X_{r-1:r-1}|\mathtt{T}}$ can be easily derived from the injection of operator $L_{X_{r-2:r-2}|\mathtt{T}}$. Hence,

$$L^{-1}_{X_{r-2:r-2}|\mathtt{T}} K = \Delta_{\mathtt{T}} H_{X_{1:n-r+1}|\mathtt{T}}.$$

The left side of this equation is a specified kernel, which maps a function $g \in \mathcal{G}(\mathcal{X}_h)$ to $\int_{\mathcal{X}_h} c^1_{r,n} f^{\mathtt{T}}(\cdot) f^{X|\mathtt{T}}_{1:n-r+1}(z|\cdot) g(z) dz \in \mathcal{G}(\mathcal{T})$. Based on the one-to-one mapping between the operator and its kernel, $f^{\mathtt{T}}(\tau) f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)$ is identified for any $\tau \in \mathcal{T}$. As the conditional density $f^{X|\mathtt{T}}_{1:n-r+1}(z|\tau)$ has been identified previously, the marginal density of the latent factor, $f^{\mathtt{T}}(\tau)$, is then specified. $\square$

### *A.7. Proof of Corollary 2.5*

*Proof.* We show that when the number of potential bidders is observed and has a large support, we can identify the conditional truncation probability. Specifically, the distribution of the number of active bidders conditional on the number of potential bidders is a mixture of the Binomial distribution with the success probability being the truncation $1 - F^{X|\mathtt{T}}(R|\tau)$. Note that, the mixture weight is the marginal distribution of UH, which is already identified from the eigenfunction-decomposition. That is,

$$\begin{aligned}\Pr(n|N) &= \int_{\tau\in\mathcal{T}} \Pr(n|N,\tau)dF^{\mathtt{T}}(\tau)\\ &= \int_{\tau\in\mathcal{T}} C_{N,n}(1-F^{X|\mathtt{T}}(R|\tau))^n(F^{X|\mathtt{T}}(R|\tau))^{N-n}dF^{\mathtt{T}}(\tau),\end{aligned} \tag{A.12}$$

where $\Pr(n|N)$ is estimable from the data, $C_{N,n}$ is a constant, $F^{\mathtt{T}}(\tau)$ is identified, and $F^{X|\mathtt{T}}(R|\tau)$ is the unknown and the object of interest, which is continuous because the unobserved heterogeneity $\tau$ is continuous.

It is worth emphasizing that our goal is to identify the UH-specific truncated probability $F^{X|\mathtt{T}}(R|\tau)$ given that the mixture distribution $F^{\mathtt{T}}(\tau)$ can be viewed as known. In contrast, the identification results developed in the mixture literature mainly focus on identifying the mixture distribution of the success probability, which takes any value in the interval $[0,1]$. That is, let $p$ denote a general success probability, where $p \in [0,1]$, and $G(p)$ denote the cumulative distribution of $p$, and the goal is to identify $G(p)$ for any $p$.

We show that our identification problem can be viewed as the dual problem of that in the conventional mixture model. Specifically, define a random variable $P$, which is a function of the unobserved heterogeneity $\mathtt{T}$ using the truncated probability, $P \equiv 1 - F^{X|\mathtt{T}}(R|\mathtt{T})$. Suppose the conditional truncating probability varies with UH $\tau$. That is, for any $p \in [0,1]$, there is at most one $\tau$ such that $p = 1-F^{X|\mathtt{T}}(R|\tau)$. We can then reverse this relationship and represent the unobserved $\tau$ as a function of the success probability $p$, i.e., $\tau = F^{-1}_{X|\mathtt{T}}(1-p,R)$. Therefore, we can reformulate the above identification problem to resemble the problem in the mixture literature and thus can apply their results directly. That is,

$$\begin{aligned}\Pr(n|N) &= \int_{p\in[0,1]} C_{N,n}p^n(1-p)^{N-n}dF^{\mathtt{T}}(F^{-1}_{X|\mathtt{T}}(1-p,R))\\ &= \int_{p\in[0,1]} C_{N,n}p^n(1-p)^{N-n}dG(p),\end{aligned} \tag{A.13}$$

where the first equality holds by change of variables $\tau = F^{-1}_{X|\mathtt{T}}(1-p,R)$ and $p = 1 - F^{X|\mathtt{T}}(R|\tau)$, and the second equality holds by redefining $G(p) = F^{\mathtt{T}}(F^{-1}_{X|\mathtt{T}}(1-p,R))$, which is the cumulative distribution of $p$. Therefor, for some $p \in [0,1]$ that there does not exist an unobserved value $\tau$ such that $\tau = F^{-1}_{X|\mathtt{T}}(1-p,R)$, $G(p) = 0$. Note that $G(p)$ is unknown because $F^{X|\mathtt{T}}(R|\mathtt{T})$ is unknown and $F^{-1}_{X|\mathtt{T}}(1-p,R)$ is unknown.

We then show that the first $N$ moments of the mixture distribution $G(p)$ is identifiable from the distribution of the active bidders conditional on the number of potential bidders. The conditional distribution provides information on the first $N$ moments. Specifically,

$$\begin{aligned}
\Pr(n|N) &= E[C_{N,n}p^n(1-p)^{N-n}] \\
&= E[C_{N,n}p^n \sum_{r=0}^{N-n} C_{N-n,r}(-1)^r p^r] \\
&= \sum_{r=0}^{N-n} C_{N-n,r}C_{N,n}(-1)^r E[p^{r+n}] \\
&= \sum_{i=n}^{N} C_{N-n,i-n}C_{N,n}(-1)^{i-n}E[p^i] \\
&= \sum_{i=1}^{N} C_{N-n,i-n}C_{N,n}(-1)^{i-n}E[p^i], \qquad \text{(A.14)}
\end{aligned}$$

where the first equality is by change of variables, the second equality holds by expressing $(1-p)^{N-n}$, the fourth equality holds by letting $i = r + n$, and the last equality holds by $C_{N-n,i-n} = 0$ for $i - n \leq 0$. We can then rewrite the above equation in matrix form

$$P_N = Am_N, \qquad \text{(A.15)}$$

where $P_N \equiv \{\Pr(n = 1|N), ..., \Pr(n = N|N)\}^T$ is an $N \times 1$ column vector, $A \equiv \{C_{N-n,i-n}C_{N,n}(-1)^{i-n}\}_{i,n}$ is an $N \times N$ square matrix, and $m_N \equiv \{E[p], E[p^2], ..., E[p^N]\}^T$ is an $N \times 1$ column vector collecting the $N$ moments. Note that $A$ is invertible because it is an upper triangular matrix with nonzero diagonal elements. Therefore, we can identify the first $N$ moments of distribution $G(p)$.

If $N$ has a large support, following the exiting literature [22] that the distribution of a bounded variable is uniquely determined by its moments, we can identify all moments of $G(p)$ so that $G(p)$ is identified. Consequently, the truncated distribution $1 - F^{X|\mathtt{T}}(R|\tau)$ is identified. Specifically, for a given $p$, we have $G(p) = F^{\mathtt{T}}(\tau)$, so that we can recover $\tau$ by inverting the CDF of $F^{\mathtt{T}}$ since it is identified. Moreover, given that $p = 1 - F^{X|\mathtt{T}}(R|\tau)$, we can have $F^{X|\mathtt{T}}(R|\tau) = 1 - p$. Therefore, we can recover the value of the truncation distribution $F^{X|\mathtt{T}}(R|\tau)$ for any given $\tau$. □

## Appendix B: Properties of the sieve estimators

In this part, we show the consistency of the sieve MLEs and further investigate the consistency of estimates using B-splines and/or Bernstein polynomials.

We provide sieve estimators to approximate the joint distributions of bids and unobserved heterogeneity. This is different from the estimator proposed in [29], where they approximate the conditional distribution of bids and the marginal distribution of the unobserved heterogeneity. Our sieve estimator for approximating the joint distribution has the following advantages: First, HS concerns the relationship between the dependent variable and the independent variable, and the relationship between the latent variable and the instrumental variable. In particular, they consider a parametric model for the conditional density of the dependent variable given the latent variable and nonparametric models for the conditional density of the independent variable and that of the latent variable. For the latter two, they need to consider different sieve estimates. Consequently, for each value of the latent variable and the instrument, normalization is needed, which is more computationally burdensome. In contrast, we are concerned about the bid/value distribution given the latent variable. The order statistics of bids follow the same parent distribution. Hence, we can first use a bivariate sieve to approximate the joint density of the bid and the latent variable. The densities of interest can then be directly derived from the joint sieve estimator. Only one normalization step is needed when implementing our sieve estimation procedure. Second, in [29], the density of the dependent variable given the latent is assumed to be parametric, while the conditional densities of the independent variable and the latent variable are modeled in a nonparametric way. Model misspecification may be encountered due to the parametric assumption. However, our sieve estimator is under a fully nonparametric framework, and thus we can circumvent the issue of model misspecifications. In practice, concerns about slow convergence might be raised because of this fully nonparametric structure. However, numerical studies demonstrate the resulting estimator from the sieve estimator has a promising performance. Lastly, we adopt a new technique to derive the consistency of our estimator. More specifically, we establish a concentration inequality based on the bracketing entropy and, from there, we derive the consistency of our sieve estimator, while [29] uses a covering number.

### B.1. Asymptotic properties

*Proof of Lemma 3.1.* Since $(\hat{f}_m^{X|\mathtt{T}}, \hat{f}_m^{\mathtt{T}})$ maximizes $\boldsymbol{P}_m[\log G(x,y,z;f)]$ over the sieve space, where $\boldsymbol{P}_m$ denotes the empirical measure of the data $(x_i,y_i,z_i)_{i=1}^m$, it follows that

$$\boldsymbol{P}_m\left\{\log G(x,y,z;\hat{f}_m^{X|\mathtt{T}}, \hat{f}_m^{\mathtt{T}})\right\} \geq \boldsymbol{P}_m\left\{\log G(x,y,z;f_m^{X|\mathtt{T}}, f_m^{\mathtt{T}})\right\}.$$

Therefore, we have

$$\begin{aligned} m^{-1/2}\boldsymbol{G}_m\left\{\log\frac{G(x,y,z;\hat{f}_m^{X|\mathtt{T}},\hat{f}_m^{\mathtt{T}})}{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})}\right\} &\geq \boldsymbol{P}\left\{\log\frac{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})}{G(x,y,z;f_0^{X|\mathtt{T}},f_0^{\mathtt{T}})}\right\} \\ &+ \boldsymbol{P}\left\{\log\frac{G(x,y,z;f_0^{X|\mathtt{T}},f_0^{\mathtt{T}})}{G(x,y,z;\hat{f}_m^{X|\mathtt{T}},\hat{f}_m^{\mathtt{T}})}\right\}. \quad \text{(B.16)} \end{aligned}$$

□

*Proof of Theorem 3.2.* Under Assumption 7, to find the convergence rate of the sieve MLE, we only need to quantify the convergence rate of the estimation error in the sieve space.

Under Assumption 9, by Theorem 19.35 of [47], we have

$$\sqrt{m}\|\boldsymbol{P}_m - \boldsymbol{P}\|_{\mathcal{F}_m} \leq O_p(1) p_m Q_m^{c_0} \log p_m/\sqrt{m}$$

for some positive constant $c_0$. Let (I) and (II) denote the two terms of the right-hand side of (B.16), respectively. Since the functional $G$ is Lipschitz continuous with respect to each component, we have that

$$(I) \geq -O_p(1)\left\{\|f_m^{X|\mathtt{T}} - f_0^{X|\mathtt{T}}\|_{L_\infty} + \|f_m^{\mathtt{T}} - f_0^{\mathtt{T}}\|_{L_\infty}\right\} \geq -O_p(p_m^{-\beta}).$$

Note that (II) is the Kulback-Leibler information. We consider the Taylor expansion of it. Obviously, the first term in this expansion vanishes while the second-order term in the expansion has a lower bound,

$$O(e^{-c_1 Q_m})\big\|G(x,y,z;f_0^{X|\mathtt{T}}, f_0^{\mathtt{T}}) - G(x,y,z;\hat{f}_m^{X|\mathtt{T}}, \hat{f}_m^{\mathtt{T}}))\big\|_{L_2(P)}^2,$$

for some positive constant $c_1 > 1$.

Since we assume that the joint density is bounded, the joint probability measure $P$ is equivalent to the product of the Lebsgue measure in $[0,1]^3$. Therefore, by combining the results above, we have

$$\begin{aligned}&\int_{0\leq x\leq y\leq z\leq 1}\left\{G(x,y,z;f_0^{X|\mathtt{T}}, f_0^{\mathtt{T}}) - G(x,y,z;\hat{f}_m^{X|\mathtt{T}}, \hat{f}_m^{\mathtt{T}})\right\}^2 dxdydz\\&\leq O_p(1)B(m,p_m,Q_m),\end{aligned}$$

where $B(m,p_m,Q_m) = e^{c_2 Q_m} p_m \log p_m/\sqrt{m} + e^{c_2 Q_m}/p_m^{\beta}$, and $c_2 > 1$ is a constant.

Lastly, we show the consistency of $\hat{f}_m^{X|\mathtt{T}}$ and $\hat{f}_m^{\mathtt{T}}$. To do this, we can consider the square $L_2$ distance between $\int_{0\leq x\leq y\leq z\leq 1} G(x,y,z;f_0^{X|\mathtt{T}}, f_0^{\mathtt{T}})dxdz$ and $\int_{0\leq x\leq y\leq z\leq 1} G(x,y,z;\hat{f}_m^{X|\mathtt{T}}, \hat{f}_m^{\mathtt{T}})]^2 dxdz$. This upper bound still holds for this distance. After some simple algebra, it follows that

$$\int\left\{\hat{f}_m^{X|\mathtt{T}}(y|\tau)\hat{f}_m^{\mathtt{T}}(\tau) - f_0^{X|\mathtt{T}}(y|\tau)f_0^{\mathtt{T}}(\tau)\right\}^2 d\tau \leq O_p(1)B(m,p_m,Q_m).$$

We can further justify that $\|\hat{f}_m^{X,\mathtt{T}} - f_0^{X,\mathtt{T}}\|_{L^2}^2$ is bounded by $O_p(1)B(m,p_m,Q_m)$. Then, it is obvious that $B(m,p_m,Q_m)^{1/2}$ is the convergence rate of both $\hat{f}_m^{X|\mathtt{T}}(x|\tau)$ and $\hat{f}_m^{\mathtt{T}}(\tau)$. □

### *B.2. Popular sieve spaces in auctions*

In this subsection, we evaluate the regularity assumptions for B-splines and Bernstein polynomials, which are popular in empirical applications due to their flexibility and the ease with which shape can be imposed. See, e.g., [14] and [31] for using Bernstein polynomials in auctions.

*B.2.1. B-Splines*

We want to show that the proposed MLE using B-splines as a sieve base satisfies the regularity conditions in Theorem 3.2. We first express the joint density as

$$f^{X,\mathtt{T}}(x,\tau) = \frac{\exp\{\eta(x,\tau)\}}{\int\int \exp\{\eta(x,\tau)\}dxd\tau},$$

which is always positive and integrates to 1, qualifying it as a density function. As a result, the conditional and marginal densities can be expressed as $f^{X|\mathtt{T}}(x|\tau) = \frac{\exp\{\eta(x,\tau)\}}{\int \exp\{\eta(x,\tau)\}dx}$ and $f^{\mathtt{T}}(\tau) = \frac{\int_x \exp\{\eta(x,\tau)\}dx}{\int\int \exp\{\eta(x,\tau)\}dxd\tau}$, respectively.

Next, we employ a tensor product of B-spline basis functions to approximate $\eta(x,\tau)$. We first define an extended partition on the interval [0, 1], given by

$$\triangle_e = \{s_{-p+1} = \cdots = s_{-1} = 0 = s_0 < s_1 < \cdots < s_{K_m+1} = 1 = \cdots = s_{p+K_n}\},$$

where $p$ is the order of the spline basis and $K_m$ is the number of interior knots. Let $\{B_j^p(t)\}_{j=1}^{K_m+p}$ be a normalized B spline basis of order $p$ (degree $p-1$) associated with $\triangle_e$. The sieve space for the parameter $\eta(x,\tau)$ is defined as

$$\begin{aligned} &S_m(p,K_m,Q_m) = \\ &\Bigg\{\eta(x,\tau) : \sum_{i_1,i_2=1}^{K_m+p} |b_{i_1,i_2}| \le Q_m, \eta(x,\tau) \\ &\quad = \sum_{i_1,i_2=1}^{K_m+p} b_{i_1,i_2} B_{i_1}^p(x) B_{i_2}^p(\tau), \sum_{i_1,i_2=1}^{K_m+p} b_{i_1,i_2} B_{i_1}^p(0) = 0\Bigg\}. \end{aligned} \tag{B.17}$$

Here, $Q_m$ is a constant that depends on sample size $m$. The first two conditions in $S_m(p,K_m,Q_m)$ ensure that the sieve space is a compact set in a finite-dimensional space, and the third condition is equivalent to $\eta(0,\tau) = 0$, which is needed to ensure the identifiability of $\eta$.

We now prove that the sieve estimator using B-splines as a sieve base is consistent. Specifically, we show that the sieve base of B-splines satisfies Assumption 7, which is critical for consistency. We introduce the following two regularity conditions:

- (A1) For a known integer $k \ge 2$, the true conditional density $f_0(x,z)$ satisfies $\log f_0(x,\tau) \in W^{k,\infty}([0,1]^2)$, where $W^{k,\infty}([0,1]^2)$ is a Sobolev space consisting of the functions defined on $[0,1]^2$ with bounded $k$th derivative.
- (A2) $(Q_m, p_m)$ satisfies $Q_m = O(\log\log m)$ and $p_m = O(m^\alpha)$ with $0 < \alpha < 1/4$.

Assumption A1 characterizes the smoothness of the joint density function; Assumption A2 is equivalent to Assumption 8. Together with Assumption 6, these assumptions ensure that the logarithm of the true joint density function can be approximated sufficiently well by the tensor product of B-spline basis functions, i.e., Assumption 7 is satisfied. This is also one important reason why we prefer choosing B-spline as the univariate basis functions to construct the sieve space. We summarize this result in the following proposition.

**Lemma B.1.** *Under Assumptions A1 and 6, Assumption 7 is satisfied. That is, there exist $f_m^{X|\mathtt{T}}(x|\tau)$ and $f_m^{\mathtt{T}}(\tau)$, both of which are represented in terms of basis functions in the sieve space, such that*

$$\|f_m^{X|\mathtt{T}}(x|\tau) - f_0^{X|\mathtt{T}}(x|\tau)\|_{L_\infty([0,1]^2)} = O(p_m^{-k}),$$
$$\|f_m^{\mathtt{T}}(\tau) - f_0^{\mathtt{T}}(\tau)\|_{L_\infty([0,1]^2)} = O(p_m^{-k}).$$

*Proof of Lemma B.1.* Following [45], we define a linear operator $\mathcal{Q}_p$, which is a mapping from $W^{k,\infty}([0,1]^p)$ to the sieve space. More specifically, for any $g \in W^{k,\infty}([0,1]^p)$,

$$\mathcal{Q}_p[g] = \sum_{i_1,\ldots,i_p=1}^{K_m+L} \Gamma_{i_1,\ldots,i_p}[g] B_{i_1}^L(x_1)\cdots B_{i_p}^L(x_p),$$

where $\Gamma_{i_1,\ldots,i_p}$ are the linear functionals in $L_\infty([0,1]^p)$. This mapping satisfies that

$$\sum_{i_1,\ldots,i_p=1}^{K_m+L} |\Gamma_{i_1,\ldots,i_p}[g]| \le (2L+1)^p 9^{p(L-1)} ||g||_{L_\infty([0,1]^p)},$$

and by Theorem 12.7 of [45],

$$\|\mathcal{Q}_p[g] - g\|_{L_\infty([0,1]^p)} \le \frac{C(L)}{K_m^k} \|g\|_{W^{k,\infty}([0,1]^p)}.$$

Then, we define $\eta_m(x,\tau) = \mathcal{Q}_2[\log f_0(x,\tau)] - \mathcal{Q}_2[\log f_0(x,\tau)]|_{x=0}$, an element in the sieve space. Hence,

$$f_m^{X|\mathtt{T}}(x|\tau) = \frac{\exp\{\eta_m(x,\tau)\}}{\int_0^1 \exp\{\eta_n(x,\tau)\}dx}, \quad f_m^{\mathtt{T}}(\tau) = \frac{\int_0^1 \exp\{\eta_m(x,\tau)\}dx}{\int_0^1\int_0^1 \exp\{\eta_m(x,\tau)\}dxd\tau}.$$

As a result,

$$\|f_m^{X|\mathtt{T}}(x|\tau) - f_0^{X|\mathtt{T}}(x|\tau)\|_{L_\infty([0,1]^2)} \le O(1)\|\log f_0 - \mathcal{Q}_2[\log f_0]\|_{L_\infty([0,1]^2)} \le O(K_m^{-k}).$$

Moreover, the same bound holds for $\|f_m - f_0\|_{L_\infty([0,1]^2)}$ and $\|f_m^{\mathtt{T}} - f_0^{\mathtt{T}}\|_{L_\infty([0,1]^2)}$. □

We then prove that Assumption 9 is satisfied for sieve MLE using B-splines. If the sieve space is $\mathcal{S}_m$, constructed as in Equation B.17, and Assumption 6 and conditions A1-A2 are met, then the following space

$$\left\{\log\frac{G(x,y,z;\tilde{f}_m^{X|\mathtt{T}},\tilde{f}_m^{\mathtt{T}})}{G(x,y,z;f_m^{X|\mathtt{T}},f_m^{\mathtt{T}})} : \tilde{f}_m^{X|\mathtt{T}} = \frac{\exp\{\tilde{\eta}_m(x,\tau)\}}{\int_0^1 \exp\{\eta_m(x,\tau)\}dx},\right.$$
$$\left.\tilde{f}_m^{\mathtt{T}} = \frac{\int_0^1 \exp\{\tilde{\eta}_m(x,\tau)\}dx}{\int_0^1\int_0^1 \exp\{\tilde{\eta}_m(x,\tau)\}dxd\tau}, \tilde{\eta}_m \in S_m\right\}$$

satisfies Assumption 9. A similar result can be found in [51].

**Corollary B.2.** *Under Assumption 6 and the regularity conditions (A1) and (A2), the proposed MLE using B-splines as sieve bases is consistent.*

*B.2.2. Berstein polynomials*

We approximate the joint density function of $\mathtt{T}$ and $X$ using a mixture of beta distributions as follows:

$$f(x,\tau) = \sum_{i,j} \theta_{ij}\beta_i(x)\beta_j(\tau),$$

where $\beta_i(x)$ represents a beta density function with parameters $\alpha = i, \beta = p_m + 1 - i$, and $p_m$ is the number of components in the mixture.[28] We restrict $\theta_{ij} \geq 0$ and $\sum_{i,j\in\{1,\ldots,p_m\}} \theta_{ij} = 1$ to ensure that $f$ is a joint density function. That is,

$$\int_{x,\tau} f(x,\tau)dxd\tau = \sum_{i,j\in\{1,\ldots,p_m\}} \theta_{ij} = 1.$$

Given the approximation of the joint distribution, we can represent the marginal density function of $x^*$ as

$$f^{\mathtt{T}}(\tau) = \int_x f(x,\tau)dx = \int_x \sum_{i,j} \theta_{ij}\beta_i(x)\beta_j(\tau)dx = \sum_j \left\{ \left( \sum_i \theta_{ij} \right) \beta_j(\tau) \right\},$$

the conditional probability density function as

$$f^{X|\mathtt{T}}(x|\tau) = \frac{f(x,\tau)}{f^{\mathtt{T}}(\tau)} = \frac{\sum_{i,j}\theta_{ij}\beta_i(x)\beta_j(\tau)}{\sum_j[(\sum_i \theta_{ij})\beta_j(\tau)]},$$

and the conditional cumulative distribution function as

$$F^{X|\mathtt{T}}(x|\tau) = \int_{-\infty}^{x} f^{X|\mathtt{T}}(x|\tau)dx = \frac{\sum_{i,j}\theta_{ij}B_i(x)\beta_j(\tau)}{\sum_j[(\sum_i \theta_{ij})\beta_j(\tau)]},$$

where $B_i(x)$ represents the CDF of the beta distribution $\beta_i(x)$.

According to these derivations, given $\mathtt{T} = \tau$, we can write the density functions as mixtures of Beta distributions:

$$f^{\mathtt{T}}(\tau) = \sum_{j=1}^{p_m} w_{1j}\beta_j(\tau), \quad f^{X|\mathtt{T}}(x|\tau) = \sum_{i=1}^{p_m} w_{2i}(\tau)\beta_i(x)$$

for some nonnegative weights $\{w_{1j}, j = 1,\ldots,p_m\}$ and $\{w_{2i}(\tau), i = 1,\ldots,p_m\}$ that satisfy $\sum_{j=1}^{p_m} w_{1j} = 1$ and $\sum_{i=1}^{p_m} w_{2i}(\tau) = 1$. Let $\hat{f}_m^{X|\mathtt{T}}(x|\tau)$ and $\hat{f}_m^{\mathtt{T}}(\tau)$ denote the sieve MLE of $f_0^{X|\mathtt{T}}(x|\tau)$ and $f_0^{\mathtt{T}}$, respectively.

[28] One can allow different numbers for each of the two dimensions. We leave the optimal choice for future research.

The following result is from Theorem 5 of [43]. If we assume that $p_m = m^\alpha$ for some $\alpha \in (0,1)$, and the true joint density $f_0(x,\tau)$ is continuous and bounded away from 0, then we have for every $\epsilon > 0$,

$$P\{d_H(f_0^{X|\mathtt{T}}(x|\tau), \hat{f}_m^{X|\mathtt{T}}(x|\tau)) > \epsilon\} < 4\exp(-mc_1\epsilon^2),$$
$$P\{d_H(f_0^{\mathtt{T}}(\tau), \hat{f}_m^{\mathtt{T}}(\tau)) > \epsilon\} < 4\exp(-mc_2\epsilon^2)$$

for some positive constants $c_1$ and $c_2$, where

$$d_H(f,g) = \left[\int_0^1 \{\sqrt{f(x)} - \sqrt{g(x)}\}^2 dx\right]^{1/2}$$

is the Hellinger metric between two density functions with domain $[0,1]$. It follows that both $\hat{f}_m^{X|\mathtt{T}}(x|\tau))$ and $\hat{f}_m^{\mathtt{T}}(\tau)$ are consistent. A key technique to obtain this result is to make use of the fact that, when the sieve space $\mathcal{F}_m$ is taken as the mixture of Beta distributions with $p_m$ components, then $N_{[]}(\epsilon, \mathcal{F}_m, d_H) = O((1/\epsilon)^{p_m})$. See Proposition 2 of [43]. Thus, Assumption 9 is satisfied for the sieve space comprised of Beta mixtures.

Additionally, as stated in [18], if $g(x)$ is a continuously differentiable probability density on (0,1] with bounded second derivative, then there exists a beta mixture density, such that

$$\sup_{0<x\leq 1}\left|g(x) - \sum_{j=1}^{p_m}\theta_j\beta_j(x)\right| = O(p_m^{-1}).$$

Moreover, if $\sum_{j=1}^{p_m}\theta_j = 1$ with each $\theta_j \in [0,1]$, then

$$\sup_{0<x\leq 1}\left|\sum_{j=1}^{p_m}\theta_j\beta_j(x)\right| \leq p_m.$$

Therefore, if $\mathcal{B}_m$ is constructed using a mixture of Beta distributions, and the true joint density $f_0^{X,\mathtt{T}}$ is continuously differentiable and has bounded second derivatives, Assumptions 7-9 are satisfied with suitable conditions on $p_m$ and $Q_m$, as in Assumption 8.

**Corollary B.3.** *Assume that the true joint density $f_0^{X,\mathtt{T}}$ is continuously differentiable and has bounded second derivatives. Under Assumption 6 and suitable conditions on $p_m$ and $Q_m$, the proposed MLE using Berstein polynomials as a sieve base is consistent.*

### *B.3. Some sieve examples*

For reference, we provide a few alternative sieve bases: 1) Trigonometric linear series as the base, where the space of Trigonometric polynomials on the real line of degree $p$ or less can be represented as:

$$\text{TriPol}(p) = \left\{a_0, \sum_{k=1}^{p}[a_k\cos(2k\pi x) + b_k\sin(2k\pi x)], x\in[0,1]; a_k, b_k \in \mathbb{R}\right\};$$

2) Hermite polynomials as the base, where the space of Hermite polynomials on the real line of degree $p$ or less is represented as:

$$\mathrm{HPol}(p) = \left\{ \sum_{k=1}^{p+1} a_k H_k(x) \exp\left(-\frac{x^2}{2}\right), x \in \mathbb{R} : a_k \in \mathbb{R} \right\},$$

where $H_k(x)$ is the probabilists' Hermite polynomials.

## Appendix C: Identification using four order statistics

Here, we extend the main identification results to the situation with four order statistics that are not necessarily consecutive. The identification is also conducted on different segments of the support due to the special feature of order statistics. The identification process consists of several steps: 1) An eigenfunction-decomposition argument identifies a linear integral operator, which is defined using both the conditional distribution on a higher order statistic and the UH as kernel function on the low segment. Therefore, we can identify the conditional distribution in such a segment. 2) We use the fact that the operator defined on the low segment is identified to identify the operator defined on the high segment and so identify the conditional distributions. 3) Both identifications are up to scales, ordering, and location. We exploit similar features as in Theorem 1 to solve these problems.

Suppose we observe four order statistics that are not necessary consecutive. Denote the four order statistics as $r_i$, where $i = 1, \ldots, 4$ and $1 \leq r_1 < r_2 < r_3 < r_4 \leq n$. For simplicity of notation, let $x_i$ denote the realized values of order statistics $X_{r_i:n}$ and omit $n$ in all notation. The joint distribution of the observed four order statistics can be represented as

$$\begin{aligned}
&f_{r_1,r_2,r_3,r_4:n}(x_1, x_2, x_3, x_4) \\
&= \int_\tau f_{r_1,r_2,r_3,r_4:n}(x_1, x_2, x_3, x_4|\tau) f^{\mathrm{T}}(\tau) d\tau \\
&= \mathbb{1}(x_1 \leq x_2 \leq x_3 \leq x_4) \\
&\times \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) f_{r_4|r_3:n}(x_4|x_3, \tau) f_{r_2,r_3:n}(x_2, x_3|\tau) f^{\mathrm{T}}(\tau) d\tau.
\end{aligned}$$

The last equality holds due to the Markov property in order statistics. This joint distribution of four order statistics has a semi-separable structure, in the sense that we can separate the observed joint density function into three density functions. This, again, is similar to that in the measurement error literature, but it has an extra restriction by the nature of order statistics $\mathbb{1}(x_1 \leq x_2 \leq x_3 \leq x_4)$, which cannot be separated but can be controlled by dividing the support accordingly.

Once the correlation between order statistics is controlled by restricting the variation of the order statistics into the associated sub-support, we can derive the equivalence between the linear operator defined in the data and the unknown densities that we are yet to identify. Given that the injective and distinct assumptions are satisfied, we can then identify the latent density uniquely. Please refer to the identification details in Appendix A.

To control for the correlation between order statistics, we follow the identification argument in the situation with three consecutive order statistics. Specifically, we divide the support into four segments and only exploit the variations of $x$ in the predetermined segments: $x_1 \in \mathcal{X}_l \equiv \{x : x \le c_1\}, x_2 \in \mathcal{X}_{m1} \equiv [c_1, c_2], x_3 \in \mathcal{X}_{m2} \equiv [c_2, c_3], x_4 \in \mathcal{X}_h \equiv \{x : x \ge c_2\}$. The separable structure of the joint distribution $f_{r_1,r_2,r_3,r_4:n}(x_1, x_2, x_3, x_4)$ reappears then. Specifically, if $x_1 \in \mathcal{X}_l, x_2 \in \mathcal{X}_{m1}, x_3 \in \mathcal{X}_{m2}, x_4 \in \mathcal{X}_h$, the joint distribution can be expressed as

$$\begin{aligned} & f_{r_1,r_2,r_3,r_4:n}(x_1, x_2, x_3, x_4) \\ = & \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) f_{r_4|r_3:n}(x_4|x_3, \tau) f_{r_2,r_3:n}(x_2, x_3|\tau) f^{\mathtt{T}}(\tau) d\tau. \end{aligned}$$

**Step 1** We then exploit the equivalence of linear integral operators to identify the conditional distribution. Particularly, we can derive the following operator equivalence, fixing $x_2 \in \mathcal{X}_{m1}$ and $x_3 \in \mathcal{X}_{m2}$:

$$\begin{aligned} & [J_{x_2,x_3} g](x_1) \\ \equiv & \int_{x_4 \in \mathcal{X}_h} f_{r_1,r_2,r_3,r_4:n}(x_1, x_2, x_3, x_4) g(x_4) dx_4 \\ = & \int_{x_4 \in \mathcal{X}_h} \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) f_{r_4|r_3:n}(x_4|x_3, \tau) f_{r_2,r_3:n}(x_2, x_3|\tau) f^{\mathtt{T}}(\tau) g(x_4) dx_4 \\ = & \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) f_{r_2,r_3:n}(x_2, x_3|\tau) f^{\mathtt{T}}(\tau) \int_{x_4 \in \mathcal{X}_h} f_{r_4|r_3}(x_4|x_3, \tau) g(x_4) dx_4 d\tau \\ = & \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) f_{r_2,r_3:n}(x_2, x_3|\tau) f^{\mathtt{T}}(\tau) [H_{X_{r_4}|X_{r_3}=x_3,\mathtt{T}} g](x_4) \tau \\ \equiv & \int_\tau f_{r_1|r_2:n}(x_1|x_2, \tau) [\Delta_{x_2,x_3,\mathtt{T}} H_{X_{r_4}|X_{r_3}=x_3,\mathtt{T}} g](\tau) d\tau \\ = & [L_{X_{r_1}|x_2,\mathtt{T}} \Delta_{x_2,x_3,\mathtt{T}} H_{X_{r_4}|x_3,\mathtt{T}} g](x_1). \end{aligned} \tag{C.18}$$

Equation (C.18) implies that the operators from both sides are equivalent for any $x_2 \in \mathcal{X}_{m1}, x_3 \in \mathcal{X}_{m2}$. That is,

$$J_{x_2,x_3} = L_{X_{r_1}|x_2,\mathtt{T}} \Delta_{x_2,x_3,\mathtt{T}} H_{X_{r_4}|x_3,\mathtt{T}}. \tag{C.19}$$

Since such equivalence holds for any $x_2 \in \mathcal{X}_{m1}, x_3 \in \mathcal{X}_{m2}$, we first have the following equations at two different values of $(X_{r_2}, X_{r_3})$: $(c_1, x_3)$ and $(x_2, x_3)$, where $x_2 \in \mathcal{X}_{m1}$ and $x_3 \in \mathcal{X}_{m2}$, resulting in four matrix equations with common components:

$$\begin{aligned} J_{c_1,x_3} &= L_{X_{r_1}|c_1,\mathtt{T}} \Delta_{c_1,x_3,\mathtt{T}} H_{X_{r_4}|x_3,\mathtt{T}}, \\ J_{x_2,x_3} &= L_{X_{r_1}|x_2,\mathtt{T}} \Delta_{x_2,x_3,\mathtt{T}} H_{X_{r_4}|x_3,\mathtt{T}}, \end{aligned}$$

which share a common operator $H_{X_{r_4}|x_3,\mathtt{T}}$.

Similarly, we then have the following equations at two different values of $(X_{r_2}, X_{r_3})$, $(x_2, c_3)$ and $(c_1, c_3)$:

$$\begin{aligned} J_{x_2,c_3} &= L_{X_{r_1}|x_2,\mathtt{T}} \Delta_{x_2,c_3,\mathtt{T}} H_{X_{r_4}|c_3,\mathtt{T}}, \\ J_{c_1,c_3} &= L_{X_{r_1}|c_1,\mathtt{T}} \Delta_{c_1,c_3,\mathtt{T}} H_{X_{r_4}|c_3,\mathtt{T}}, \end{aligned}$$

which share a common operator $H_{X_{r_4}|c_3,\mathtt{T}}$.

We impose the following injective assumptions on all four operators:

**Assumption 10.** *(Injective) there exists one division of the domain such that the operators $L_{X_{r_1}|c_1,\mathtt{T}}$, $L_{X_{r_1}|x_2,\mathtt{T}}$, $H_{X_{r_4}|x_3,\mathtt{T}}$, and $H_{X_{r_4}|c_3,\mathtt{T}}$ are injective for $\mathcal{G} = \mathcal{L}^1$.*

With such an injective assumption being satisfied, we obtain the following main equation:

$$J_{c_1,x_3} J^{-1}_{x_2,x_3} J_{x_2,c_3} J^{-1}_{c_1,c_3} = L_{X_{r_1}|c_1,\mathtt{T}} \Delta_{c_1,x_2,x_3,c_3} L^{-1}_{X_{r_1}|c_1,\mathtt{T}}, \tag{C.20}$$

where the left-hand side matrix can be computed directly from the data, and the right-hand side matrix is the linear integral operator $L_{X_{r_1}|c_1,\mathtt{T}}$ with the conditional density, with diagonal operator $\Delta_{c_1,x_2,x_3,c_3}$ defined as

$$\Delta_{c_1,x_2,x_3,c_3} = \Delta_{c_1,x_3,\mathtt{T}} \Delta^{-1}_{x_2,x_3,\mathtt{T}} \Delta_{x_2,c_3,\mathtt{T}} \Delta^{-1}_{c_1,c_3,\mathtt{T}}.$$

Equation (C.20) indicates that the operator $J_{c_1,x_3} J^{-1}_{x_2,x_3} J_{x_2,c_3} J^{-1}_{c_1,c_3}$ can be represented as an eigenvalue-eigenfunction decomposition for the unknown operators $L_{X_{r_1}|c_1,\mathtt{T}}$ and $\Delta_{c_1,x_2,x_3,c_3}$ being the eigenvalues and eigenfunctions, respectively. The eigenfunctions $L_{X_{r_1}|c_1,\mathtt{T}}$, indexed by the latent factor, provide the unobserved conditional densities of order statistic $X_{r_1:n}|X_{r_2:n} = c_1, \mathtt{T}$.

For unique decomposition, we further impose restrictions on the relationship between the observed measurement $X$ in segment $\mathcal{X}_m$ and the latent factor $\mathtt{T}$. Specifically,

**Assumption 11.** *(Distinct) There exists one division of the domain such that the set*

$$\left\{ (x_2, x_3) : \frac{f_{r_2,r_3:n}(c_1, x_3|\tau_1) f_{r_2,r_3:n}(x_2, x_3|\tau_1)}{f_{r_2,r_3:n}(x_2, c_3|\tau_1) f_{r_2,r_3:n}(c_1, c_3|\tau_1)} \neq \frac{f_{r_2,r_3:n}(c_1, x_3|\tau_2) f_{r_2,r_3:n}(x_2, x_3|\tau_2)}{f_{r_2,r_3:n}(x_2, c_3|\tau_2) f_{r_2,r_3:n}(c_1, c_3|\tau_2)} \right\},$$

*has positive probability for all $\tau_1, \tau_2 \in \mathcal{T}$ whenever $\tau_1 \neq \tau_2$.*

With both assumptions satisfied, we can identify operator $L_{X_{r_1}|c_1,\mathtt{T}}$ up to scales from Equation (C.20) using eigenfunction decomposition. Additionally, we can identify the conditional density $f_{r_1|r_2:n}(x_1|c_1, \tau)$ using the fact that the identified operator is associated with this density. We further pin down the scales using the fact that $\int_{x_1 \in \mathcal{X}_l} f_{r_1|r_2:n}(x_1|c_1, \tau) dx_1 = 1$. Once the scales are pinned down, we identify the conditional density in segment "low," i.e., $f_{r_1|r_2:n}(x_1|c_1, \tau)$

for all $x_1 \leq c_1$. Note that the conditional distribution $f_{r_1|r_2:n}(x_1|c_1,\tau)$ is the same as the density of $r_1$th order statistics from a sample of size $(r_2-1)$ based on the parent distribution that is truncated on the right at $c_1$, i.e., $\frac{f^X(x|\tau)}{F^X(c_1|\tau)}$. Therefore, we identify this truncated distribution $\frac{f^X(x|\tau)}{F^X(c_1|\tau)}$, indicating that we identify the parent density $f^X(x|\tau)$ in segment "low" up to an unknown scale $F^X(c_1|\tau)$ for all $x_1 \leq c_1$.

**Step 2** Using the identified operator $L_{X_{r_1}|c_1,\mathtt{T}}$ in segment "low," we first identify an operator defined to be associated with the density in both segments "middle" and "high" using the joint density of the first three OS:

$$f_{r_1,r_2,r_3:n}(x_1,x_2,x_3) = \int_\tau f_{r_1|r_2:n}(x_1|x_2,\tau) f_{r_3|r_2:n}(x_3|x_2,\tau) f_{r_2:n}(x_2,\tau) d\tau.$$

We then exploit the equivalence of the linear integral operator when fixing $X_{r_2} = c_1$ to identify the conditional distribution. Particularly, we can derive the following operator equivalence for $X_{r_2} = c_1$ and any $X_{r_3} = y \geq c_1$. Specifically,

$$\begin{aligned}
[J_{c_1}g](x_1) &\equiv \int_{y\geq c_1} f_{r_1,r_2,r_3:n}(x_1,c_1,y)g(y)dy \\
&= \int_{y\geq c_1}\int_\tau f_{r_1|r_2:n}(x_1|c_1,\tau) f_{r_3|r_2:n}(x_3|c_1,\tau) f_{r_2:n}(x_2,\tau) d\tau dy \\
&= \int_\tau f_{r_1|r_2:n}(x_1|c_1,\tau) f_{r_2:n}(x_2,\tau) \int_{y\geq c_1} f_{r_3|r_2:n}(y|c_1,\tau)g(y)dyd\tau \\
&= \int_\tau f_{r_1|r_2:n}(x_1|c_1,\tau)[f_{r_2:n}(x_2,\tau) M_{X_{r_3}|c_1,\mathtt{T}}g](\tau)d\tau \\
&= [L_{X_{r_1}|c_1,\mathtt{T}}\Delta_{c_1,\mathtt{T}} M_{X_{r_3}|c_1,\mathtt{T}}g](x_1),
\end{aligned}$$

where $M_{X_{r_3}|c_1,\mathtt{T}}g](\tau) \equiv \int_{y\geq c_1} f_{r_3|r_2:n}(y|c_1,\tau)g(y)dy$ and $[\Delta_{c_1,\mathtt{T}}g](\tau) \equiv f_{r_2:n}(x_2,\tau)g(\tau)$. We obtain the equivalence of operators in the following:

$$J_{c_1} = L_{X_{r_1}|c_1,\mathtt{T}}\Delta_{c_1,\mathtt{T}} M_{X_{r_3}|c_1,\mathtt{T}}.$$

Therefore, we can identify the operator $M_{X_{r_3}|c_1,\mathtt{T}}$ up to scales, since the operator $L_{X_{r_1}|c_1,\mathtt{T}}$ is identified and operator $\Delta_{c_1,\mathtt{T}}$ is a diagonal. In addition, we identify the conditional density $f_{r_3|r_2:n}(x_3|c_1,\tau)$ up to scales. We can pin down the scales using the fact that $\int_{y\geq c_1} f_{r_3|r_2}(y|c_1,\tau)dx = 1$, so we identify fully $f_{r_3|r_2}(x_3|c_1,\tau)$ for all $x \geq c_1$. Note that the conditional distribution $f_{r_3|r_2}(x|c_1,\tau)$ is the same as the density of $(r_3-r_2)$th order statistics for a sample of size $(n-r_2)$ from a distribution that is truncated on the right at $c_1$, i.e., $\frac{f^X(x|\tau)}{1-F^X(c_1|\tau)}$. Therefore, we identify this truncated distribution $\frac{f^X(x|\tau)}{1-F^X(c_1|\tau)}$ and $f^X(x|\tau)$ up to an unknown scale $[1-F^X(c_1|\tau)]$ for all $x \geq c_1$.

**Step 3** To summarize, we identify the conditional distribution density $f^X(x|\tau)$ for $x \leq c_1$ up to an unknown scale $F^X(c_1|\tau)$ for all $x_1 \leq c_1$ and identify the conditional distribution $f^X(x|\tau)$ for $x \geq c_1$ up to an unknown scale $[1 - F^X(c_1|\tau)]$. We then pin down the unknown $F^X(c_1|\tau)$ using the smoothness of the conditional density, i.e.,

$$\frac{f^X(x|\tau)}{F^X(c_1|\tau)} = \frac{f^X(x|\tau)}{1 - F^X(c_1|\tau)},$$

which admits a unique and explicit solution for $F^X(c_1|\tau)$.

Note that the identification argument described above generates a continuous conditional distribution without knowing the associated value of the latent true factor $\mathtt{T}$. The identification is up to ordering and location. Pinning down the exact value of such location calls for extra restrictions, which typically depend on the context of the latent factor. We use the same restriction as in the previous section to pin down the location.

We summarize the identification result in the following theorem.

**Theorem C.1.** *If Assumptions 1, 2, 10, 11, and 5 are satisfied, the conditional density distribution $f^X(x|\mathtt{T})$ for $x \in \mathcal{X}$ and the marginal distribution for the latent variable $f^{\mathtt{T}}(\tau)$ for $\tau \in \mathcal{T}$ are identified using any four order statistics.*

## Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

## Funding

The first author was supported by Social Sciences and Humanities Research Council (SSHRC) 430-2023-00061.

The second author was supported in part by the Discovery grant (RGPIN-2020-04602) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

[1] Allen, J., Clark, R., Hickman, B. and Richert, E. (2023). Resolving failed banks: Uncertainty, multiple bidding & auction design. *Review of Economic Studies* rdad062. MR4743461

[2] An, Y. (2017). Identification of first-price auctions with non-equilibrium beliefs: A measurement error approach. *Journal of Econometrics* **200** 326–343. MR3684982

[3] Andreyanov, P. and Caoui, E. H. (2022). Secret reserve prices by uninformed sellers. *Quantitative Economics* **13** 1203–1256. MR4480426

[4] Aradillas-López, A., Gandhi, A. and Quint, D. (2013). Identification and inference in ascending auctions with correlated private values. *Econometrica* **81** 489–534. MR3043341

[5] Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review* **100** 724–762.

[6] Athey, S. and Haile, P. A. (2002). Identification of standard auction models. *Econometrica* **70** 2107–2140. MR1939892

[7] Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics* **6** 5549–5632.

[8] Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* **101** 1228–1240. MR2328309

[9] Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica* 289–314. MR1612238

[10] Cho, J., Luo, Y. and Xiao, R. (2022). Deconvolution From Two Order Statistics Technical Report, Working Paper, `https://ssrn.com/abstract=3733211`.

[11] Chung, K. L. (2000). *A Course in Probability Theory, 3rd edition*. Academic Press. MR1796326

[12] Coey, D., Larsen, B. J., Sweeney, K. and Waisman, C. (2021). Scalable optimal online auctions. *Marketing Science*.

[13] Compiani, G., Haile, P. and Sant'Anna, M. (2020). Common values, unobserved heterogeneity, and endogenous entry in US offshore oil lease auctions. *Journal of Political Economy* **128** 3872–3912.

[14] Compiani, G., Haile, P. and Sant'Anna, M. (2020). Common values, unobserved heterogeneity, and endogenous entry in US offshore oil lease auctions. *Journal of Political Economy* **128** 3872–3912.

[15] David, H. A. and Nagaraja, H. N. (2004). *Order Statistics*. John Wiley & Sons. MR1994955

[16] Dunford, N. and Schwartz, J. (1971). *Linear Operators*. New York: Wiley.

[17] Freyberger, J. and Larsen, B. J. (2022). Identification in ascending auctions, with an application to digital rights management. *Quantitative Economics* **13** 505–543. MR4438543

[18] Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics* **29** 1264–1280. MR1873330

[19] Grenander, U. (1981). *Abstract Inference*. John Wiley & Sons. MR0599175

[20] Guerre, E. and Luo, Y. (2022). Nonparametric identification of first-price auction with unobserved competition: a density discontinuity framework. Working Paper.

[21] Guerre, E., Perrigne, I. and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica* **68** 525–574. MR1769378

[22] Gut, A. (2005). *Probability: A Graduate Course*. New York, Springer. MR2125120

[23] Haile, P. A., Hong, H. and Shum, M. (2003). Nonparametric tests for common values at first-price sealed-bid auctions. Working Paper.

[24] Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of English auctions. *Journal of Political Economy* **111** 1–51.

[25] Hernández, C., Quint, D. and Turansick, C. (2020). Estimation in English auctions with unobserved heterogeneity. *The RAND Journal of Economics* **51** 868–904.

[26] Hortaçsu, A. and Perrigne, I. (2021). Empirical perspectives on auctions. In *Handbook of Industrial Organization*, **5** 81–175. Elsevier.

[27] Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* **144** 27–61. MR2439921

[28] Hu, Y., McAdams, D. and Shum, M. (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics* **174** 186–193. MR3045027

[29] Hu, Y. and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* **76** 195–216. MR2374986

[30] Komarova, T. (2013). A new approach to identifying generalized competing risks models with application to second-price auctions. *Quantitative Economics* **4** 269–328. MR3082775

[31] Kong, Y. (2020). Not knowing the competition: evidence and implications for auction design. *The RAND Journal of Economics*.

[32] Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies* **78** 293–327. MR2807728

[33] Krasnokutskaya, E. and Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review* **101** 2653–2686.

[34] Li, T., Perrigne, I. and Vuong, Q. (2000). Conditionally independent private information in OCS wildcat auctions. *Journal of Econometrics* **98** 129–161. MR1790650

[35] Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* **65** 139–165. MR1625869

[36] Liu, N. and Luo, Y. (2017). A Nonparametric Test for Comparing Valuation Distributions in First-Price Auctions. *International Economic Review* **58** 857–888. MR3696350

[37] Lu, J. and Perrigne, I. (2008). Estimating risk aversion from ascending and sealed-bid auctions: The case of timber auction data. *Journal of Applied Econometrics* **23** 871–896. MR2649058

[38] Luo, Y. (2020). Unobserved heterogeneity in auctions under restricted stochastic dominance. *Journal of Econometrics* **216** 354–374. MR4083104

[39] Luo, Y. and Takahashi, H. (2022). Bidding for Contracts under Uncertain Demand: Skewed Bidding and Risk Sharing Technical Report, University of Toronto, Department of Economics.

[40] Luo, Y. and Xiao, R. (2023). Identification of auction models using order statistics. *Journal of Econometrics* **236** 105457. MR4602955

[41] Mbakop, E. (2017). Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity Technical Report, Working Paper.

[42] Menzel, K. and Morganti, P. (2013). Large sample properties for estimators based on the order statistics approach in auctions. *Quantitative Economics* **4** 329–375. MR3082776

[43] Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 79–100. MR1881846

[44] Roberts, J. W. (2013). Unobserved heterogeneity and reserve prices in auctions. *The RAND Journal of Economics* **44** 712–732.

[45] Schumacker, L. (1981). *Spline Functions: Basic Theory*. New York, Wiley Interscience. MR0606200

[46] Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics* **25** 2555–2591. MR1604416

[47] Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, Cambridge University. MR1652247

[48] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics*. New York, Springer. MR1385671

[49] Wu, Y. and Zhang, Y. (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *The Annals of Statistics* **40** 1609–1636. MR3015037

[50] Xiao, R. (2018). Identification and estimation of incomplete information games with multiple equilibria. *Journal of Econometrics* **203** 328–343. MR3770830

[51] Zeng, D. (2005). Likelihood approach for marginal proportional hazards regression in the presence of dependent censoring. *The Annals of Statistics* **33** 501–521. MR2163149