Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity *

Eric Mbakop
For the latest draft, click here †

November 15, 2017

Abstract

This paper derives novel nonparametric identification results for auction models with incomplete bid data and finite unobserved heterogeneity (UH). By exploiting the Markov property of order statistics, I show that the joint distribution of bidders' valuations and the UH is point identified from an incomplete set of bids. The result holds if the econometrician either observes (any) five order statistics of the bids in each auction or only three along with an instrument, and without imposing any functional form restriction on how the UH affects valuations. This data structure is encountered in many empirical settings, such as ascending auctions in which the winner's bid is usually not observed. I establish these results under weak distributional assumptions. For second price auctions, the result holds generically over the space of possible distributions of valuations and UH, and for first price auctions, it holds when the conditional distribution of valuations varies monotonically with the UH in the reverse hazard rate order. I show that identification can be extended to settings where the number of potential bidders is unobserved, as is often the case in online auctions. Finally, I provide easily implementable nonparametric estimation procedures, and simulation results show that they perform well for samples of moderate size.

*I thank my committee, Ivan Canay, Robert Porter and Alexander Torgovitsky for their help and encouragement over the course of this project. Eric Auerbach, Vivek Bhattacharya, Sofya Budonova, Denis Chetverikov, Joel Horowitz, Gaston Illanes, Caleb Kwon, Matthew Leisten, Charles Manski, Matthew Masten, Mar Reguant, Alex Theisen, Daniel Wilhelm, and Gabriel Ziegler provided useful thoughts and comments.

†Department of Economics, Northwestern University. ericmbakop2011@u.northwestern.edu

1 Introduction

This paper studies the identification of the distribution of bidders' valuation in settings where the econometrician observes an incomplete set of bids from each auction in the data set, and does not observe all the variables that affect the distribution of bidders' valuation and that are commonly observed by all the auction participants− the auction level unobserved heterogeneity (UH). In general, identification of the distribution of bidders' valuations in auction models is important as it allows the researcher to determine (among other things) bidders' surplus, sellers' profits and to do counterfactual analysis of alternative auction mechanisms. Failing to account for UH in the econometric analysis (when it is present) can lead to incorrect inference of the structural parameters and to erroneous policy recommendation (see Krasnokutskaya (2011)). The papers in the literature that address the problem of UH in auction data, with the exception of a few, are mainly concerned with settings where the econometrician observes the bids of all auction participants in the data set. Moreover, the methods that they propose do not extend to settings where the econometrician only observes a subset of the order statistics of the bids−a case that is relevant for many empirical applications. Incomplete bid data arises naturally (for instance) in ascending auctions, where the winner's bid (exit price) is usually not observed[1].

The papers in the literature that study identification of auction models with incomplete bid data and unobserved heterogeneity (Roberts (2013) and Freyberger and Larsen (2017)) all rely on the availability of some auxiliary data: such as public reserve prices, secret reserve prices or instruments more generally. Moreover, to achieve identification, these papers either make strong assumptions on how the auxiliary variable is related to the UH (as in Roberts (2013)) or they rely on the assumption that the UH affects valuations (hence bids) and the auxiliary variable in an additively separable way (as in Freyberger and Larsen (2017)). From these papers, it is not clear whether identification is possible without relying on the availability of auxiliary variables− which are sometimes not easy to obtain in empirical applications− or without making strong functional form assumptions on how the UH relates to the observed variables−which may lead to severely biased estimates of the parameters of interest if the model is misspecified.

In an environment where bidders' valuations for the auctioned object are private, independent and symmetric (drawn from the same distribution) given the realization of a discrete UH, I provide in this paper two novel identification results that address the latter questions for both first and second price auctions. The first identification result shows that the conditional distribution of bidders' valuations given the UH as well as the marginal distribution of the UH are point identified if the econometrician has access to at least (any) five order statistics of the bids from each auction

[1]Recall that in the stylized push-button model of the English auction (see Milgrom and Weber (1982)), prices rise continuously (or in small increments) from a low value and each bidder chooses when to irreversibly exit the auction by releasing a pressed button. The auction ends when only one bidder remains and she obtains the object at a price equal to the exit price of her last competitor. See Roberts (2013) for an empirical example that uses such auctions.

in the data set. The result holds without making any functional form assumption on how the UH affects the distribution of valuations. In fact, the conditional distribution of valuations given the UH is allowed to vary in an unrestricted way (as in Hu, McAdams, and Shum (2013)); all that is required is that a "full rank" condition on the distribution of the observed order statistic of the bids is satisfied. In addition, I show that the required rank condition is "generic" (hence mild) in the context of second price auctions (SPA). This identification result is the first in the auction literature to show that identification of the distribution of valuations is possible in an auction environment with incomplete bid data and UH without relying on the availability of some auxiliary variable (in addition to bids). It thus answers in part an identification question raised in section 3.2 of Athey and Haile (2002), concerning the possibility of identification in auction models with UH from only a strict subset of the order statistics of the bids.

The second identification result in the present paper shows that identification (in the same setting) is possible with only three order statistics of the bids (from each auction in the data set) if the econometrician also observes an instrument. Unlike Roberts (2013) and Freyberger and Larsen (2017), I allow the UH to change the distribution of the instrument in an unrestricted way, all that is required −in analogy to the usual relevance condition on the instrument in the linear IV model− is that the set of conditional distributions of the instrument given different values of the UH satisfy a full rank condition. As in the first identification result, I also require the distributions of the observed order statistics of the bids to satisfy a full rank condition. In the context of SPA, I show again that the required full rank condition on the distribution of order statistics of the observed bids holds generically. In the context of first price auctions (FPA), I provide a simple and easily interpretable monotonicity condition on the distribution of bidders' valuation given different values of the unobserved heterogeneity, that implies the desired full rank condition on the distribution of the observed order statistics of the bids. The result thus shows that when the econometrician has access to an instrument, identification is possible without relying on the strong functional form assumptions used in Roberts (2013) and Freyberger and Larsen (2017). However, for my identification result to hold, I require that the econometrician has access to at least three order statistics of the bids (contrast to the result of Roberts (2013) that only requires one order statistic of the bids, and to the result of Freyberger and Larsen (2017) that require only two order statistics).

In the context of SPA, I extend my identification results to settings with UH and an unobserved number of potential bidders. The latter setting is particularly relevant to online auctions where not all potential bidders place bids, and where substituting the number of actual bidders (those that place bids) for the number of potential bidder may lead to incorrect inference of the bidders' valuations when the two differ. In the latter setting (which is actually the one considered in Freyberger and Larsen (2017)), I show that identification is again possible under similar mild rank conditions on the distribution of the observed order statistics of the bids.

The main observation that I use throughout the paper to establish these new identification results is the fact that order statistics satisfy the "Markov property". Loosely speaking the Markov property of order statistics states that, conditional on any intermediate order statistics, any two non-consecutive order statistics (from *i.i.d* draws from a continuous distribution) are independent. This property allows me to conclude that nonconsecutive order statistics of the bids are independent once I condition on the UH and on any intermediate order statistic of the bids. This is particularly useful as it allows me to *represent* the distribution of the observed order statistics of the bids in a form that allows me to exploit the results and proof techniques from the mixture literature. This paper is the first one in the auction literature (to my knowledge) that uses the Markov property of order statistics for identification. The paper contributes to the mixture literature as well, by establishing identification for a mixture model where the distribution of the observed variables are correlated within the components of the mixture.

1.1 Literature review

There is an extensive literature that studies auction models in the presence of unobserved heterogeneity. These papers include, among others, Krasnokutskaya (2011), Li, Perrigne, and Vuong (2000), Athey and Haile (2002), Hu, McAdams, and Shum (2013), Roberts (2013), Armstrong (2013), Aradillas-Lopez, Gandhi, and Quint (2013) and Freyberger and Larsen (2017). Models with UH provide a good alternative way to model auction data where the observed bids are correlated even after controlling for observable covariates that plausibly affect bidders' valuations. Athey and Haile (2002) (for SPA), Li, Perrigne, and Vuong (2000) (for FPA) and Krasnokutskaya (2011) (for FPA) are among some of the earlier papers to study the nonparametric identification of auction models in the presence of UH. In all these papers, the UH (which is assumed to be a continuous random variable) is assumed to either have a multiplicative or an additive effect on bidders' valuations, and the identification arguments rely on results from the measurement error literature, and require (in a setting with symmetry) observation of (at least) the same two bidders across an i.i.d sample of auctions. Hu, McAdams, and Shum (2013) relaxes the assumptions in Krasnokutskaya (2011) by allowing the distribution of valuations to depend on the unobserved heterogeneity in an unrestricted way. All that is required is for the conditional distribution of valuations to be strictly monotone (in a first order stochastic sense) in the UH. However, to establish identification, the econometrician needs to observe bids from (at least) the same three bidders across an i.i.d sample of auctions.

The identification arguments in Athey and Haile (2002), Krasnokutskaya (2011) and Hu, McAdams, and Shum (2013) are not applicable to settings with incomplete bid data, where the econometrician only observes a subset of the order statistics of the bids across an i.i.d sample of auctions. That is the scenario with which the present paper is concerned. Moreover, Athey and Haile (2002) show that for FPA and SPA with private values where bidders' valuations are allowed to be correlated in an unrestricted way, the joint distribution of bidders' valuations is not identified from any strict subset of the order statistics of the bids (even if ones assumes that the joint distribution of valuation is symmetric). I show below that if the correlation of bidders' valuations arises through a model of conditionally independent private values (CIPV) with finite UH, then the joint distribution of bidders' valuations and the UH can be identified from some strict subset of order statistics of the bids. This structure is reasonable in the setting described in the introduction, where bidders have independent private values conditional on some covariates that are commonly observed by the bidders but unobserved by the econometrician. Like Hu, McAdams, and Shum (2013), I do not make any functional-form assumption on how the UH determines valuations.

Other papers in the auction literature that study auction models with UH and incomplete bid data include (among others) Roberts (2013), Aradillas-Lopez, Gandhi, and Quint (2013), Armstrong (2013) and Freyberger and Larsen (2017). In a private value setting with continuous UH, Roberts (2013) establishes (among other things) the identification of the conditional distribution of valuations given the UH from the joint distribution of any order statistics of the bids and the reserve price. Although Roberts (2013) allows the UH to affect the distribution of valuations in an unrestricted way, the reserve price is required to be a strictly monotonic function of the UH (and some other observable covariates) for his identification argument. In the present paper, when an instrument is used to obtain identification, I allow it to be related to the UH in a much weaker way (which will be made clear below); the instrument is allowed (in particular) to be a non-degenerate random variable after conditioning on the UH (and all observable covariates), which makes it applicable to settings where the sellers may have some private information about the object being auctioned that they use to set reserve prices (see the discussion following Remark 3.4). Freyberger and Larsen (2017) establish the identification of the joint distribution of bidders' valuations and the UH, in a setting with incomplete bid data where the number of potential bidders is also unobserved. Since their results rely on classical measurement error arguments, they assume that the UH affect the distribution of bidders' valuations and the instrument (the secret reserve price) in an additively separable way. The results of the present paper do not rely on such strong functional-form restrictions. Armstrong (2013) and Aradillas-Lopez, Gandhi, and Quint (2013) establish partial identification results (for FP and ascending auctions respectively) in a CIPV setting with UH (which they both allow to be of unrestricted dimension) from the distribution of the highest bid. Their partial identification results concern, however, lower dimensional parameters like seller's profit and bidder's surplus, whereas the present paper is concerned with a much deeper structural parameter: the distribution of bidders' valuations.

The present paper is also related to the literature on the nonparametric identification of mixtures models. Related papers from that literature include (among others) Hall and Zhou (2003), Elizabeth, Matias, and Rhodes (2009), Bonhomme, Jochmans, and Robin (2014), Bonhomme, Jochmans, and Robin (2016), Kasahara and Shimotsu (2014) and Kasahara and Shimotsu (2009). The setup in these papers is one where the econometrician observes an i.i.d sample of d covariates that are conditionally independent given the realizations of some finite UH, and the goal is to study the identification of the conditional distribution of the covariates with respect to the UH as well as the marginal distribution of the UH. It is shown (in the papers cited above) that the conditional distribution of the covariates with respect to the UH as well as the marginal distribution of the UH are point identified (under mild "rank conditions") if d is greater than or equal to 3. In the setting of the present paper (CIPV with finite UH) the distribution of the observed bids is a finite mixture. The results from the mixture literature are, however, not directly applicable. This is due to the fact that order statistics of bids are necessarily correlated even after conditioning on the UH. I overcome this obstacle by exploiting the Markov property of order statistics.

The rest of the paper is organized as follows. Section 2 introduces the model and discusses some of the assumptions that are needed for the identification results. Section 3 states all the identification results. Section 4 introduces estimators based on the constructive identification argument of section 3, and establishes their asymptotic properties. Section 5 contains a Monte Carlo exercise. All the proofs omitted from the main text, as well as a subsection on the mathematical notation used throughout the paper, are provided in the appendix.

2 Model

I now describe the model that is used throughout most of the paper. In each auction t, a single and indivisible object is auctioned to $I_t \geq 2$ risk neutral bidders indexed by i. At each auction t, bidders learn their private values $V_{it} \sim F_{V_{it}}$ which can depend on a set of auction level covariates (X_t, U_t), but not on the random variable I_t. Here $X_t \in \mathbb{R}^d$ (X_t is allowed to have continuous or discrete components) and U_t is discrete with finite support $\mathcal{U}$[2]. It is assumed that both X_t and U_t are observed by the bidders, but that the econometrician only observes X_t. The covariates X_t and U_t can respectively be thought of as observed and unobserved characteristics of the auctioned object that affect bidders' valuations. I will refer to the variable U_t as the auction level unobserved heterogeneity (UH). Before stating the assumptions of the model, I provide an illustrative example.

Example 2.1. Consider an auction for used cars where all the auction participants are allowed to inspect the cars prior to placing their bids. There are many characteristics of the cars being auctioned (and of the auction itself) that may reasonably be assumed to be commonly observed by the bidders and to affect their valuations. These car attributes include among others the car's make,

[2]Some of the identification results in this paper (theorems 4.4 and 4.2) can be extended to a setting with continuous UH, by replacing the full rank conditions in assumptions 3.3 and 3.2 with completeness conditions on the corresponding operators (see Hu and Schennach (2008)). It is not clear however whether results similar to those in sections 3.1 and 3.2 can be established in the setting with continuous UH.

model, mileage, and transmission type. The subset of these characteristics that the econometrician observes will constitute the vector X_t, and U_t will denote the remaining subset of characteristics that the econometrician does not observe (which I assume are all discrete variables). It can be the case for instance that the econometrician observes all the relevant car attributes to the exception of the condition of the car's body, which is modelled as a categorical variable with three values: $U_t \in \mathcal{U} := \{good, fair, bad\}$.

The main informational assumption that I maintain throughout the paper is that the bidders' valuations are independent and symmetric conditional on (X_t, U_t). For notational simplicity, I will omit X_t onwards; all arguments and results can thus be understood as being made conditional on the observed covariates X_t. Formally, we have:

Assumption 2.2. (Conditional IPV) The joint distribution of bidders' valuations F_V in an auction with I_0 bidders satisfies

$$F_V(v_1, \cdots, v_{I_0}) = \sum_{u \in \mathcal{U}} P(U = u) \prod_{i=1}^{I_0} F_{V|U}(v_i | U = u), \tag{2.1}$$

where $F_{V|U}$ denotes the common marginal distribution of private values given U, which I assume to have a closed interval (that may vary with the value of $U = u$) $[c_u, d_u]$ for support[3], with $d_u > c_u$, and a continuous density $f_{V|U=u}$ which is strictly positive at every point in (c_u, d_u).

Assumption 2.3. (Exogenous entry) $I_t \perp (V_{it}, U_t)$. [4]

I will assume throughout that there is no reserve price (or a non-binding one) and that the observed bids are the equilibrium bids of the auction format under consideration, and satisfy:

$$B_{it} = \beta(v_{it}, U_t, I_t), \tag{2.2}$$

for some strictly increasing and continuously differentiable (in its first argument) function β. In the case of SPA, I will assume that players play their weakly-dominant strategy that consists in bidding their valuations ($\beta(v_{it}, u_t, I_t) = v_{it}$). Note that for second price auctions, the model described above is observationally equivalent to a model where bidders do not observe the variable U (see Li, Perrigne, and Vuong (2000)); In both models the bids are equal to bidders' valuations, and whether or not the variable U is observed by the bidders is irrelevant for their bids. Therefore, in the case of a SPA, our identification results presented below will extend to settings where the variable U

[3]The lower bounds c_u are assumed to be non-negative, and I allow the upper and lower bounds to vary with different values of the unobserved heterogeneity.

[4] It will be clear from the proof of the identification results presented below that this assumption is not necessary for identification, and all the arguments in the proof of identification will still hold if they are made conditional on auctions of a fixed size I_0. The main importance of this assumption is that it will allow us to pool auctions of different sizes for estimation.

is also unobserved by the bidders[5]. The direct extension of our identification results to the setting where the variable U is also unobserved by the bidders will, however, not be possible for other auction formats (FPA for instance) where whether or not U is observed has behavioral implication for the bidders. When I consider FPA, I will assume that for each value of U, the bidders play the unique symmetric, increasing and differentiable Bayesian Nash equilibrium strategy, which exists under the assumptions made on $F_{V|U}$ in 2.1 (see Athey and Haile (2002)).

As mentioned in the introduction, I will consider a scenario where only a subset of the order statistics of the bids is observed by the econometrician. With that end in mind, I will denote by $B_t^{(i)}$ ($i = 1, \cdots, I_t$) the i^{th} largest order statistic in auction t (which has I_t bidders); we have for instance $B^{(1)} = \max\{B_1, B_2, \cdots, B_{I_t}\}$ and $B^{(I_t)} = \min\{B_1, B_2, \cdots, B_{I_t}\}$. Let $i_1 < i_2 < \cdots < i_r$ denote the indices of the observed order statistics of the bids, where it is assumed that $r < I_t$ for each auction t.

Example 2.4. Ascending auctions constitute an example of an auction format where incomplete bid data arise naturally. In the stylized model of ascending auctions, the push-button auctions, prices rise continually (from a very small value) and each of the I bidders chooses when to (irreversibly) exit the auction by releasing a pressed button. The auction ends when the second to last bidder exits, and the last remaining bidder is awarded the object at a price that is equal to the exit price of her last competitor. A dominant strategy equilibrium for the bidders (in the private value environment) in this setting is to exit the auction when the selling price reaches their reservation value (see Athey and Haile (2002)). Therefore, by design, we can never observe the bid (which are the exit prices in this case) of the bidder with the highest valuation, and we observe at most the lowest $I - 1$ order statistics of the bids.

Example 2.5. Another reason why the econometrician may only observe an incomplete set of order statistics of the bids, can simply be due to the fact that only a subset of the order statistics of the bids are recorded in the data. It is for instance not uncommon in large FP sealed bid auctions, where all the bids are observed by the auctioneer at the time of the auction, to only have records of the top two (or more) bids in the data.

Since the model described above is observationally equivalent to one where a different labelling is used to denote the different values of the UH, I normalize the support of U to $\mathcal{U} = \{1, \cdots, N\}$, where N denotes the cardinality of $\mathcal{U}$ which I assume to be possibly unknown to the econometrician. By assumption 2.1, the distribution of the observed bids satisfies (see 2.3)

$$F_{B^{(i_1)}, \cdots, B^{(i_r)}|I}(b_{i_1}, \cdots, b_{i_r}|I) = \sum_{n=1}^{N} P(U = n) F_{B^{(i_1)}, \cdots, B^{(i_r)}|U,I}(b_{i_1}, \cdots, b_{i_r}|U = n, I). \tag{2.3}$$

[5]When the variable U is unobserved by the bidders, the above model is an affiliated private value model, with a particular affiliation structure given by Assumption 2.2 (see Milgrom and Weber (1982) or Li, Perrigne, and Vuong (2002))

The left-hand side (LHS) of 2.3 is identified from observation of an i.i.d sample of the corresponding order statistics of the bids (and I_t), and the goal is to identify all the terms that appear on the right-hand side (RHS) of 2.3 [6]. Since the distribution of any order statistics of the bids (conditional on a fixed value of I) is sufficient to identify the distribution of bidders' valuations in an auction model (both first and second price) with symmetric independent private values and no unobserved heterogeneity (see Athey and Haile (2002)), the identification of $F_{B^{(i_1)},\cdots,B^{(i_r)}|U,I}$ will imply (using results in Athey and Haile (2002)) that $F_{V|U}$ (see 2.1) is identified. Indeed, in the case of SPA, identification of distribution of an order statistics of the bids $F_{B^{(i_j)}|U,I}$ (for some $j = 1, \cdots, r$) imply the identification of the distribution of the corresponding order statistic of valuations (since the bids in this case correspond to players' valuations). Furthermore, the distribution of any order statistic of valuations (from independent draws) identifies the underlying distribution of valuation; this follows from the fact that the CDF of any order statistic of valuations is a known monotone transformation of the underlying CDF of valuations (see Lemma 7.2). For FPA, the marginal distribution of bids $F_{B|U,I}$ is identified from the distribution of any order statistic of the bids $F_{B^{(i_j)}|U,I}$ (for some $j = 1, \cdots, r$) (see Lemma 7.2), and the marginal distribution of valuations $F_{V|U}$ is identified from the marginal distribution of bids $F_{B|U,I}$ using the first order condition of player's maximization problem (see Guerre, Perrigne, and Vuong (2000)):

$$V = \zeta(B, F_{B|U,I}, I) = B + \frac{1}{I-1}\frac{F_{B|U,I}(B)}{f_{B|U,I}(B)} \tag{2.4}$$

where $f_{B|U,I}$ denotes the density of B conditional on U. In equation 2.4, the random variable B that appears on the RHS has a distribution given by $F_{B|U,I}$ and the variable V that appears on the LHS has distribution given by $F_{V|U}$. Hence if $F_{B|U,I}$ is identified for some fixed value of the random variable I, then the function $\zeta(\cdot, F_{B|U,I}, I)$ in 2.4 is identified, and the distribution $F_{V|U}$ is identified as the distribution of variable $V = \zeta(B, F_{B|U,I}, I)$, where $B \sim F_{B|U,I}$. Therefore, for both FP and SPA, our identification problem reduces to studying under what conditions the terms on the RHS of 2.3 are identified from the distribution on its LHS.

The structure of 2.3 is very much different from that of mixture models considered in the literature on nonparametric identification of mixtures (see Elizabeth, Matias, and Rhodes (2009), Bonhomme, Jochmans, and Robin (2014), Bonhomme, Jochmans, and Robin (2016) and Kasahara and Shimotsu (2014)). The mixture model considered in these papers is one where the observed variables, say $X_1, X_2, \cdots, X_p$ are conditionally independent given the realizations of some finite unobserved heterogeneity Θ (say with support $\{1, \cdots, N\}$), which yields the following expres-

[6] Here and in what follows, identification is understood to mean up to label swapping of the various components of the mixture.

sion (similar to 2.1):

$$P(X_1 \le x_1, X_2 \le x_2, \cdots, X_p \le x_p) = \sum_{n=1}^{N} P(\Theta = n) \prod_{j=1}^{p} P(X_j \le x_j | \Theta = n). \tag{2.5}$$

The multiplicative structure on the RHS of 2.5 is however not possible in our setting where the observed variables are order statistics, as they are necessarily correlated (after conditioning on the U_t and I_t). Indeed, conditional on the largest order statistic (out of I_0 independent draws from some distribution) taking a specific value, all other order statistics are constrained to take smaller values. In the next section, I will show how the Markov property of order statistics (see Lemma 3.1) can be exploited to rewrite equation 2.3 in a form that is somewhat similar to 2.5.

3 Identification

Before proceeding to the identification argument, I first state a lemma that recalls the Markov property of order statistics. The original statement and proof of this result can be found in Kolmogorov (1933) (see also Aron and Navada (2003) for a more recent treatment).

Lemma 3.1. (Markov Property) *Let W_i ($i = 1, \ldots, I_0$) represent i.i.d draws from some continuous distribution F, then the corresponding order statistics $W^{(i)}$ ($i = 1, \cdots, I_0$) satisfy*

$$W^{(k)} | \left(W^{(k+1)}, \cdots, W^{(I_0)} \right) \sim W^{(k)} | W^{(k+1)}$$

for any $1 \le k \le I_0 - 2$.

Lemma 3.1 implies for instance that for a fixed number of bidders I_0, the observed bids $B^{(i_1)}$ and $B^{(i_3)}$ are conditionally independent given $B^{(i_2)}$ and U (recall that $i_1 < i_2 < i_3$). I will use such arguments below to rewrite equation 2.3 in a form that allows me to exploit techniques from the mixture literature and to identify all variables on the RHS of 2.3. I now proceed to introduce some additional assumptions that I will require for identification.

I establish below identification of my model under two distinct scenarios that are determined by the structure of the data that is available to the econometrician. In the first scenario, I will assume that the econometrician can observe an instrumental-like variable that is conditionally independent of the distribution of valuations given the UH. In this case I establish identification of the RHS of 2.3 if the econometrician observes in addition three order statistics of the bids across an i.i.d sample of auctions, if these order statistics of bids satisfy a full rank condition, and if the instrument satisfies exclusion-restriction and relevance-like conditions. In the second scenario, I will assume that the econometrician can observe at least five order statistics of the bids across an i.i.d sample of auctions. This arises, for example, if the data contains all the exit prices in a push button English auction where at least six bidders participate, or it can also arise in large auctions where only few of the top bids (at least five) are recorded in the data (see Example 2.4 and Example 2.5). I show that with this data structure, a full rank condition on the conditional distribution of the order statistics of the bids suffices to identify all terms on the RHS of 2.3. Formally, the assumptions are:

Assumption 3.2 (**F**or three order statistics of the bids plus an instrument)**.** For $i_1 < i_2 < i_3 \leq I$, and for some $a \in \mathbb{R}_+$, the following conditions on the instrument Z and on the players' bids and valuations are satisfied:

$$(V_1, \cdots, V_I) \perp Z|U, \tag{3.1}$$

and the cumulative distribution functions within each of the following sets are linearly independent:

$$\{F_{B^{(i_1)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_1)}|B^{(i_2)}=a,U=N,I}\}, \{F_{B^{(i_3)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,U=N,I}\}, \tag{3.2}$$

and

$$\{F_{Z|U=1}, \cdots, F_{Z|U=N}\}. \tag{3.3}$$

Assumption 3.3 (**F**or five order statistics of the bids)**.** For $i_1 < i_2 < \cdots < i_5 \leq I$, and for some $a, b \in \mathbb{R}_+$ with $a > b$, the cumulative distribution functions within each of the following sets are linearly independent

$$\{F_{B^{(i_1)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_1)}|B^{(i_2)}=a,U=N,I}\},$$

$$\{F_{B^{(i_3)}|B^{(i_2)}=a,B^{(i_4)}=b,U=1,I}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,B^{(i_4)}=b,U=N,I}\}$$

and

$$\{F_{B^{(i_5)}|B^{(i_4)}=b,U=1,I}, \cdots, F_{B^{(i_5)}|B^{(i_4)}=b,U=N,I}\}.$$

Remark 3.4. For Assumption 3.3 to hold, for either FPA or SPA, it is necessary that the interval $[b, a]$ is contained in the intersection of the support of the marginal distribution of bids given different values of U. For SPA, this amounts to requiring that $[b, a] \subset \cap_{n=1}^{N}(c_n, d_n)$, where (c_n, d_n) represents the interior of the support of the marginal distribution of valuations given $U = n$ (see 2.1 and 2.2). Hence 3.3 precludes settings in which the support of the marginal distribution of bids given different values of the UH have intersection with empty interior. A similar remark applies to 3.2.

Remark 3.5. Assumption 3.2 allows for the instrument Z to be discrete. However, for condition 3.3 to hold, it is necessary that Z has at least N support points.

I refer to the variable Z in Assumption 3.2 as an "instrumental-like" variable, because condition 3.1 can be interpreted as an exclusion-restriction, as it allows for the variable Z to affect the distribution of the observed bids only through the variable U, and condition 3.3 can be seen as a relevance condition, as it requires—in some sense—that Z be correlated with U. Note, however, that the variable U which plays the role of the endogenous variable in the analogy is not observed. One good example of an instrument Z in our setting is a secret reserve price[7]. Indeed, suppose that the secret reserve price is given by $Z = h(U, \eta)$, for some function h, and where η represents the seller's own private information which is conditionally (on UH) independent of bidders valuations: that is $\eta \perp (V_1, \cdots, V_I)|U$. Then 3.1 is clearly satisfied, and 3.3 can be expected to hold for some choices of functions h and of variable η. It can easily be shown that 3.3 holds if $Z = h(U, \eta) = U + \eta$ and if we assume in addition that $\eta \perp U$. Such a model for the reserve price is reasonable if the seller also observes U (in addition to some private signal η), and uses that information to set the reserve price (see Roberts (2013) and Freyberger and Larsen (2017)). Another example of an instrument suggested by Hu, McAdams, and Shum (2013) in the context of timber auction, where the U represents the quality of timber for sale, is the average amount of rainfall or the soil quality, which is arguably correlated with timber quality, but only affects bidders' valuations through timber quality. Another example of an instrument in the setting of Example 2.1 above, is the average yearly amount of salt used to melt ice on the roads in the locality (zip code) of provenance of the car, as the amount of snow used on the road is negatively correlated with the condition of the car's body and can reasonably be assumed to affect bidders' valuations of the car through the car's body.

Note that unlike Roberts (2013) and Freyberger and Larsen (2017) who use an instrument to establish identification in a setting with incomplete bid data and UH (in the setting of Freyberger and Larsen (2017) the number of bidders I is also unobserved by the econometrician), we put little restrictions on how the instrument and UH are related. In Freyberger and Larsen (2017), since they rely on classical deconvolution arguments for identification, the instrument is related to the UH in an additively separable way: $Z = U + \eta$, with η independent of all other variables in the model. The additive separability assumption greatly restricts the way in which the UH affect the distribution of the instrument, as it is only allowed to shift its mean[8]. In 3.3, the UH is allowed to shift the distribution of the instrument in a much more complex way. In Roberts (2013), however, the identification argument relies on a control function approach, and the instrument Z (which is the reserve price in his setting) is essentially modelled as some strictly monotonic function of U: That is $Z = h(U)$ for some unknown and strictly monotone h. Clearly, modelling the seller's reserve price as a deterministic function of U excludes settings where the seller also possesses some additional information that she uses to set the reserve prices, which may lead to the distribution of $Z|U$ being non-degenerate.

[7] A secret (as opposed to a public) reserve price is one that is not observed by the bidders when they place their bids.

[8] Their identification argument is also valid for the multiplicative separable specification $Z = U\eta$, where both U and η are positive. Indeed an application of the logarithm function transforms this specification into an additive separable one, and the identification results in that context apply. Note that the multiplicative separable specification only allows the UH to shift the mean of the logarithm of the instrument.

Remark 3.6. The full rank conditions that appear in Assumption 3.3 and Assumption 3.2 are similar to the full rank conditions that are needed for identification in the mixture literature and in the misclassification literature (see Elizabeth, Matias, and Rhodes (2009), Bonhomme, Jochmans, and Robin (2014), Bonhomme, Jochmans, and Robin (2016) and Kasahara and Shimotsu (2009), Hu, McAdams, and Shum (2013), Hu (2008), An, Hu, and Shum (2010)). These rank conditions require that the UH induces sufficiently heterogeneous variations on the conditional distribution of observed order statistics given U. As will be shown in the proof, the rank conditions are needed to guarantee the invertibility of certain operators. Moreover, I will show that 3.3 and 3.2 have testable implications; Indeed, they imply that certain identified matrices have finite rank. In the case of SPA, since bids are equal to valuations, assumptions 3.3 and 3.2 are assumptions on the underlying distribution of valuations; a primitive of the model. Hence these assumptions can be seen as "low-level" conditions in the case of SPA. In the case of FPA, however, these assumptions put restrictions on the distributions of observable bids (not a primitive of the model), and 3.3 and 3.2 are interpretable in this case as "high-level" conditions. Intuitively, for conditions 3.2 and 3.3 to hold, the re-normalization to subintervals of the conditional distribution of valuations has to sufficiently vary across different values of U.

Example 3.7. For both FP and SPA, Assumption 3.3 and condition 3.2 in Assumption 3.2 hold for many familiar parametric distributions. It can be shown for instance that it holds if the conditional distribution of valuations given the UH is log-normally distributed. That is $V|U \sim Z$ with $\log(Z) \sim N(U, 1)$ or $\log(Z) \sim N(0, U)$, and with any pair of positive numbers $0 < b < a$ and any five order statistics. The same conclusion also applies if the conditional distribution of valuation given the UH is exponential: That is $V|U \sim exp(U)$.

Example 3.8. In the case of FPA and SPA, 3.3 and 3.2 fail to hold for instance when $V|U = n$ is distributed uniformly on the interval $[0, n]$, for $n = 1, \cdots, N$. In the case of SPA, this failure is due to the fact that the conditional distribution of valuations conditional on valuations being less than a (and on $U = n$), for any $a \in (0, 1)$, is equal to the uniform distribution on $[0, a]$ (independently of n). Therefore, by Lemma 7.1 the elements of the set

$$\{F_{B^{(i_3)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,U=N,I}\}$$

appearing in condition 3.2 (for any $a \in (0, 1)$) are all equal, as well as those within the set

$$\{F_{B^{(i_5)}|B^{(i_4)}=b,U=1,I}, \cdots, F_{B^{(i_5)}|B^{(i_4)}=b,U=N,I}\}$$

appearing in 3.3 (for any $b \in (0, 1)$), and linear independence fails to hold. Therefore, Assumptions 3.2 and 3.3 fail to hold for this example (see Remark 3.4). In the case of FPA, the bidding strategy in auctions of type $U = n$ is given by $\beta(v) = (I/(I+1))v$ for $v \in [0, n]$, and the conditional distribution of bids conditional on bids being less than any a, for any $a \in (0, I/(I+1))$, is given by the uniform distribution on $[0, a]$ (independently of n). Therefore, by Lemma 7.1, as above, linear independence of the elements of the set

$$\{F_{B^{(i_3)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,U=N,I}\},$$

or

$$\{F_{B^{(i_5)}|B^{(i_4)}=b,U=1,I}, \cdots, F_{B^{(i_5)}|B^{(i_4)}=b,U=N,I}\},$$

fails to hold for all values of a and b in the interior of the common intersection of the supports (see Remark 3.4)) which is equal to $(0, I/(I+1))$.

I show in Section 3.2 below that in the case of SPA, failure of 3.3 or 3.2$-$as in Example 3.8$-$is in some sense "pathological", and both conditions hold generically. Loosely speaking, this means that given all the restrictions that we put on the conditional (on UH) distribution of valuations, an $N-$tuple of distributions $(F_{V|U=1}, \cdots, F_{V|U=N})$ taken at random from the appropriate space (which imposes all the other restrictions of the model) satisfies the desired conditions with "probability" one. The appropriate definitions and the exact statement of the result will be provided in Section 3.2. I now state the main identification results of this section.

Theorem 3.9. *Suppose that the econometrician observes an i.i.d sample* $\{B_t^{(i_1)}, B_t^{(i_2)}, B_t^{(i_3)}, Z_t, I_t\}_{t=1}^T$ *and that assumption 3.2 holds. Then the conditional distributions of players' bids given different realizations of the unobserved heterogeneity* U *and of the number of bidders* I, $F_{B|U,I}$, *as well as the marginal distribution of the unobserved heterogeneity are identified.*

Theorem 3.10. *Suppose that the econometrician observes an i.i.d sample* $\{B_t^{(i_1)}, \cdots, B_t^{(i_5)}, I_t\}_{t=1}^T$ *and that assumption 3.3 holds. Then the conditional distributions of players' bids given different realizations of the unobserved heterogeneity* U *and of the number of bidders* I, $F_{B|U,I}$, *as well as the marginal distribution of the unobserved heterogeneity are identified.*

Remark 3.11. Theorem 3.10 is the first positive identification result in a model with incomplete bid data and unobserved heterogeneity that does not rely on the availability of some additional auxiliary data such as an instrument (contrast to Roberts (2013) and Freyberger and Larsen (2017)). Also, the identification argument does not exploit variations in the number of bidders I, and the argument can be made conditional on a fixed value of I (contrast to Quint (2015)). Hence Theorem 3.10, in parts, answers an identification question raised in section 3.2 of Athey and Haile (2002), concerning the possibility of identification in a model of UH from an incomplete set of bids. The main new technical tool used to establish these identification results is the Markov property of order statistics (Lemma 3.1). The use of five order statistics in 3.10 to obtain identification mirrors the result of Hu and Shum (2012) where five periods of observation are needed to establish identification of the law of motion of a Markov process with some unobserved state variables, and the identification arguments are somewhat similar.

Remark 3.12. Note that by the exogenous entry assumption (2.3) and the fact that bids are equal to valuations in SPA, the identified conditional distributions of bids $F_{B|U,I}$ are independent of the number I of bidders and are equal to the conditional distributions of bidders' valuations given U: $F_{V|U}$. In the context of FPA however, although the exogenous entry Assumption implies that the conditional distribution of valuations given U and I is independent of I, the identified conditional distributions $F_{B|U,I}$ will be dependent on I, as the bidding strategies in FPA are functions of the level of competition I. In fact, $F_{B|U,I}$ will be increasing in I in the first order stochastic dominance sense, since players shade less when the competition is greater in FPA. This shows that both in FPA and SPA the exogenous entry assumption has some testable implications when the variable I takes at least 2 values with positive probability.

Remark 3.13. A closer look at the proof of Theorem 3.10 reveals that Assumption 3.3 is stronger than needed for identification and can be relaxed. Indeed, the conclusion of 3.10 still holds if condition 3.3 is replaced with the weaker assumption that only requires that the CDFs within any two of the three sets in 3.3 are linearly independent and those within the third set are distinct. However, the identification argument under this weaker form of Assumption 3.3 is more involved. Similarly, the result of Theorem 3.9 still holds if we only require that the CDFs within any two of the three sets in 3.2 and 3.3 are linearly independent, and that those within the third set are distinct (see proof of 3.20). In section 3.1, I introduce an easily interpretable condition on the conditional distributions of valuations (given U) in the context of FPA, and show that it implies the weaker form of Assumption 3.2 alluded to above, thus providing a low-level condition that is sufficient for identification in the setting of Theorem 3.9.

Heuristic. I provide now the heuristics for the main steps involved in the proof of 3.10. The argument used to establish 3.9 is similar. The following argument is done conditional on the random variable I (the number of bidders) taking a fixed value I_0 in its support, and, for notational simplicity, I will omit I from the conditioning set. The proof proceeds in three main steps. In the first step, I use multiple applications of Lemma 3.1 and the law of iterated expectations, to express the joint distribution of any five order statistics of the observed bids (in auctions with I_0 bidders[9])$-$ which I assume without loss of generality to be the first five order statistics $-$ as follows

$$F_{\left(B^{(1)},B^{(3)},B^{(5)}|B^{(2)},B^{(4)}\right)}(b_1,b_3,b_5|B^{(2)}=a,B^{(4)}=b) \tag{3.4}$$

$$=\sum_{n=1}^{N} P(U=n|B^{(2)}=a,B^{(4)}=b)F_{1n}(b_1)F_{3n}(b_3)F_{5n}(b_5) \tag{3.5}$$

where $F_{1n}(b_1) := F_{B^{(1)}|B^{(2)}}(b_1|B^{(2)}=a,U=n)$, $F_{3n}(b_3) := F_{B^{(3)}|B^{(2)},B^{(4)}}(b_3|B^{(2)}=a,B^{(4)}=b,U=n)$ and $F_{5n}(b_5) := F_{B^{(5)}|B^{(4)}}(b_5|B^{(4)}=b,U=n)$. Equation 3.4 is now in a form that is similar to 2.5, and note that by Remark 3.4 the probabilities $\{P(U=n|B^{(2)}=a,B^{(4)}=b)\}_{n=1}^{N}$ are all non-zero.

[9]Since we consider auctions with incomplete bid data, I_0 is necessarily greater than or equal to 6.

In the second step of the proof, I apply the arguments from the mixture literature to identify[10] all the terms on the RHS of 3.4. Assumption 3.3 can be seen as the natural analogue of the linear independence condition in Theorem 8 of Elizabeth, Matias, and Rhodes (2009). In the third step of the proof, I show how the objects of interest− the conditional distribution of bids given different values of the UH, $\{F_{B|U}(\cdot|U=n)\}_{n=1}^{N}$, and the marginal distribution of the UH−can be identified from the terms on the RHS of 3.4. In this last step, I repeatedly use Lemma 7.1 and Lemma 7.2 to identify different "portions" of the marginal distribution of bids given the UH from the conditional distribution of order statistic of bids that appear on the RHS of 3.4. The application of Lemma 7.1 implies, for instance, that F_{5n} is equal to the distribution of the largest order statistic out of $I_0 - 4$ independent draws from the distribution of bids conditional on $U = n$ and on bids being less than b. It then follows from Lemma 7.2 that $\phi^{-1}_{1:I_0-1} \circ F_{5n}$ identifies the distribution of bids conditional on $U = n$ and on bids being less than b:

$$\frac{F_{B|U}(\cdot|U=n)}{F_{B|U}(b|U=n)}$$

.

Once the conditional distributions of bids given different values of the UH are identified, recall that in the case of SPA, the identification of $F_{B|U}$ implies the identification of the conditional distribution of valuations $F_{V|U}$ (since bids are equal to valuations), and for FPA, the conditional distributions of valuation given the UH, $F_{V|U}$, are identified from $F_{B|U}$ through 2.4. This concludes the identification argument.

□

3.1 Low level condition for FPA

In the setting of FPA, I provide in this section a condition on the distribution of bidders' valuations that is sufficient for identification when the data available to the econometrician is as in Theorem 3.9. I replace the identifying assumption 3.2 − a condition on the observed bids (a high-level condition for FPA) − used to establish Theorem 3.9 by an assumption on the conditional distribution of bidders' valuations given U − a condition on a primitive of the model (a low-level condition). Moreover, I show that the new low level condition is sufficient for identification of the parameters in Theorem 3.9. I begin by recalling the definition of the reverse hazard rate order (see Shaked M. (2007)), which is a stochastic order that I use in the statement of Assumption 3.16 below[11].

[10] Recall that identification in this context is defined up to label swapping.

[11] Reverse hazard rate dominance is equivalent to the notion of *conditional stochastic dominance* in Maskin and Riley (2000) (see 1.B.43 in Shaked M. (2007)).

Definition 3.14 (Reverse hazard rate dominance)**.** Let F be a continuously differentiable CDF with density f. The reverse hazard rate (RHR) function of the distribution F is defined by

$$r_F(t) = \frac{f(t)}{F(t)}$$

for all values of t strictly greater than the lower bound of the support of F, and is equal to zero otherwise. Given two random variables $X \sim F$ and $Y \sim G$, we say that X dominates Y (alternatively F dominates G) in the RHR order, and write $X \succeq_{rh} Y$ (alternatively $F \succeq_{rh} G$), iff

$$r_F(t) \geq r_G(t) \tag{3.6}$$

for all $t \in \mathbb{R}$. Moreover, we say that X *strictly* dominates Y in the RHR order and write $X \succ_{rh} Y$ (or $F \succ_{rh} G$) if inequality 3.6 holds for all t and is strict for some values of t. Finally, we say that X *strongly* dominates Y in the RHR order, if inequality 3.6 holds for all t, and is strict for all t strictly greater than the lower bound of the support of G and strictly less than the upper bound of the support of F.

Remark 3.15. Note that if $F \succeq_{rh} G$, then it is necessarily true that the lower bound (resp. upper bound) of the support of G is less than or equal to the lower bound (resp. upper bound) of the support of F. Also, it is easy to show that (see 1.B.43 in Shaked M. (2007)) X dominates Y in the reverse hazard rate order if and only if $[X|X \leq t]$ first-order stochastically dominate $[Y|Y \leq t]$ for all t strictly greater than the maximum between the lower bound of the support of X and the lower bound of the support of Y [12]. Finally, reverse hazard rate order is implied by the likelihood ratio order and implies first order stochastic dominance [13] (see Theorem 1.B.32 and Theorem 1.C.1 in Shaked M. (2007), and Lemma 3.1 in Maskin and Riley (2000)).

I now state the main identifying assumption for Theorem 3.20 below.

Assumption 3.16. The instrument Z satisfies conditions 3.1 and 3.3. The conditional distributions of bidders' valuations, denoted by F_n ($F_n := F_{V|U=n}$), are supported on the finite intervals $[c_n, d_n]$ with a common lower bound: $c_n = c$ independently of n. The distributions F_n are increasing in n, for $n = 1, \cdots, N$, in the RHR order: That is, whenever $n > n'$, we have

$$F_n \succeq_{rh} F_{n'}, \tag{3.7}$$

and we assume in addition that the inequality between the RHR functions are strict for values of t near the lower bound of the support, i.e, there exists $\delta > 0$ such that

$$r_{F_n}(t) > r_{F_{n'}}(t) \tag{3.8}$$

for all $t \in (c, \delta)$, and for all $n > n'$.

[12] In an asymmetric auction setting with two types of bidders, Maskin and Riley (2000) refer to a notion similar to the latter property as *conditional stochastic dominance*, and Athey, Levin, and Seira (2011) refer to 3.6 as the *hazard rate order*.

[13] We say that F dominates G in the likelihood ratio order, and write $F \succeq_{lr} G$, if the ratio of their densities $\frac{f(t)}{g(t)}$ is non-decreasing over the union of their supports.

Remark 3.17. By Remark 3.15 and Footnote 13, 3.7 and 3.8 hold for instance if the ratio of densities $f_n(t)/f_{n'}(t)$ is non-decreasing on (c, d_n) and strictly increasing on (c, δ) whenever $1 \leq n' < n \leq N$. Also, it is easy to show that 3.7 and 3.8 imply that $F_n(t)/F_{n'}(t)$ is non-decreasing in t (for $t > c$), and is strictly increasing on the interval $(0, \delta)$ (see p.37 in Shaked M. (2007)), where δ is as in Assumption 3.16.

The condition on the distribution of bidders valuations in Assumption 3.16 is inspired by Proposition 3 in Hu, McAdams, and Shum (2013), where monotonicity of the conditional distribution of valuations in the first order stochastic dominance (FOSD) sense is used to establish a full rank condition on the conditional distributions of observed bids in FPA. In the setting of this paper, however, monotonicity of $F_{V|U}$ in U with respect to FOSD does not imply condition 3.2 on the conditional distribution of the observed order statistics of the bids, as shown by Example 3.8. Therefore, Monotonicity with respect to a stronger stochastic order is necessary to guarantee the full rank conditions 3.2 on the distribution of the observed bids. Note that when the conditional distributions of valuations are as in Example 3.8, condition 3.7 is satisfied, whereas 3.8 is not. Indeed, in the setting of Example 3.8, the RHR function of F_n is given by

$$r_n(t) = 1/t$$

for $t \in (0, n]$, and $r_n(t) = 0$ otherwise. I show in Proposition 3.18 below, that condition 3.7 coupled with 3.8 imply that bidders in auctions corresponding to larger values of the unobserved heterogeneity consistently bid more aggressively (closer to their valuation) than those in auctions with smaller values of U, and using the first order condition of bidders' maximization problem, I show that this implies that the distribution of bids in auctions that correspond to larger values of the unobserved heterogeneity *strongly* dominates in the RHR order the distribution of bids corresponding to lower values of the UH.

Proposition 3.18. *Suppose that Assumption 3.16 is satisfied, and let $G_{n,I}$ denote the marginal distribution of players' bids in auctions of type $U = n$ with I participants ($G_{n,I} := F_{B|U=n,I}$), where $I \geq 2$. Then $G_{n,I}$ strongly dominates $G_{n',I}$ in the reverse hazard rate order whenever $1 \leq n' < n \leq N$, and the upper bound of the support of $G_{n,I}$, denoted $\tilde{b}_{n,I}$, is strictly increasing in n.*

Proof. Fix $I \geq 2$ auction participants, and consider the different first-price auctions that correspond to different values of the variable U. When $U = n$, the (unique) symmetric, differentiable and strictly increasing Bayesian Nash equilibrium strategy of the corresponding FPA, denoted by $\beta_{n,I}$, is given by (see Riley and Samuelson (1981) or Guerre, Perrigne, and Vuong (2000))

$$\beta_{n,I}(v) = v - \int_c^v \left(\frac{F_n(u)}{F_n(v)} \right)^{I-1} du \tag{3.9}$$

for $v \in [c, d_n]$, where F_n, c and d_n are as in Assumption 3.16. Let $1 \leq n' < n \leq N$. By Remark 3.17, for all $v > c$, the distribution of bidders' valuations conditional on valuations being less than

v in auctions of type $U = n$, strictly first order stochastically dominates the distribution of bidders' valuations conditional on valuations being less than v in auctions of type $U = n'$: That is

$$\frac{F_n(u)}{F_n(v)} \le \frac{F_{n'}(u)}{F_{n'}(v)} \tag{3.10}$$

for any $c \le u \le v$, and note that the inequality is strict whenever $c < u < \min\{\delta, v\}$. Combining 3.9 and 3.10 yields

$$\beta_{n,I}(v) > \beta_{n',I}(v) \tag{3.11}$$

for all $v > c$. The latter is easily shown to imply

$$\tilde{b}_{n,I} > \tilde{b}_{n',I} \tag{3.12}$$

whenever $1 \le n' < n \le N$. Let $g_{n,I}$ denote the density of the marginal distribution of players' bids in auctions of type $U = n$. Making the change of variable $v = \beta_{n,I}^{-1}(b)$ into the first-order condition of bidders' optimization problem (see Proposition 6 in Laffont and Vuong (1996) for details) we get

$$r_{n,I}(b) = \frac{1}{(I-1)(\beta_{n,I}^{-1}(b) - b)} \tag{3.13}$$

where $r_{n,I}(b) = g_{n,I}(b)/G_{n,I}(b)$ (for $b > c$) denotes the RHR function of the distribution $G_{n,I}$. Combining 3.11 and 3.13 yields

$$r_{n,I}(b) > r_{n',I}(b) \tag{3.14}$$

for all $b \in (c, \tilde{b}_{n',I}]$, and $n > n'$. Moreover, since $r_{n',I}(b) = 0$ for all $b \in (\tilde{b}_{n',I}, \tilde{b}_{n,I}]$, we conclude that inequality 3.14 holds for all $b \in (c, \tilde{b}_{n,I})$ and $G_{n,I}$ *strongly* dominates $G_{n',I}$ in the RHR order.

□

I now establish a corollary of the preceding proposition that shows that a weaker version of condition 3.2 is satisfied when Assumption 3.16 holds. Note that the distributions in the set 3.16 below are only shown to be distinct (contrast to the condition on the same set of distributions in 3.2).

Corollary 3.19. *Fix $1 \le i_1 < i_2 < i_3 \le I$, and suppose that Assumption 3.16 is satisfied. Then for any $a \in (c, \tilde{b}_{1,I})$ — where $\tilde{b}_{1,I}$ denotes the upper bound of the support of bids when there are I auction participants and $U = 1$ (see Proposition 3.18) — the elements of the set*

$$\{F_{B^{(i_1)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_1)}|B^{(i_2)}=a,U=N,I}\} \tag{3.15}$$

are linearly independent, and the elements of the set

$$\{F_{B^{(i_3)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,U=N,I}\} \tag{3.16}$$

are distinct.

Proof. I first establish the claim concerning 3.15. Note that since densities of valuations are assumed to be strictly positive on the interior of their support (recall 2.1) and that bidding strategies are continuously differentiable and strictly increasing on the interior of the support of bidders' valuations, the support of a player's bid in an auction of type $U = n$ with I participants is given by $[c, \tilde{b}_{n,I}]$. Therefore, by Proposition 3.18, for any $a \in (c, \tilde{b}_{1,I})$, the support of a player's bid in an auction of type $U = n$ with I participants and conditional on the bid being larger than a, $[B|B > a, U = n, I]$, is equal to $[a, \tilde{b}_{n,I}]$. By Lemma 7.1, the distribution of $[B^{(i_1)}|B^{(i_2)} = a, U = n, I]$ is the same as the distribution of the i_1^{th} order statistic out of $i_2 - 1$ draws from the distribution of $[B|B > a, U = n, I]$. The latter observation combined with Lemma 7.2 imply that the support of $[B^{(i_1)}|B^{(i_2)} = a, U = n, I]$ is given by the interval $[a, \tilde{b}_{n,I}]$. Since $\tilde{b}_{n,I}$ is strictly increasing by Proposition 3.18, it easily follows that the elements of $\{F_{B^{(i_1)}|B^{(i_2)}=a,U=1,I}, \cdots, F_{B^{(i_1)}|B^{(i_2)}=a,U=N,I}\}$ are linearly independent.

I now establish the claim concerning 3.16. By Proposition 3.18 the marginal distribution of bids in auctions of type $U = n$ with I participants, $G_{n,I}$, is increasing in n in the *strong* RHR order. Therefore, for any $a \in (c, \tilde{b}_{1,I})$, the distribution of a player's bid in an auction of type $U = n$ with I participants and conditional on the bid being less than or equal to a, $[B|B \leq a, U = n, I]$, is *strictly* increasing in n in the FOSD sense [14]. Indeed, $G_{n,I}$ increasing in n in the *strong* RHR order implies that $G_{n,I}(t)/G_{n',I}(t)$ is strictly increasing in t on $(c, \bar{b}_{n,I})$, whenever $1 \leq n' < n \leq N$. By Lemma 7.1 the distribution of $[B^{(i_3)}|B^{(i_2)} = a, U = 1, I]$ is the same as the distribution of the $(i_3 - i_2)^{th}$ order statistic out of $I - i_2$ draws from the distribution of $[B|B \leq a, U = n, I]$. Combining the latter observation with Lemma 7.2, and the fact that all the functions $\phi_{i:I}$ in 7.2 are strictly increasing, we conclude that $[B^{(i_3)}|B^{(i_2)} = a, U = n, I]$ *strictly* first order stochastically dominate $[B^{(i_3)}|B^{(i_2)} = a, U = n', I]$ whenever $1 \leq n' < n \leq N$, and the elements of the set 3.16 are distinct. □

I now state the main theorem of this section, the proof is given in the Appendix.

Theorem 3.20. *Suppose that the econometrician observes an i.i.d sample* $\{B_t^{(i_1)}, B_t^{(i_2)}, B_t^{(i_3)}, Z_t, I_t\}_{t=1}^T$ *from FPA and that assumption 3.16 holds. Then the conditional distributions of players' bids given different realizations of the unobserved heterogeneity U and of the number of bidders I,* $F_{B|U,I}$*, as well as the marginal distribution of the unobserved heterogeneity are identified.*

Remark 3.21. Note that Assumption 3.16 puts a natural order on the components of the mixture and makes it possible to unambiguously identify each mixture component. Indeed, Proposition 3.18 implies that the mean of the distributions $F_{B|U,I}$ is strictly increasing in U[15]. Hence the identification result in Theorem 3.20 holds in the "classical" sense (not up to a permutation of the mixture

[14] I say that a distribution F *strictly* dominates a distribution G in the FOSD sense if $F(t) \leq G(t)$ for all $t \in \mathbb{R}$, with the latter inequality being strict for some values of t.

[15] Recall that strong RHR dominance implies strict first-order stochastic dominance.

components as in Theorem 3.9 and Theorem 3.10).

3.2 Genericity of identification for SPA

In this section, in the setting of SPA, I show that given all the (other) maintained assumptions on our model, the set of underlying distribution of valuations for which the distribution of observed bids satisfy condition 3.2 and 3.3 is "large" or "generic". This, in some sense, provides some justification for stating that Assumptions 3.2 and 3.3 are mild in the context of SPA.

For finite dimensional spaces, a property is said to be generic if the set of parameter values for which it fails to hold is a set of Lebesgue measure zero (a Lebesgue null set). This definition, however, does not readily extend to infinite dimensional Banach spaces, as there is no natural analogue of the Lebesgue measure on such spaces[16] (see Hunt, Sauer, and Yorke (1992)). There are two main notions of genericity in infinite dimensional spaces: the topological notion and the measure theoretic notion (see Anderson and Zame (2001)). The results in this paper will be stated in terms of the measure theoretic notion of genericity, as it is the natural extension (to infinite dimensional spaces) of the finite dimensional notion of genericity alluded to above. The starting point for this notion of genericity is based on the observation that in $\mathbb{R}^d$, a Borel set A has Lebesgue measure zero if and only if there exists a compactly supported probability measure, μ, such that $\mu(A+x)=0$ for all $x \in \mathbb{R}^d$ (see Hunt, Sauer, and Yorke (1992)). The latter equivalent characterization of a Lebesgue null set has a natural extension to infinite dimensional spaces: We say that a Borel subset A of an infinite dimensional Banach space $\mathcal{X}$ is shy if there exists a compactly supported regular Borel probability measure μ on $\mathcal{X}$ such that $\mu(A+x)=0$ for all $x \in \mathcal{X}$ (see Anderson and Zame (2001) and Hunt, Sauer, and Yorke (1992))[17]. Hence, shy sets are the infinite dimensional analogue of Lebesgue null sets, and we say that a set A is prevalent if its complement is shy[18].

As argued by Anderson and Zame (2001), however, the latter definition is not satisfactory for many economic applications, as the parameter space under consideration is often a much smaller subset of the ambient vector space[19]. Anderson and Zame (2001) provide an extension of the concept of shyness and prevalence to convex subsets of vector spaces. The definition (which I simplify to the setting of this paper) is as follows: Let $\mathcal{X}$ be a Banach space, and let $\mathcal{C} \subset \mathcal{X}$ be a closed con-

[16] The important property of the Lebesgue measure being that it is a non-zero translation invariant Borel measure which assigns finite mass to open balls.

[17] The definitions and results in Hunt, Sauer, and Yorke (1992) and Anderson and Zame (2001) are stated in terms of *completely metrizable topological vector spaces*. This level of generality, however, will not be needed for our results.

[18] See Hunt, Sauer, and Yorke (1992) for an extension of the concepts of shyness and prevalence to sets that are not necessarily Borel measurable.

[19] Consider for instance the question of how generic the property of being invertible is for symmetric 2×2 covariance matrices. The correct parameter space with respect to which genericity of invertibility should be established in this example is the set of symmetric 2×2 matrices, a set that has Lebesgue measure zero in the space of all 2×2 matrices (it is a set of dimension 3 in a space of dimension 4).

vex subset of $\mathcal{X}$. Let $c \in \mathcal{C}$. A set $E \subset \mathcal{C}$ which is Borel measurable is said to be shy in $\mathcal{C}$ at c if for each $r > 0$ and $\delta \in (0,1)$ there is a Borel regular probability measure with compact support such that $supp(\mu) \subset [\delta(\mathcal{C} - c) + c] \cap B(c,r)$,[20] and $\mu(E + x) = 0$ for all $x \in \mathcal{X}$. By definition, E is shy in $\mathcal{C}$ if it is shy at c for all $c \in \mathcal{C}$. A subset $F \subset \mathcal{C}$ is prevalent if its complement is shy in $\mathcal{C}$. [21] Anderson and Zame (2001) provide a series of results that show that the (relative) notion of shyness given above satisfies all the properties that we should expect of a measure theoretic definition of relative smallness. Papers in the economic literature that study this notion (or the topological notion) of genericity (in infinite dimensional spaces) include among others Heifetz and Neeman (2006), Chen and Xiong (2013)—who study genericity issues related to auction theory/mechanism design— Andrews (2017) (see also Connault (2016))— for genericity issues related to identification.

Let $\mathcal{X}$ be the vector space of all $N-$ tuple of continuous function on $[0,1]$ equipped with the norm $\|\mathbf{F}\| = \max_{1\leq i\leq N} \max_{x\in[0,1]} |F_i(x)|$, where $\mathbf{F} = (F_1, \cdots, F_N)$ is an element of $\mathcal{X}$. Let $\mathcal{C}$ be the subset of $\mathcal{X}$ that consists of all $N-$tuples of continuous CDFs. It can be easily shown that $\mathcal{X}$ is a Banach space, and that $\mathcal{C}$ is a closed convex subset of $\mathcal{X}$. Given a CDF ϕ and an element $\mathbf{F} \in \mathcal{C}$, let $\phi \circ \mathbf{F} \in \mathcal{C}$ be defined by $\phi \circ \mathbf{F} = (\phi(F_1), \cdots, \phi(F_N))$. Given a closed subinterval $S = [b,a]$ (with $b < a$) of $[0,1]$ and an element $\mathbf{F} = (F_1, \cdots, F_N) \in \mathcal{C}$, let *the renormalization of* $\mathbf{F}$ *on* S, denoted $\mathbf{F}_S$, be the element of $\mathcal{X}$ such that its i^{th} element is equal to

$$\min\{\max\{\frac{F_i(\cdot) - F_i(b)}{F_i(a) - F_i(b)}, 0\}, 1\}$$

if $F_i(a) - F_i(b) > 0$ and is equal to the zero function otherwise. That is, the renormalization of $\mathbf{F}$ on S is the $N-$tuple of functions which replaces each element of $\mathbf{F}$ by the corresponding conditional distribution on $[b,a]$ when the latter exists or by the zeroth function otherwise. Note that for $\mathbf{F} \in \mathcal{C}$, $\mathbf{F}_S \notin \mathcal{C}$ if and only if one of the elements of $\mathbf{F}$ does not vary on S, in which case the corresponding element in $\mathbf{F}_S$ is equal to zero. Also, when $S = [0,1]$, $\mathbf{F}_S = \mathbf{F}$. let $\phi_{i:n}$ be the distribution of the i^{th} order statistic out of n independent draws from a uniform distribution on $[0,1]$ (see 7.2), and let $\mathcal{A}$ be the collection of all such CDFs for all i and n:

$$\mathcal{A} := \{\phi_{i:n} | \text{for some } \ i \text{ and } n \ \in \mathbb{N} \text{ such that } \ 1 \leq i \leq n\}.$$

Given an element $\mathbf{F} \in \mathcal{C}$ we define the rank of $\mathbf{F}$, denoted $rank(\mathbf{F})$, to be the dimension of the vector space spanned by the elements of $\mathbf{F}$. Let $\mathbf{G^*} \in \mathcal{C}$ be such that $rank(\phi \circ \mathbf{G^*}) = N$ for all $\phi \in \mathcal{A}$ (I show in the appendix that such an element $\mathbf{G^*}$ exists). When the set S is a singleton, define the renormalization of $\mathbf{F}$ on S, for $\mathbf{F} \in \mathcal{C}$, by: $\mathbf{F}_S = \mathbf{G^*}$. I now state the main result of this section.

[20]B(c,r) denote the ball centered at c of radius r, and $\mathcal{C} - c := \{x - c | x \in \mathcal{C}\}$.

[21]When the ambient space $\mathcal{X}$ is finite dimensional, a subset E of a closed convex set $\mathcal{C}$ is shy with respect to the above definition if and only if it has measure zero with respect to the Lebesgue measure on the smallest hyperplane that contains $\mathcal{C}$ (see Anderson and Zame (2001))

Proposition 3.22. *For all $0 \leq b \leq a \leq 1$, the subset $E_{a,b}$ of $\mathcal{C}$ defined by*

$$E_{a,b} = \{\boldsymbol{F} \in \mathcal{C} |\, \min\{rank(\phi \circ \boldsymbol{F}_{[0,b]}), rank(\phi \circ \boldsymbol{F}_{[b,a]}), rank(\phi \circ \boldsymbol{F}_{[a,1]})\} < N, \text{ for some } \phi \in \mathcal{A}\}$$

is shy in $\mathcal{C}$.

Corollary 3.23. *For SPA, if we assume that the conditional distributions of bidders' valuations given different values of the UH have support contained on some compact set, say $[0, 1]$ for instance, then the set of all such N-tuples of distributions that satisfy condition 3.2 (or Assumption 3.3) is prevalent in $\mathcal{C}$.*

Interpretation Although the genericity results established in this section give some justification in stating that assumptions 3.2 and 3.3 are mild, one should be careful with their interpretation. The genericity claim is justified if any element of $\mathcal{C}$ is a plausible candidate for the set of conditional distributions of valuations (given different values of U) in our model. In that case, assuming that the identifying assumptions 3.2 and 3.3 hold will "almost always" be correct. However, for a specific application, it might be the case that more structure is imposed on the model. For instance, if the unobserved heterogeneity U represents some measure of quality of the auctioned object, it might be natural to assume that distributions $F_{V|U}$ that correspond to higher values of U first order stochastically dominate those that correspond to lower values of U (see Hu, McAdams, and Shum (2013)). In that case, the correct set $\mathcal{C}'$ with respect to which genericity should be establish is smaller than C; Indeed, $\mathcal{C}'$ consists of the elements of $\mathcal{C}$ that are totally ordered in the first order stochastic sense. Since the results of this section only deal with the genericity of the statement in $\mathcal{C}$, they are not applicable in the latter case, and it is possible for the set of distributions that satisfy 3.2 and 3.3 to now be non-generic relative to $\mathcal{C}'$.

Remark 3.24. By considering the setting where $a = 1$, $b = 0$, and by only considering the element $\phi_{1:1}(x) = x$ of $\mathcal{A}$, Proposition 3.22 shows that the set of linearly independent N-tuples of distributions is shy in $\mathcal{C}$. This is precisely the condition required by Theorem 8 of Elizabeth, Matias, and Rhodes (2009) to establish identification of mixtures of the type 2.5; it is shown In Elizabeth, Matias, and Rhodes (2009) that the mixture model 2.5 is identified if one observes at least three continuous covariates, and the set of distributions of covariate j across different values of Θ, $\{P(X_j \leq x_j|\Theta = n)\}_{n=1}^{N}$, is linearly independent (for at least three values of j). Proposition 3.22 strengthens the conclusion of Theorem 8 by showing that this linear independence assumption holds generically, and thus provides a counterpart to Theorem 4 in Elizabeth, Matias, and Rhodes (2009) which shows (under some mild conditions) that the mixture model 2.5 is generically identified if the observed covariates are discrete. I state the foregoing observation in the following corollary.

Corollary 3.25. *The mixture model 2.5 is generically identified whenever $p \geq 3$ and the covariates $\{X_i\}_{i=1}^{p}$ are continuously distributed.*

3.3 Identification for SPA when the number of bidders is unobserved

In this section, in the context of SPA, I show how the identification result in Theorem 3.9 can be extended to a setting where the number of potential bidders I is unobserved. As in Theorem 3.9, I will assume that the econometrician observes at least three order statistics of the bids $\{B^{(i_p)}\}_{p=1}^{3}$ ($1 \leq i_1 < i_2 < i_3 < I$) and an instrument Z. However, I will now assume that the level of competition I is unobserved. For instance, when $i_1 = 1$, $i_2 = 2$ and $i_3 = 3$, this will mean that the econometrician observes the top three bids from each auction in the sample (with at least three bidders), but does not know how many bidders participated in each auction− all that she can infer is that there are at least three bidders in each auction in the sample. The latter setting is in particular relevant to online auctions (see Song (2004)). I will assume that I has finite support $supp(I) = \mathcal{I}$.

Assumption 3.26 (**F**or three order statistics of the bids plus an instrument)**.** The instrument Z satisfies conditions 3.1 and 3.3. For $1 \leq i_1 < i_2 < i_3$, and for some $a \in \mathbb{R}_+$, the conditional distribution functions of the observed order statistics of the bids (given U) satisfy the following conditions:
the distributions within the set

$$\{F_{B^{(i_1)}|B^{(i_2)}=a,U=1}, \cdots, F_{B^{(i_1)}|B^{(i_2)}=a,U=N}\} \tag{3.17}$$

are linearly independent, and the distributions within the set

$$\{F_{B^{(i_3)}|B^{(i_2)}=a,U=1}, \cdots, F_{B^{(i_3)}|B^{(i_2)}=a,U=N}\} \tag{3.18}$$

are distinct.

Remark 3.27*.* Note that the distributions appearing in 3.17 and 3.18 do not involve conditioning on the unobserved I (compare to 3.2). However, Lemma 7.1 implies that the distributions in 3.17 are independent of I. We have for instance that

$$F_{B^{(i_1)}|B^{(i_2)}=a,U=n,I} = \phi_{i_1:i_2-1} \circ [F_{B|U=n}]_{[a,+\infty]} \tag{3.19}$$

where $\phi_{i_1:i_2-1}$ is as in Lemma 7.2 and $[F_{B|U=1}]_{[a,+\infty]}$ denotes the distribution of $F_{B|U=1}$ truncated at the left at a. Therefore the first set of distribution in 3.2 and the distributions in 3.17 are identical. Note however that the second set of distributions in 3.2 and the set of distributions in 3.18 are different. Indeed, by the law of iterated expectation and the exogenous entry assumption (2.3), we have

$$F_{B^{(i_3)}|B^{(i_2)}=a,U=1} = \sum_{j\in\mathcal{I} \text{ and } j\geq i_3} P(I=j|I\geq i_3) F_{B^{(i_3)}|B^{(i_2)}=a,U=1,I=j}$$

and Lemma 7.1 and Lemma 7.2 yield

$$F_{B^{(i_3)}|B^{(i_2)}=a,U=1} = \sum_{j\in\mathcal{I} \text{ and } j\geq i_3} P(I=j|I\geq i_3) \phi_{i_3-i_2:j-i_2} \circ [F_{B|U=1}]_{[0,a]} \tag{3.20}$$

where $[F_{B|U=1}]_{[0,a]}$ denotes the distribution $[F_{B|U=1}]_{[0,a]}$ truncated at the right at a. Therefore, the elements in 3.18 are mixtures of the corresponding elements in the second set of condition 3.2 (over different values of I), and condition 3.18 only requires that these mixtures are distinct for different values of the UH (see Remark 3.13).

Remark 3.28. Using an argument similar to the one used to establish Corollary 3.19, it is easy to show that assumption 3.26 is satisfied, for instance, for any three order statistics of the bids if: the intersection of the supports of the conditional distributions of valuations given different values of U has non-empty interior, $F_{V|U=n}$ *strongly* dominate in the RHR order $F_{V|U=n'}$ whenever $1 \leq n' < n \leq N$, and the upper bound of the support of $F_{V|U}$, d_n, is strictly increasing in U.

I now state the main result of this section. Its proof, which I provide in the appendix, involves three main steps: In the first step (similarly to the first step of the proof of Theorem 3.20) I show how Assumption 3.26 can be used to identify the conditional distribution of the instrument Z given U for all values of U, i.e, $\{F_{Z|U=n}\}_{n=1}^{N}$. In the second step, I show how the distributions $\{F_{Z|U=n}\}_{n=1}^{N}$ can be used to identify the joint distributions of the observed order statistics of the bids conditional on different values of the unobserved heterogeneity U, i.e, $\{F_{B^{(i_1)},B^{(i_2)},B^{(i_3)}|U=n}\}_{n=1}^{N}$), and the marginal distribution of the unobserved heterogeneity. In the final step, I use the observation of Song (2004) to identify the conditional distributions of bids, $F_{B|U}$, from the distribution of any pair of observed order statistics [22].

Theorem 3.29. *Suppose that the econometrician observes an i.i.d sample* $\{B_t^{(i_1)}, B_t^{(i_2)}, B_t^{(i_3)}, Z_t\}_{t=1}^{T}$ *from SPA and that assumption 3.26 holds. Then the conditional distributions of players' valuations given different realizations of the unobserved heterogeneity* U*,* $F_{V|U}$*, as well as the marginal distribution of the unobserved heterogeneity are identified.*

Remark 3.30. A simple modification of the proof of Proposition 3.22 shows that assumption 3.26 holds generically.

Remark 3.31. The identification argument of Theorem 3.29 does not extend to FPA. In particular, the argument relies on the observation that under exogenous entry (Assumption 2.3), the underlying marginal distribution of the players' bids (given a value of UH) does not depend on I (as bids are equal to values in SPA), and by Lemma 7.1 the conditional distribution of any order statistics of the bids, say $B^{(1)}$, given a lower order statistic, say $B^{(2)}$, and UH is independent of I (see equation 3.19). However, in the context of FPA, since the bidders' (common) strategy depends on the level of competition (see equation 3.9), the marginal distribution of the players' bids varies with I, even under the exogenous entry assumption. One possible approach to take would be to make the variable I part of the UH; however such an approach would require that one distinguishes

[22] The conditional distribution of $B^{(i_1)}$ given $B^{(i_2)} = a$ and U is independent of I and given by equation 3.19. The distribution $[F_{B|U}]_{[a,+\infty]}$ can then be identified by inverting relation 3.19, and taking the limit of such distributions as a approaches the lower bound of the support of $B^{(i_2)}$ identifies $F_{B|U}$

variations in the distributions of the mixture components that arise from changing I while holding other components of the UH fixed, in order to identify I and the underlying distribution of valuations through 2.4. Extending the identification result in 3.29 to FPA, would be valuable for empirical applications where the number of potential bidders also constitutes part of the UH (see An, Hu, and Shum (2010)). I leave the investigation of this question to future work.

4 Estimation

In this section, I provide estimators for the parameters identified in Theorem 3.9 and establish their statistical properties. The estimators build on the constructive identification arguments provided in the appendix. I leave the investigation of inference in the settings of Theorem 3.10 and Theorem 3.29 for future research. In what follows I assume that the cardinality of the UH N is known, as it can be consistently estimated using the sequential test of Kasahara and Shimotsu (2014). I first discuss the estimation of the distribution of the unobserved heterogeneity, $\delta = (P(U=1), \cdots, P(U=N))^T$.

Let $\Phi^1 = (\phi^1_1, \cdots, \phi^1_N)$, $\Phi^3 = (\phi^3_1, \cdots, \phi^3_N)$ and $\Phi^z = (\phi^z_1, \cdots, \phi^z_N)$ be $N-$component vectors of functions defined respectively on the support of $B^{(i_1)}|B^{(i_2)} = a$, $B^{(i_3)}|B^{(i_2)} = a$ and Z. Let the components $\{\phi^1_1, \cdots, \phi^1_N\}$ of Φ^1 (a similar statement applies to the elements of Φ^3 and Φ^z) be indicator functions of N sets that form a partition of the support of $B^{(i_1)}|B^{(i_2)} = a$. I discuss in the appendix how the sets that form the partition can be chosen. Let the matrices $\{A^j\}_{j=0}^N$ (all of dimension $N \times N$), be defined by:

$$A^0 = E\{\Phi^1(B^{(i_1)})\Phi^3(B^{(i_3)})^T | B^{(i_2)} = a\} f_{B^{(i_2)}}(a) \tag{4.1}$$

and

$$A^j = E\{\phi^z_j(Z)\Phi^1(B^{(i_1)})\Phi^3(B^{(i_3)})^T | B^{(i_2)} = a\} f_{B^{(i_2)}}(a) \tag{4.2}$$

for $j = 1, \cdots, N$. Note that the matrices $\{A^j\}_{j=0}^N$ are identified from the data. Moreover, as shown in the proof of Theorem 3.9, the partitions can be chosen such that the matrices $\{A^j\}_{j=0}^N$ are non-singular. Non-parametric kernel estimators for the matrices $\{A^j\}_{j=0}^N$ are given by

$$\hat{A}^0 = \frac{1}{T}\sum_{t=1}^{T} K_h\left(a - B_t^{(i_2)}\right)\Phi^1(B_t^{(i_1)})\Phi^3(B_t^{(i_3)})^T$$

and

$$\hat{A}^j = \frac{1}{T}\sum_{t=1}^{T} K_h\left(a - B_t^{(i_2)}\right)\Phi^z_j(Z_t)\Phi^1(B_t^{(i_1)})\Phi^3(B_t^{(i_3)})^T.$$

where $K_h(\cdot) := (1/h)K(\cdot/h)$ for a kernel function K, and $h > 0$ represents the kernel regularization parameter. I assume that the kernel K satisfies the following assumption:

Assumption 4.1. The kernel $K(\cdot)$ is a compactly supported bounded symmetric kernel of order 2.

Let the matrices $\{\hat{C}^j\}_{j=1}^N$ be defined by

$$\hat{C}^j = \hat{A}^j(\hat{A}^0)^{-1}, \tag{4.3}$$

and let the matrix $\hat{Q}$ be defined by

$$\hat{Q} = \arg\min_{Q \in \mathcal{Q}} \sum_{j=1}^{N} \|off(Q^{-1}\hat{C}^j Q)\|_F^2, \tag{4.4}$$

where $\mathcal{Q}$ denotes the collection of all invertible $N \times N$ probability matrices (columns are non-negative and sum to 1)[23]. Finally, let $\hat{M}$ denote the $N \times N$ matrix with its j^{th} row given by the diagonal elements of the matrix $\hat{D}^j$, where

$$\hat{D}^j = diag(\hat{Q}^{-1}\hat{C}^j\hat{Q}). \tag{4.5}$$

From the proof of Theorem 3.9, a natural estimator of δ (recall that $\delta = (P(U=1), \cdots, P(U = N))^T$) is given by

$$\hat{\delta} = \hat{M}^{-1}\hat{d}, \tag{4.6}$$

where

$$\hat{d} = \frac{1}{T}\sum_{t=1}^{T} \Phi^z(Z_t)$$

denotes an estimator of $E\Phi^z(Z)$. The theorem below provides the asymptotic distribution of $\hat{\delta}$; its proof is given in the appendix.

Theorem 4.2. *Suppose that the conditional distributions of bids $\{f_{B|U=k}\}_{k=1}^N$ are twice continuously differentiable on their supports, and let $K(\cdot)$ satisfy Assumption 4.1. Then provided that $Th \to \infty$ and $Th^5 \to 0$, the asymptotic distribution of the estimated distribution of the unobserved heterogeneity $\hat{\delta}$ is given by*

$$\sqrt{hT}(\hat{\delta} - \delta) \xrightarrow{d} N(0, R\Sigma_M R^T) \tag{4.7}$$

where $R = d^T M^{-T} \otimes_K M^{-1}$ and the covariance matrix Σ_M is as in 7.4.

Remark 4.3. Note that the conclusion of 4.2 differs from its counterpart in Bonhomme, Jochmans, and Robin (2016) and Bonhomme, Jochmans, and Robin (2014) (Theorem 2 and Corollary 2 in Bonhomme, Jochmans, and Robin (2014)), where it is shown that the mixture weights can be estimated at the parametric (root-n) rate. The main difference is due to the fact that from assumption 3.2, the identification of δ relies on the matrices $\{A^j\}$ which are only "identified locally" at a. In principle, one can recover the root-n rates in those papers by exploiting the fact that, by continuity, 3.2 holds for all a that belong to an interval contained in the interior of the intersection of the supports of the marginal distribution of bids conditional on different values of the unobserved heterogeneity (see Remark 3.4). I leave the investigation of this extension for future research.

[23] To solve this joint approximate diagonalization problem, as in Bonhomme, Jochmans, and Robin (2016), I use the algorithm of Luciani and Albera (2014)

I now turn to the estimation of the marginal distribution of bids conditional on the UH, $F_{B|U}$. For $p \in \{1,2,3\}$, let the stochastic process $\hat{y}^p(s)$ (defined for $s \in \mathbb{R}$) be defined by

$$\hat{y}^p(s) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{B_t^{(i_p)} \leq s\}\Phi^z(Z_t),$$

and let $\hat{\Delta}$ denote the $N \times N$ diagonal matrix, with diagonal elements given by $\hat{\delta}$. As shown in the proof of 3.9, a natural estimator of the vector

$$x^{(p)}(s) = (F_{B^{(i_p)}|U=1}(s), \cdots, F_{B^{(i_p)}|U=N}(s))^T$$

of conditional distributions of the i_p^{th} order statistic given different values of the UH, is given by

$$\hat{x}^{(p)}(s) = \hat{\Delta}^{-1}\hat{M}^{-1}\hat{y}^p(s). \tag{4.8}$$

The following theorem provides the asymptotic distribution of the process $\{\hat{x}^{(p)}(s)|s \in \mathbb{R}\}$.

Theorem 4.4. *Suppose that the conditional distributions of bids $\{f_{B|U=k}\}_{k=1}^N$ are twice continuously differentiable on their supports, and let $K(\cdot)$ satisfy Assumption 4.1. Then provided that $Th \to \infty$ and $Th^5 \to 0$, the asymptotic distribution of the estimator of the distribution of the i_p order statistics of bids conditional on the unobserved heterogeneity, $\hat{x}^{(p)}$ ($p \in \{1,2,3\}$), is given by*

$$\sqrt{hT}(\hat{x}^{(p)}(s) - x^{(p)}(s)) \rightsquigarrow \Gamma(s)N(0,\Sigma_M) \tag{4.9}$$

where the weak convergence is meant in the functional sense, the continuous matrix valued function Γ is given by

$$\Gamma(s) = \{x^{(p)}(s)^T \otimes_K \Delta^{-1}\}\left(\left(\sum_{j=1}^{N}(e_j^N \otimes_K e_j^N) \otimes_K (e_j^N)^T\right)R + \Delta \otimes_K M^{-1}\right),$$

and the matrix R and the covariance matrix Σ_M are as in Theorem 4.2.

Remark 4.5. Although I provide closed-form expressions for the covariance matrices that appear in Theorem 4.2 and 4.4 (see 7.4), for which consistent estimators can be easily constructed, the implementation of the estimators in 4.2 and 4.4 requires a suboptimal choice of bandwidth ($Th^5 \to 0$) to remove the asymptotic bias. It would be nice to have a data-driven procedure that performs this "undersmoothing" in practice. Furthermore, a close inspection of the proofs reveals that the covariance matrices in Theorem 4.2 and 4.4 depend on the partition functions Φ^1, Φ^3 and Φ^z, and their choice can be guided by efficiency considerations. I leave the investigation of such issues to future research.

5 Simulations

This section provides the results of a Monte Carlo study of the finite sample properties of the estimators of Theorem 4.2 and Theorem 4.4. The synthetic data, which simulates an ascending auction, is generated according to the following model:

$$U \in \{1,2,3\} \text{ and } U \sim \delta = (\delta_1, \delta_2, \delta_3),$$

$$F_{V|U} \sim Gamma(U,U) \text{ restricted to the interval } [0,2],$$

$$F_{Z|U} \sim Beta(U,1).$$

Contingent on the realization of U, I draw four independent draws from the distribution $F_{V|U}$, and I save the lowest three values ($i_1 = 1$, $i_2 = 2$, $i_3 = 3$) in the synthetic data set, as well as the realization of an independent draw from the distribution $F_{Z|U}$ of the instrument. I consider auction data of size $T \in \{300, 500, 1000\}$, and each Monte Carlo experiment is based on 500 repetitions. In the simulations, I assume that the number of support points of U ($N = 3$) is known, and I let the components of Φ^z be given by the indicator functions of the partition of the support $[0,1]$ into three intervals of equal lengths, i.e,

$$\Phi^z(\cdot) = (\mathbf{1}_{[0,1/3)}(\cdot), \mathbf{1}_{[1/3,2/3)}(\cdot), \mathbf{1}_{[2/3,1]}(\cdot)).$$

Analogously, I let the components of Φ^1 (resp. Φ^3) be given by the indicator functions of a partition of the interval $[a,2]$ (resp. $[0,a]$) into three sub-intervals of equal size, where a is as in Assumption 3.2. It is easy to show that condition 3.2 holds for all values of a in $(0,2)$ in our design; I use the value of $a = 1$ for my estimates. To estimate the matrices $\{A^j\}_{j=0}^3$ in equations 4.1 and 4.2, I use the triangular kernel $K(x) = (1 - |x|)\mathbf{1}_{[-1,1]}(x)$. The estimator in equation 4.6 (resp. 4.8) is not constrained to be a probability vector (resp. cumulative distribution functions); the imposition of such restriction on our estimators may lead to improved finite sample performance. In the simulations, I use these natural constraints as a selection criterion for the bandwidth: I estimate $\hat{x}^{(p)}$ and $\hat{\delta}$ for various values of h in the interval $[.1, .2]$, and I choose the value of h that minimizes

$$\sum_{i=1}^{3} \int_0^2 (\max\{-\hat{x}_i^{(1)}(t), 0\} + \max\{\hat{x}_i^{(1)}(t), 1\}) dt + ||\hat{\delta}||_1,$$

where $\hat{x}_i^{(1)}$ denotes the i^{th} component of $\hat{x}^{(1)}$. The later criterion penalizes estimates of $\hat{x}^{(1)}$ that are negative or larger than 1 on some sub-interval of $[0,2]$, as well as estimates of $\hat{\delta}$ that are too large. To deal with label swapping, as in Bonhomme, Jochmans, and Robin (2014), I estimate in each replication the means of the mixture components and label them according to the rank of their means (for instance, the component with the lowest estimated mean is associated with the corresponding component in the population: $U = 1$). The results of the simulations are presented below. Figure 1 and 2 show the outcome of the simulation when $T = 300$ and $T = 1000$, and for $\delta = (.3, .3, .4)$. Figure 3 shows the outcome of the simulation results when the distribution of U is given by $\delta = (.1, .3, .6)$, and for $T = 1000$. Each figure shows the average over the 500 Monte Carlo repetitions of the estimates of the distributions of the third order statistic ($x^{(2)}$ from 4.4) for different values of the UH (the solid black lines), as well as the population distributions of the third order statistic for different values of U (the dashed blue lines). The results for the estimates of $x^{(1)}$ and $x^{(3)}$ are similar. The stars in the figures represent the 25^{th} and 75^{th} percentiles (across the Monte Carlo replications) of the estimated cdfs at the corresponding points.

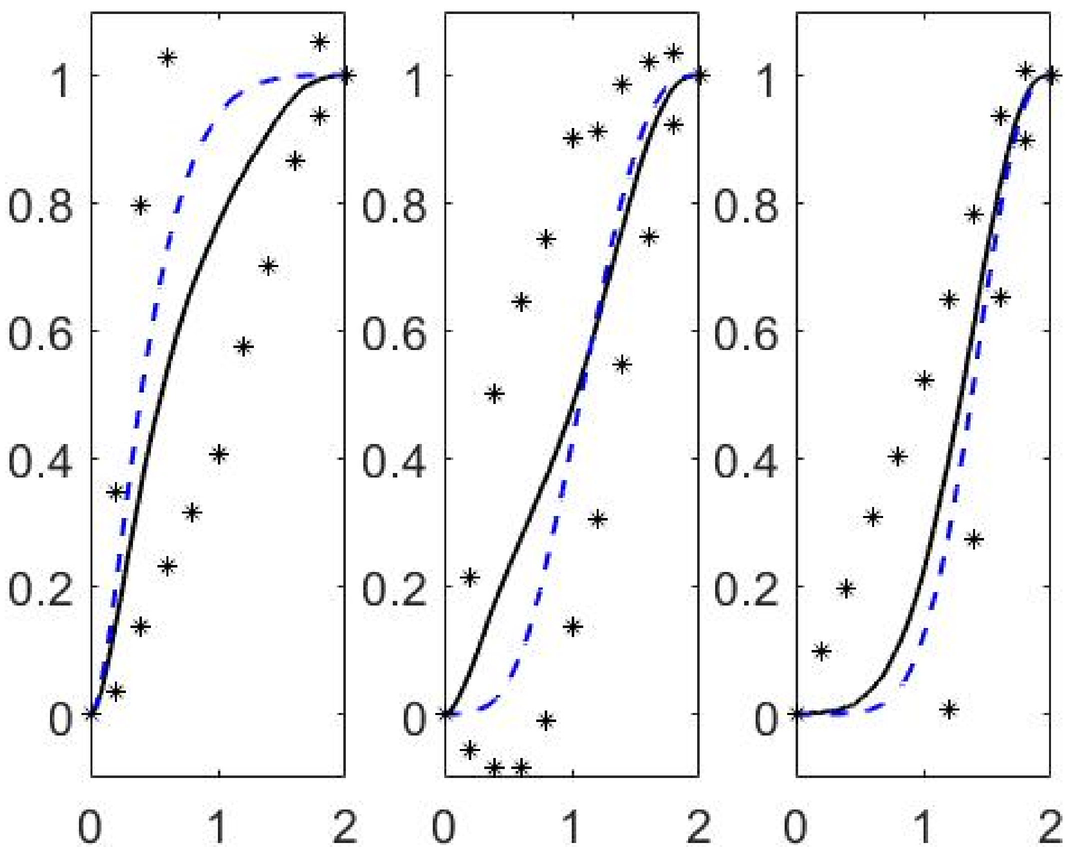

Figure 1: Estimates of the components of $x^{(2)}$ ($N = 300$, $\delta = (.3, .3, .4)$)

The performance of the estimator $\hat{x}^{(2)}$ is evaluated in terms of the root integrated mean squared error (RIMSE) (defined by $\sqrt{E||\hat{F} - F||_2^2}$) and provided in Table 1. To evaluate the performance of $\hat{\delta}$, I provide in Table 2 the average (over the 500 Monte Carlo repetitions) of its components, as well as their 25^{th} and 75^{th} percentiles.

6 Conclusion

This paper studies the identification of auction models with incomplete bid data in a setting where bidders' valuations are independent conditional on some auction level UH with finite support. By exploiting the Markov property of order statistics, this paper shows that the joint distribu-

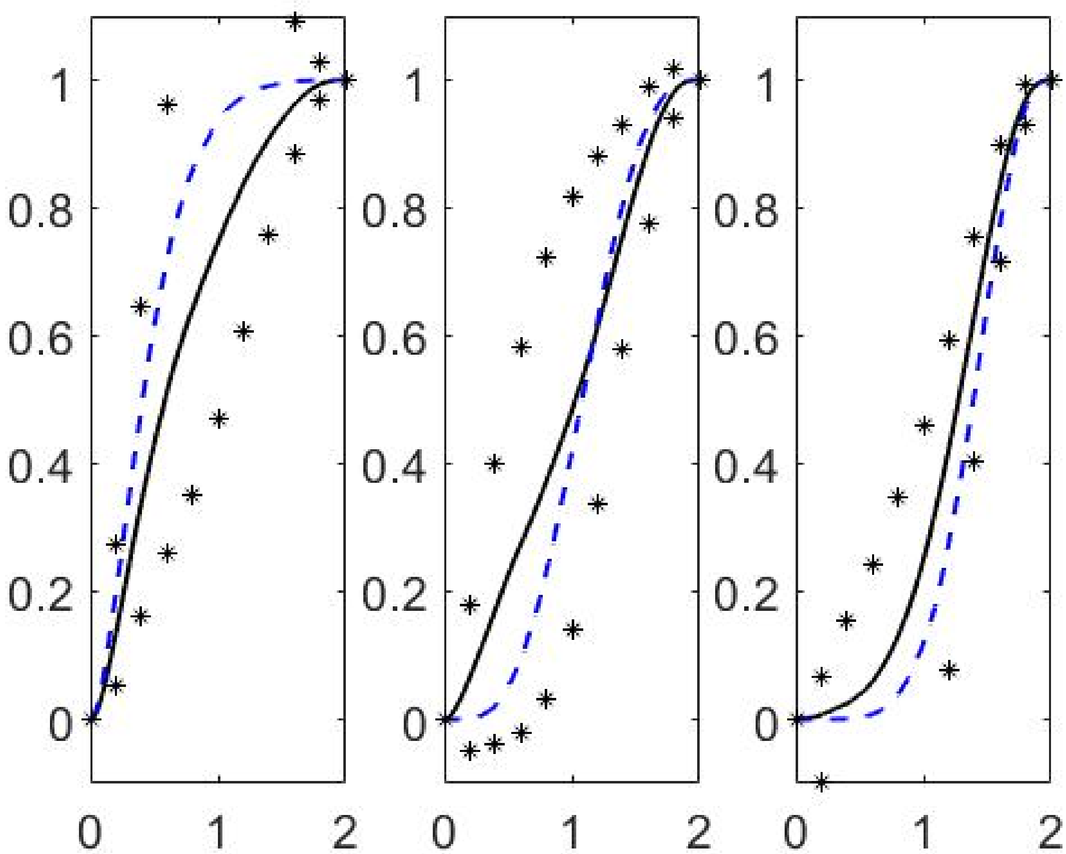

Figure 2: Estimates of the components of $x^{(2)}$ ($N = 1000$, $\delta = (.3, .3, .4)$)

$\delta = (.3, .3, .4)$			
	U=1	U=2	U=3
T=300	.282	.222	.223
T=500	.271	.215	.217
T= 1000	.2631	.200	.192
$\delta = (.1, .3, .6)$			
N=300	.459	.183	.147
N=500	.446	.163	.126
N= 1000	.444	.169	.114

Table 1: RIMSE of $\hat{x}^{(2)}$

tion of bidders' valuations and UH is point identified in both first and second price auction models

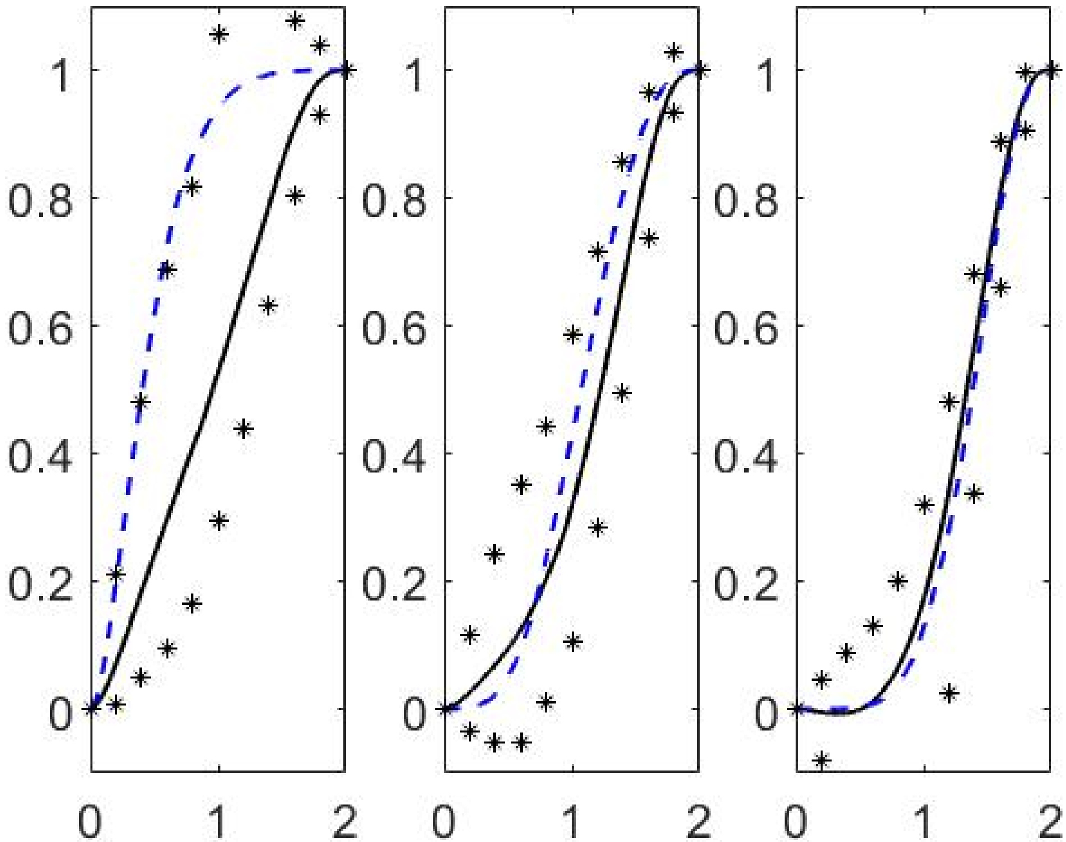

Figure 3: Estimates of the components of $x^{(2)}$ ($N = 500$, $\delta = (.1, .3, .6)$)

without relying on the availability of auxiliary auction data; all that is required is that the econometrician observes at least five order statistics of the bids in the auctions in her data set. When the econometrician has access to an instrument, the paper shows that observing at least three order statistics of the bids suffices for point identification, and identification still holds even if the econometrician does not observe the number of potential bidders, a setting that is relevant for online auctions. All the results are established under mild assumptions and without imposing any functional form restriction. I provide estimators that are based on the constructive identification arguments, and simulation results show that the estimators perform well for samples of moderate size.

[h!]

	Mean of $\hat{\delta}$			25^{th} to 75^{th} percentiles of $\hat{\delta}$		
	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
T=300	.235	.294	.343	[.042,.398]	[.102,.478]	[.149,.522]
T=500	.245	.327	.362	[.077,.397]	[.137,.492]	[.178,.531]
T= 1000	.264	.317	.398	[.072,.416]	[.100, .502]	[.202,.572]

Table 2: Mean and percentiles of $\hat{\delta}$ ($\delta = (.3, .3, .4)$)

References

An, Yonghong, Yingyao Hu, and Matthew Shum (2010). "Estimating first-price auctions with an unknown number of bidders: A misclassification approach". In: *Journal of Econometrics* 157.2, pp. 328–341.

Anderson, Robert and William Zame (2001). "Genericity with Infinitely Many Parameters". In: *The B.E. Journal of Theoretical Economics* 1.1, pp. 1–64.

Andrews, Donald W.K. (2017). "Examples of L2-complete and boundedly-complete distributions". In: *Journal of Econometrics* 199.2, pp. 213–220.

Aradillas-Lopez, Andres, Amit Gandhi, and Daniel Quint (2013). "Identification and Inference in Ascending Auctions With Correlated Private Values". In: *Econometrica* 81.2, pp. 489–534.

Armstrong, Timothy B. (2013). "Bounds in auctions with unobserved heterogeneity". In: *Quantitative Economics* 4.3, pp. 377–415.

Arnold, Barry (1992). *A First Course in Order Statistics, 3rd edition*. New York: John Wiley & Sons.

Aron, David H. and Nagaraja H. Navada (2003). *Order Statistics, 3rd edition*. New York: John Wiley & Sons.

Athey, Susan and Philip A. Haile (2002). "Identification of Standard Auction Models". In: *Econometrica* 70.6, pp. 2107–2140.

Athey, Susan, Jonathan Levin, and Enrique Seira (2011). "Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions". In: *The Quarterly Journal of Economics* 126.1, pp. 207–257.

Bonhomme, Stphane, Koen Jochmans, and Jean-Marc Robin (2014). "Nonparametric Estimation of Finite Mixtures from Repeated Measurements". In: *Journal of the Royal Statistical Society Series B* 78, pp. 211–229.

— (2016). "Estimating Multivariate Latent-Structure Models". In: *Annals of Statistics* 44.2, pp. 540–563.

Chen, YiChun and Siyang Xiong (2013). "Genericity and Robustness of Full Surplus Extraction". In: *Econometrica* 81.2, pp. 825–847.

Connault, Benjamin (2016). *Hidden Rust Models*. Working Paper.

Elizabeth, Allman, Catherine Matias, and John Rhodes (2009). "Identifiability of parameters in latent structure models with many observed variables". In: *Annals of Statistics* 37.6A, pp. 3099–3132.

Freyberger, Joachim and Bradley Larsen (2017). *Identification in ascending auctions, with an application to digital rights management*. NBER Working Papers.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong (2000). "Optimal Nonparametric Estimation of First-Price Auctions". In: *Econometrica* 68.3, pp. 525–574.

Hall, Peter and Xiao-Hua Zhou (2003). "Nonparametric estimation of component distributions in a multivariate mixture". In: *Annals of Statistics* 31.1, pp. 201–224.

Heifetz, Aviad and Zvika Neeman (2006). "On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design". In: *Econometrica* 74.1, pp. 213–233.

Hu, Yingyao (2008). "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution". In: *Journal of Econometrics* 144.1, pp. 27–61.

Hu, Yingyao, David McAdams, and Matthew Shum (2013). "Identification of first-price auctions with non-separable unobserved heterogeneity". In: *Journal of Econometrics* 174.2, pp. 186–193.

Hu, Yingyao and Susanne Schennach (2008). "Instrumental Variable Treatment of Nonclassical Measurement Error Models". In: *Econometrica* 76.1, pp. 195–216.

Hu, Yingyao and Matthew Shum (2012). "Nonparametric identification of dynamic models with unobserved state variables". In: *Journal of Econometrics* 171.1, pp. 32–44.

Hunt, Brian R., Tim Sauer, and James A. Yorke (1992). "Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces". In: *Bull. Amer. Math. Soc.* 27.2, pp. 217–238.

Kasahara, Hiroyuki and Katsumi Shimotsu (2009). "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices". In: *Econometrica* 77.1, pp. 135–175.

— (2014). "Non-parametric identification and estimation of the number of components in multivariate mixtures". In: *Journal of the Royal Statistical Society Series B* 76.1, pp. 97–111.

Kolmogorov, Andrey N. (1933). "Sulla determinazione empirico di una legge di distribuzione". In: *Giornale dell'Instituto Italiano d. Attuari* 4, pp. 83–91.

Krasnokutskaya, Elena (2011). "Identification and Estimation of Auction Models with Unobserved Heterogeneity". In: *Review of Economic Studies* 78.1, pp. 293–327.

Laffont, Jean-Jacques and Quang Vuong (1996). "Structural Analysis of Auction Data". In: *American Economic Review* 86.2, pp. 414–420.

Lathauwer, Lieven De, Bart De Moor, and Joos Vandewalle (2004). "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition". In: *Siam Journal on Matrix Analysis and Applications* 26, pp. 295–327.

Li, Tong, Isabelle Perrigne, and Quang Vuong (2000). "Conditionally independent private information in OCS wildcat auctions". In: *Journal of Econometrics* 98.1, pp. 129–161.

Li, Tong, Isabelle Perrigne, and Quang Vuong (2002). “Structural Estimation of the Affliated Private Value Auction Model”. In: *RAND Journal of Economics* 33.2, pp. 171–193.

Luciani, X. and L. Albera (2014). “Canonical polyadic decomposition based on joint eigenvalue decomposition”. In: *Chemom. Intell. Lab. Syst.* 132, pp. 152–167.

Magnus, Jan R. and H. Neudecker (1979). “The Commutation Matrix: Some Properties and Applications”. In: *Annals of Statistics* 7.2, pp. 381–394.

Maskin, Eric and John Riley (2000). “Asymmetric Auctions”. In: *Review of Economic Studies* 67, pp. 413–438.

Milgrom, Paul R and Robert J Weber (1982). “A Theory of Auctions and Competitive Bidding”. In: *Econometrica* 50.5, pp. 1089–1122.

Quint, Daniel (2015). *Identification in Symmetric English Auctions with Additively Separable Unobserved Heterogeneity*. Working Paper.

Riley, John G and William F Samuelson (1981). “Optimal Auctions”. In: *American Economic Review* 71.3, pp. 381–392.

Roberts, James W. (2013). “Unobserved heterogeneity and reserve prices in auctions”. In: *RAND Journal of Economics* 44.4, pp. 712–732.

Shaked M., Shanthikumar J. (2007). *Stochastic orders*. Springer.

Song, Unjy (2004). *Nonparametric estimation of an eBay auction model with an unknown number of bidders*. Working Paper.

Vaart, Aad van der and Jon A. Wellner (1996). *Weak convergence and empircal processes*. Springer Series in Statistics.

7 Appendix

7.1 Notation

I use $\otimes$ and $\otimes_K$ to denote respectively the tensor product and Kronecker product. Given a square matrix B, $diag(B)$ will denote the diagonal matrix that coincides with B on its diagonal, and $off(B)$ will denote the matrix that coincides with B on its off-diagonal and its zero on the diagonal. For a general matrix B, I will use B^+ to denotes its Moore-Penrose pseudo inverse and $\|B\|_F$ to denote its Frobenius norm. For a positive integer d, I_d will denote the $d \times d$ identity matrix, and I will use e_k^d to denote the k^{th} unit coordinate vector in $\mathbb{R}^d$. Given a matrix $B \in \mathbb{R}^{p \times q}$, let $vec(B) \in \mathbb{R}^{pq}$ denote the vector with its first block of p elements corresponding to the first column of B, its second block of p elements corresponding to the second column of B, and so on.

7.2 Omitted proofs of section 3

Before proceeding to the identification proofs, I first state two lemmas that will be used repeatedly throughout the identification argument. The first lemma describes how the distribution of one order statistic conditional on another is related to the parent distribution. For its proof, see Theorem 2.5 in Aron and Navada (2003).

Lemma 7.1. *Let $W_1, \cdots, W_I$ be independent observations from a continuous CDF F. Fix $1 \leq i < j \leq I$. Then,*

- *the conditional distribution of $W^{(i)}$ given $W^{(j)} = w$ is the same as the unconditional distribution of the ith order statistic in a sample of size $j-1$ from a new distribution, namely the original F truncated at the left at w,*
- *and the conditional distribution of $W^{(j)}$ given $W^{(i)} = w$ is the same as the unconditional distribution of the (j-i)th order statistic in a sample of size $I-i$ from a new distribution, namely the original F truncated at the right at w.*

The second lemma describes how the distribution of an order statistic is related to that of the parent distribution. Its proof can be found in Aron and Navada (2003) or Arnold (1992) (see also Athey and Haile (2002)).

Lemma 7.2. *The CDF of the i^{th} order statistic from a sample of size I from a continuous cdf F, which I denote by $F^{i:I}$, is a strictly monotonic function of the F. Indeed, $F^{i:I}(t) = \phi_{i:I}(F(t))$ where $\phi_{i:I}$ is the CDF of the i^{th} order statistic from I i.i.d draws from a uniform (on $[0,1]$) distribution, given explicitly by*

$$\phi_{i:n}(t) = \frac{n!}{(n-i)!(i-1)!}\int_0^t s^{n-i}(1-s)^{i-1}ds \tag{7.1}$$

for $t \in [0,1]$.

7.2.1 Proof of Theorem 3.10

Proof. The proof of Theorems 3.10 and 3.9 is similar in parts to that of Theorem 1 and 2 in Bonhomme, Jochmans, and Robin (2016) (see also Bonhomme, Jochmans, and Robin (2014), Hu, McAdams, and Shum (2013) and Kasahara and Shimotsu (2014)), the main complication in the present setting being due to the lack of conditional independence of the observed bids, which I overcome by exploiting the Markov property of order statistics. The following argument is done conditional on $\{I_t = I\}$ where I is a value in the support of I_t.

I begin by showing that the cardinality N of the support of the unobserved heterogeneity U is identified. Fix $0 < b < a$ and let Δ^1, Δ^2 and Δ^3 be arbitrary finite partitions of $[0,b]$, $[b,a]$ and $[a,+\infty]$ respectively, where $\Delta^1 = \{\delta^1_1, \cdots, \delta^1_{|\Delta^1|}\}$, $\Delta^2 = \{\delta^2_1, \cdots, \delta^2_{|\Delta^2|}\}$ and $\Delta^3 = \{\delta^3_1, \cdots, \delta^3_{|\Delta^3|}\}$.

Here $|A|$ denotes the cardinality of the set A. By 2.1 and the Markov property of order statistics, for $i \in \{1, \cdots, |\Delta^1|\}$ and $j \in \{1, \cdots, |\Delta^3|\}$, we have:

$$P\left(B^{(i_1)} \in \delta_i^1, B^{(i_5)} \in \delta_j^3 | B^{(i_2)} = a, B^{(i_4)} = b\right) =$$
$$\sum_{k=1}^{N} P(U = k | B^{(i_2)} = a, B^{(i_4)} = b) P\left(B^{(i_1)} \in \delta_i^1, B^{(i_5)} \in \delta_j^3 | B^{(i_2)} = a, B^{(i_4)} = b, U = k\right)$$

$$= \sum_{k=1}^{N} P(U = k | B^{(i_2)} = a, B^{(i_4)} = b) P\left(B^{(i_1)} \in \delta_i^1 | B^{(i_2)} = a, U = k\right) P\left(B^{(i_5)} \in \delta_j^3 | B^{(i_4)} = b, U = k\right)$$

Let the $M \in \mathbb{R}^{|\Delta^1| \times |\Delta^3|}$ be defined by $M_{i,j} = P\left(B^{(i_1)} \in \delta_i^1, B^{(i_5)} \in \delta_j^3 | B^{(i_2)} = a, B^{(i_4)} = b\right)$, where $i \in \{1, \cdots, |\Delta^1|\}$ and $j \in \{1, \cdots, |\Delta^3|\}$. For $k \in \{1, \cdots, N\}$, $i \in \{1, \cdots, |\Delta^1|\}$ and $j \in \{1, \cdots, |\Delta^3|\}$, let $u_k \in \mathbb{R}^{|\Delta^1|}$ and $v_k \in \mathbb{R}^{|\Delta^3|}$ be respectively defined by $[u_k]_i = P\left(B^{(i_1)} \in \delta_i^1 | B^{(i_2)} = a, U = k\right)$ and $[v_k]_j = P\left(B^{(i_5)} \in \delta_j^3 | B^{(i_4)} = b, U = k\right)$, and set $\lambda_k = P(U = k | B^{(i_2)} = a, B^{(i_4)} = b)$. The previous equation then becomes:

$$M = \sum_{k=1}^{N} \lambda_k u_k v_k^T. \tag{7.2}$$

Note that the matrix M, the vectors u_k and v_k, and the constants λ_k, depend on our choice of a, b and the partitions Δ^1 and Δ^3. Equation 7.2 implies that the identified matrix M has rank at most N, for any choice of a, b, and of the partitions Δ^1 and Δ^3. By assumption 3.3 and by *Lemma* 17 of Elizabeth, Matias, and Rhodes (2009), there exist $0 < b < a$ and partitions Δ^1 and Δ^3, such that the collection of vectors $\{u_1, \cdots, u_N\}$ and $\{v_1, \cdots, v_N\}$ each form a linearly independent set, and the coefficients λ_k are all positive. The corresponding matrix M thus has rank N. In conclusion, the maximal rank of the identified matrices M over different choices of a, b and of the partitions Δ^1, Δ^2 and Δ^3, is equal to N.

Let $0 < b < a$ and the partitions Δ^1 and Δ^3 be chosen such that the matrix M has maximal rank N. A simple modification of Lemma 17 in Elizabeth, Matias, and Rhodes (2009) shows that the partitions Δ^1 and Δ^3 can be chosen to have cardinality N: $|\Delta^1| = |\Delta^3| = N$. For this choice of partitions, I now show how the terms on the RHS of 7.2 are identified (up to permutation of indices). Let Δ^2 be a partition of $[a, b]$ such that $|\Delta_2| = N$. For $i, k \in \{1, \cdots, N\}$, let $w_k \in \mathbb{R}^N$ be defined by $[w_k]_i = P\left(B^{(i_3)} \in \delta_i^2 | B^{(i_2)} = a, B^{(i_4)} = b, U = k\right)$. By lemma 17 in Elizabeth, Matias, and Rhodes (2009) and assumption 3.3, the partition Δ^2 can be chosen such that the vectors $\{w_k\}_{k=1}^N$ are linearly independent. For $p, i, j \in \{1, \cdots, N\}$, let the matrices $M_p \in \mathbb{R}^{N \times N}$ be defined by

$$[M_p]_{i,j} := P\left(B^{(i_1)} \in \delta_i^1, B^{(i_5)} \in \delta_j^3 | B^{(i_2)} = a, B^{(i_4)} = b, B^{(i_3)} \in \delta_p^2\right).$$

Assumption 2.1 and the Markov property of order statistics yield

$$M_p = \sum_{k=1}^{N} \lambda_{p,k} u_k v_k^T, \tag{7.3}$$

where $\lambda_{p,k} = P(U = k | B^{(i_2)} = a, B^{(i_4)} = b, B^{(i_3)} \in \delta_p^2)$, and the vectors u_k and v_k are defined as in the preceding paragraph. Let $U \in \mathbb{R}^{N\times N}$ (resp. $V \in \mathbb{R}^{N\times N}$) be the matrix with k^{th} column given by u_k (resp. v_k), and let $\Lambda_p \in \mathbb{R}^{N\times N}$ (resp. $\Lambda \in \mathbb{R}^{N\times N}$) be the diagonal matrix with k^{th} diagonal element given by $\lambda_{p,k}$ (resp. λ_k). Equations 7.2 and 7.3 then respectively become

$$M = U\Lambda V^T \text{ and } M_p = U\Lambda_p V^T. \tag{7.4}$$

Since the columns of the matrices U and V are linearly independent (by the choice of partition) and the diagonal elements of Λ are non-zero [24], the matrix M is invertible. For $p \in \{1, \cdots, N\}$, define $\tilde{M}_p := M_p M^{-1}$ and $\tilde{\Lambda}_p = \Lambda_p \Lambda^{-1}$, we get

$$\tilde{M}_p = U\tilde{\Lambda}_p U^{-1}. \tag{7.5}$$

Therefore, for $p = 1, \cdots, N$, the identified matrix $\tilde{M}_p$ is similar to the diagonal matrix $\tilde{\Lambda}_p$, and the eigenvalues of $\tilde{M}_p$ identify the diagonal elements of $\tilde{\Lambda}_p$. Let $\tilde{\lambda}_{p,k}$ denote the k^{th} diagonal element of the matrix $\tilde{\Lambda}_p$, and let $D \in \mathbb{R}^{N\times N}$ denote the matrix with p^{th} column given by the diagonal elements of $\tilde{\Lambda}_p$. A simple application of Bayes' rule yields

$$\tilde{\lambda}_{p,k} = \frac{P(B^{(i_3)} \in \delta_p^2 | B^{(i_2)} = a, B^{(i_4)} = b, U = k)}{P(B^{(i_3)} \in \delta_p^2 | B^{(i_2)} = a, B^{(i_4)} = b)}.$$

Therefore, by our choice of the partition Δ^2, the matrix D has full rank, and an application of Theorem 6 in Lathauwer, Moor, and Vandewalle (2004) implies that there exists a unique (up to permutation of the columns) probability matrix[25] that simultaneously diagonalizes the matrices $\tilde{M}_p$, for $p = 1, \cdots, N$. In conclusion, the matrix U is identified up to a permutation of its columns. Let $e \in \mathbb{R}^N$ be given by $e = (1, 1, \cdots, 1)^T$. Since V is a probability matrix (see 25), the diagonal elements of Λ are identified by

$$U^{-1}Me = \Lambda V^T e = \Lambda e.$$

Finally, identification of Λ yields identification of V through

$$V^T = \Lambda^{-1}U^{-1}M.$$

I have thus shown that all terms on the RHS of 7.2 are identified up to a permutation of the indices k.

I now show that the cdfs $F_{B|U=k}$ for $k \in \{1, \cdots, N\}$, which represent the common (by 2.1 and 2.2) marginal distributions of players' bids given the UH, are identified. I begin by showing that the conditional distributions $F_{B^{(i_1)}|\{B^{(i_2)}=a,U=k\}}$, $F_{B^{(i_3)}|\{B^{(i_2)}=a,B^{(i_4)}=b,U=k\}}$ and $F_{B^{(i_5)}|\{B^{(i_4)}=b,U=k\}}$ for

[24] For assumption 3.3 to hold, a and b must belong to the interior of the supports of the distributions $F_{V|U=u}$ (see 2.1).

[25] By a probability matrix I mean any matrix with non-negative entries such that the entries of each columns sum up to 1.

$k \in \{1, \cdots, N\}$ are identified. For $t \in [a, +\infty]$, let $x^1(t)$ and $y^1(t) \in \mathbb{R}^N$ be defined by $[y^1(t)]_i := P(B^{(i_1)} \leq t, B^{(i_5)} \in \delta_i^3 | B^{(i_2)} = a, B^{(i_4)} = b)$ and $[x^1(t)]_i := P(B^{(i_1)} \leq t | B^{(i_2)} = a, U = i)$ for $i \in \{1, \cdots, N\}$. By 2.1 and the Markov property, we have

$$y^1(t) = V\Lambda x^1(t),$$

Where the identified matrices V and Λ are as in the preceding paragraph. The vector $x^1(t)$ is thus identified through:

$$x^1(t) = \Lambda^{-1}V^{-1}y^1(t).$$

Identification of the vectors $x^1(t)$ $t \in [b, a]$, yields the identification of the distributions $\{F_{B^{(i_1)}|\{B^{(i_2)}=a, U=k\}}\}_{k=1}^N$. Similarly, for $t \in [a, b]$, let $x^2(t)$ and $y^2(t) \in \mathbb{R}^N$ be defined by $[y^2(t)]_i := P(B^{(i_3)} \leq t, B^{(i_5)} \in \delta_i^3 | B^{(i_2)} = a, B^{(i_4)} = b)$ and $[x^2(t)]_i := P(B^{(i_3)} \leq t | B^{(i_2)} = a, B^{(i_4)} = b, U = i)$ for $i \in \{1, \cdots, N\}$. A similar argument to the one above yields

$$x^2(t) = \Lambda^{-1}V^{-1}y^2(t),$$

and the distributions $\{F_{B^{(i_3)}|\{B^{(i_2)}=a, B^{(i_4)}=b, U=k\}}\}_{k=1}^N$ are identified. Finally for $t \in [0, b]$, let $x^3(t)$ and $y^3(t) \in \mathbb{R}^N$ be defined by $[y^3(t)]_i := P(B^{(i_5)} \leq t, B^{(i_1)} \in \delta_i^1 | B^{(i_2)} = a, B^{(i_4)} = b)$ and $[x^3(t)]_i := P(B^{(i_5)} \leq t | B^{(i_4)} = b, U = i)$ for $i \in \{1, \cdots, N\}$. An argument similar to the one above yields

$$x^3(t) = \Lambda^{-1}U^{-1}y^3(t),$$

where the identified matrix U is as in the preceding paragraph. The latter equality yields the identification of the distributions $\{F_{B^{(i_5)}|\{B^{(i_4)}=b, U=k\}}\}_{k=1}^N$.

I now show how to recover the distributions $F_{B|U=k}$ for $k \in \{1, \cdots, N\}$, from the identified distributions $\{F_{B^{(i_1)}|\{B^{(i_2)}=a, U=k\}}\}_{k=1}^N$, $\{F_{B^{(i_3)}|\{B^{(i_2)}=a, B^{(i_4)}=b, U=k\}}\}_{k=1}^N$ and $\{F_{B^{(i_5)}|\{B^{(i_4)}=b, U=k\}}\}_{k=1}^N$. By Lemma 7.1, the distribution of $F_{B^{(i_1)}|\{B^{(i_2)}=a, U=k\}}$ is the same as the distribution of the i_1^{th} order statistic from an i.i.d sample of size $i_2 - 1$ from $[F_{B|U=k}(\cdot|U = k) - F_{B|U=k}(a|U = k)]/[1 - F_{B|U=k}(a|U = k)]$, i.e, from the parent distribution $F_{B|U=k}$ with the left tail truncated at a. Similarly, the distribution of $F_{B^{(i_3)}|\{B^{(i_2)}=a, B^{(i_4)}=b, U=k\}}$ is the same as the distribution of the $(i_3 - i_2)^{th}$ order statistic from an i.i.d sample of size $i_4 - i_2 - 1$ from the distribution $[F_{B|U=k}(\cdot|U = k) - F_{B|U=k}(a|U = k)]/[F_{B|U=k}(b|U = k) - F_{B|U=k}(a|U = k)]$. Finally, the distribution of $F_{B^{(i_5)}|\{B^{(i_4)}=b, U=k\}}$ is the same as that of the $(i_5 - i_4)^{th}$ order statistic from an i.i.d sample of size $I - i_4$ (where I denotes the number of bidders) from the distribution $F_{B|U=k}(\cdot|U = k)/F_{B|U=k}(b|U = k)$. Hence, the cdfs $F_{B|U=k}(\cdot|U = k)/F_{B|U=k}(a|U = k)$, $[F_{B|U=k}(\cdot|U = k) - F_{B|U=k}(a|U = k)]/[F_{B|U=k}(b|U = k) - F_{B|U=k}(a|U = k)]$ and $[F_{B|U=k}(\cdot|U = k) - F_{B|U=k}(b|U = k)]/[1 - F_{B|U=k}(b|U = k)]$ are (respectively) identified from $F_{B^{(i_1)}|\{B^{(i_2)}=a, U=k\}}$, $F_{B^{(i_3)}|\{B^{(i_2)}=a, B^{(i_4)}=b, U=k\}}$ and

$F_{B^{(i_5)}|\{B^{(i_4)}=b,U=k\}}$[26]. It remains to show that $F_{B|U=k}(b|U=k)$ and $F_{B|U=k}(a|U=k)$ are identified. However, a simple argument shows that $F_{B|U=k}(b|U=k)$ and $F_{B|U=k}(a|U=k)$ can be expressed as functions of the density functions corresponding to the cumulative distribution functions identified above, evaluated at the points a and b[27]. Therefore, the distributions $\{F_{B|U=k}\}_{k=1}^{N}$ are identified.

It now remains to identify the marginal distribution of the unobserved heterogeneity $\{P(U=k)\}_{k=1}^{N}$. By Bayes' rule, we have

$$P(U=k) = \frac{\lambda_k f_{B^{(i_2)},B^{(i_4)}}(a,b)}{f_{B^{(i_2)},B^{(i_4)}|U=k}(a,b)}.$$

Here, $f_{B^{(i_2)},B^{(i_4)}|U=k}$ denotes the joint density of the i_2^{th} and i_4^{th} order statistics given U, which is identified from $F_{B|U=k}$; $f_{B^{(i_2)},B^{(i_4)}}$ denotes their unconditional distribution, and is identified from the joint distribution of the observed bids; and λ_k is as in 7.2, and is identified since Λ is identified. This concludes the proof of Theorem 3.10.

□

7.2.2 Proof of Theorem 3.9

Proof. As in the proof of Theorem 3.10, the following argument is done conditional on $\{I = I_0\}$ where I_0 is a value in the support of I, and by the same argument used in the second paragraph of that proof, the cardinality N of the support of the unobserved heterogeneity U is identified. Let the matrices $\{A^j\}_{j=0}^{N}$ (all of dimension $N \times N$), be defined by as in 4.1 and 4.2. Assumption 3.2 implies that

$$A^0 = \sum_{k=1}^{N} f_{B^{(i_2)}|U=k}(a)P(U=k)E\{\Phi^1(B^{(i_1)})|B^{(i_2)}=a,U=k\} \otimes E\{\Phi^3(B^{(i_3)})|B^{(i_2)}=a,U=k\} \quad (7.6)$$

and

$$\begin{aligned} A^j = \sum_{k=1}^{N} f_{B^{(i_2)}|U=k}(a)P(U=k)E\{\phi_j^z(Z)|U=k\} \\ E\{\Phi^1(B^{(i_1)})|B^{(i_2)}=a,U=k\} \otimes E\{\Phi^3(B^{(i_3)})|B^{(i_2)}=a,U=k\} \end{aligned} \quad (7.7)$$

$j = 1, \cdots, N$. The relations of equations 7.6 and 7.7 can be written in matrix form as

$$A^0 = Q^1\Pi(Q^3)^T \quad (7.8)$$

and

$$A^j = Q^1\Pi D^j(Q^3)^T \quad (7.9)$$

[26]For instance, by Lemma 7.2 we have that $F_{B^{(i_5)}|\{B^{(i_4)}=b,U=k\}} = \phi_{i_5-i_4:I-i_4} \circ (F_{B|U=k}(\cdot|U=k)/F_{B|U=k}(b|U=k))$, and note that $\phi_{i_5-i_4:I-i_4}$ is invertible.

[27]Recall that by assumption, $f_{V|U}(\cdot|U)$ is continuous.

for $j = 1, \cdots, N$, where Q^p (for $p = 1$ or 3) denotes the $N \times N$ matrix with k^{th} column given by $E\{\Phi^p(B^{(i_p)})|B^{(i_2)} = a, U = k\}$, Π denotes the diagonal matrix with k^{th} diagonal element given by $f_{B^{(i_2)}|U=k}(a)P(U = k)$, and D^j ($j = 1, \cdots, N$) denote the $N \times N$ diagonal matrix with k^{th} diagonal element given by $E\{\phi_j^z(Z)|U = k\}$. Finally, let M denote the $N \times N$ matrix with j^{th} row given by the diagonal elements of the matrix D^j. By assumption 3.2, the components of Φ^1, Φ^3 and Φ^z can be chosen such that the matrices Q^1, Q^3, M and A^0 (see footnote 24) have full rank.

I now show that the matrix M is identified (up to a permutation of its columns) from the matrices A^j. By assumption, the matrix A^0 has full rank, and post-multiplying the other matrices A^j by the inverse of A^0 yields

$$A^j(A^0)^{-1} = Q^1 D^j (Q^1)^{-1} \tag{7.10}$$

for $j = 1, \cdots, N$. Hence the matrices $C^j \equiv A^j(A^0)^{-1}$ and D^j are similar, and the matrices $\{C^j\}_{j=1}^N$ are simultaneously diagonalized by the matrix Q^1. Moreover, since M has full rank, by Theorem 6 in Lathauwer, Moor, and Vandewalle (2004), Q^1 is the unique (up to a permutation of its columns) probability matrix that simultaneously diagonalizes the matrices $\{C^j\}_{j=1}^N$. Since the matrices $\{C^j\}_{j=1}^N$ are all identified from the data, it follows that the matrices $\{D^j\}_{j=1}^N$, and hence the matrix M, are identified.

I now show that the distribution of the unobserved heterogeneity is identified. Let $\delta = (P(U = 1), \cdots, P(U = k))^T$ and let $d = E\Phi^z(Z)$. Since $d = M\delta$, and M has full rank, the vector δ is identified by

$$\delta = M^{-1}d. \tag{7.11}$$

I finally show that the marginal (common) distributions of bids given the unobserved heterogeneity, $F_{B|U=k}$ ($k = 1, \cdots, N$), are identified. For $s \in \mathbb{R}$, let the vectors $y^p(s)$ and $x^{(p)}(s)$ (both in $\mathbb{R}^N$) for $p \in \{1, 2, 3\}$, be defined by $y^p(s) = E\mathbf{1}\{B^{(i_p)} \leq s\}\Phi^z(Z)$, and let the k^{th} element of $x^{(p)}(s)$ be defined $F_{B^{(i_p)}|U=k}(s)$. Assumption 3.2 implies

$$y^p(s) = M\Delta x^{(p)}(s),$$

Where Δ is the diagonal matrix with diagonal elements given by δ. Therefore, $x^{(p)}(s)$ is identified by

$$x^{(p)}(s) = \Delta^{-1}M^{-1}y^p(s). \tag{7.12}$$

The identification of the distributions $F_{B|U=k}$ then follows from the relations

$$F_{B^{(i_p)}|U=k}(s) = \phi_{i_p:n} \circ F_{B|U=k}(s) \tag{7.13}$$

for $p \in \{1, 2, 3\}$, and where $\phi_{i:n}$ is as in equation 7.1. $\square$

7.2.3 Proof of Theorem 3.20

Proof. The proof is similar to that of Theorem 3.9 and Theorem 3.10. The following argument is done conditional on $\{I = I_0\}$ where I_0 is a value in the support of I. By considering the rank of matrices formed by partitioning the domain of $[B^{(i_1)}|B^{(i_2)} = a]$ and Z, an argument similar to that used in the proof of Theorem 3.9 shows that N is identified. As in the definition of 4.1 and 4.2, let $\Phi^1 = (\phi_1^1, \cdots, \phi_N^1)$, $\Phi^z = (\phi_1^z, \cdots, \phi_N^z)$ and $\Phi^3 = (\phi_1^3, \cdots, \phi_M^3)$ (with $M \geq N$) be N and M-component vectors of functions defined respectively on the support of $[B^{(i_1)}|B^{(i_2)} = a]$, Z and $[B^{(i_3)}|B^{(i_2)} = a]$. Let the components $\{\phi_1^1, \cdots, \phi_N^1\}$ of Φ^1 (a similar statement applies to the elements of Φ^z) be indicator functions of N sets that form a partition of the support of $B^{(i_1)}|B^{(i_2)} = a$, and let the components of $\{\phi_1^3, \cdots, \phi_M^3\}$ be indicator functions of M sets that form a partition of the support of $B^{(i_3)}|B^{(i_2)} = a$. By conditions 3.3 and 3.15, as in the proof of Theorem 3.10, the partitions that form Φ^1 and Φ^z can be chosen such that the $N \times N$ matrices Q^1 and Q^z are invertible, where the matrix Q^1 (resp. Q^z) is such that its k^th column is given by $E\{\Phi^1(B^{(i_1)})|B^{(i_2)} = a, U = k\}$ (resp. $E\{\Phi^z(Z)|U = k\}$). And by condition 3.16 the partitions that form Φ^3 can be chosen such that the $M \times N$ matrix Q^3 with k^{th} column given by $E\{\Phi^3(B^{(i_3)})|B^{(i_2)} = a, U = k\}$, for $k = 1, \cdots, N$, has distinct columns. Let the matrices $\{A^j\}_{j=0}^M$ be defined by

$$A^0 = E\{\Phi^z(Z)\Phi^1(B^{(i_1)})^T|B^{(i_2)} = a\}f_{B^{(i_2)}}(a) \tag{7.14}$$

and, for $j = 1, \cdots, M$

$$A^j = E\{\Phi_j^3(B^{(i_3)})\phi^z(Z)\Phi^1(B^{(i_1)})^T|B^{(i_2)} = a\}f_{B^{(i_2)}}(a). \tag{7.15}$$

Then as in 7.8 and 7.9, the matrices A^j have the representation

$$A^0 = Q^z\Pi(Q^1)^T \tag{7.16}$$

and for $j = 1, \cdots, M$

$$A^j = Q^z\Pi D^j(Q^1)^T, \tag{7.17}$$

where the matrix Π is as defined in 7.8, and the $N \times N$ matrix D_j ($j = 1 \cdots, M$) is now the diagonal matrix with diagonal elements given by the j^{th} row of Q^3. As in 7.10, we have

$$A^j(A^0)^{-1} = Q^z D^j(Q^z)^{-1} \tag{7.18}$$

for $j = 1, \cdots, M$. Hence the matrices $A^j(A^0)^{-1}$ and D^j are similar, and the matrices $\{A^j(A^0)^{-1}\}_{j=1}^N$ are simultaneously diagonalized by the matrix Q^z. Moreover, since the columns of Q^3 are distinct, by Theorem 6 in Lathauwer, Moor, and Vandewalle (2004), Q^z is the unique (up to a permutation of its columns) probability matrix that simultaneously diagonalizes the matrices $\{A^j(A^0)^{-1}\}_{j=1}^N$. Therefore Q^z is identified. The rest of the identification argument proceeds as in the proof of 3.9 (the paragraph following 7.10). □

7.2.4 Proofs on Genericity

I first establish the existence of an element $\mathbf{G^*} \in \mathcal{C}$ described before the statement of Proposition 3.22. First note that all elements of $\mathcal{A}$ are polynomials (see 7.2). It is then easy to check that the element $\mathbf{G^*}$ with i^{th} element equal to the polynomial t^i ($t \in [0,1]$) has the desired property. Indeed, given $\phi \in \mathcal{A}$, $rank(\phi \circ \mathbf{G^*}) = N$, since all the elements of $\phi \circ \mathbf{G^*}$ are polynomials of distinct degrees.

Before proceeding to the proof of proposition 3.22, I first recall some results from Anderson and Zame (2001) that will be useful for the proof. By Fact 3 in Anderson and Zame (2001) (p.12), the countable union of sets that are shy in $\mathcal{C}$ is shy in $\mathcal{C}$. Hence, since $E_{a,b} = \cup_{\phi \in \mathcal{A}} E_{a,b,\phi}$ (a countable union), with $E_{a,b,\phi}$ defined by $E_{a,b,\phi} := \{\mathbf{F} \in \mathcal{C} | \min\{rank(\phi \circ \mathbf{F}_{[0,b]}), rank(\phi \circ \mathbf{F}_{[b,a]}), rank(\phi \circ \mathbf{F}_{[a,1]})\} < N\}$, it suffices to show that each $E_{a,b,\phi}$ is shy in $\mathcal{C}$ for each $\phi \in \mathcal{A}$. Let $S_1 = [0,b]$, $S_2 = [b,a]$ and $S_3 = [a,1]$. Since $E_{a,b,\phi} = E_{a,b,\phi,1} \cup E_{a,b,\phi,2} \cup E_{a,b,\phi,3}$, with $E_{a,b,\phi,i} := \{\mathbf{F} \in \mathcal{C} | rank(\phi \circ \mathbf{F}_{S_i}) < N\}$, it suffices to show that each $E_{a,b,\phi,i}$ (i=1,2,3) is shy in $\mathcal{C}$. Note that when S_i is a singleton, the set $E_{a,b,\phi,i}$ is empty, thus shy in $\mathcal{C}$. Hence it remains to prove the each $E_{a,b,\phi,i}$ is shy in $\mathcal{C}$ when S_i is a non-degenerate interval.

Given a finite dimensional subspace V (say of dimension d) of $\mathcal{X}$, let λ_V denote a Lebesgue measure on V, defined by $\lambda_V(A) = \mu_d(T(A))$, where T is an isomorphism between V and $\mathbb{R}^d$, μ_d denotes the Lebesgue measure on $\mathbb{R}^d$ and A is any Borel subset of V (see p.12 of Anderson and Zame (2001)). I now provide the definition of a notion that is used in our proof below. The original statement can be found in Anderson and Zame (2001) (p.12).

Definition 7.3. A Borel subset E of $\mathcal{C}$ is finitely shy in $\mathcal{C}$, if there is a finite dimensional subspace $V \subset \mathcal{X}$ such that $\lambda_V(\mathcal{C} + a) > 0$ for some $a \in \mathcal{X}$ and $\lambda_V(E + x) = 0$ for every $x \in \mathcal{X}$.

To establish that each set $E_{a,b,\phi,i}$ is shy in $\mathcal{C}$, I use Fact 6 of Anderson and Zame (2001) which states that: Every set which is finitely shy in $\mathcal{C}$ is shy in $\mathcal{C}$. Therefore, to prove Proposition 3.22, it suffices to prove the following proposition.

Proposition 7.4. *Suppose that S_i is a non-degenerate interval. Then each set $E_{a,b,\phi,i}$ is finitely shy in $\mathcal{C}$.*

Proof. I first consider the simple case when $\phi(t) = \phi_{1:1}(t) = t$ (for $t \in [0,1]$), since the steps of the argument are more transparent in that case. I provide the proof for a general $\phi \in \mathcal{A}$ further below.
Simple case. Let $f = (f_1, \cdots, f_N)$ be an $N-$tuple of density functions supported on S_i that are linearly independent and bounded away from zero on S_i.[28] Let $h = (h_1, h_2, \cdots, h_N)$ denote an $N-$ *tuple* of bounded functions that are linearly independent and supported on S_i, with $\int_0^1 h_i(t)dt = 0$

[28] An example of such densities would be the restriction and re-normalization of the following $N-$tuple of functions $(t+1, t^2+1, \cdots, t^N+1)$ (where $t \in [0,1]$) on the set S_i, the constant 1 being added to each component to guarantee that the re-normalized densities are bounded away from zero on S_i. Recall that a non-trivial linear combination of polynomial of distinct degrees cannot vanish on an interval.

for $i = 1, \cdots, N$.[29] Let $\mathbf{F} = (F_1, \cdots, F_N) \in \mathcal{C}$ and $\mathbf{H} = (H_1, \cdots, H_N) \in \mathcal{X}$ be defined by

$$F_i(s) = \int_0^s f_i(t)dt \quad \text{and} \quad H_i(s) = \int_0^s h_i(t)dt, \tag{7.19}$$

for $i = 1, \cdots, N$ and $s \in [0,1]$. Let the one-dimensional subspace V of $\mathcal{X}$ be defined by $V = \{\alpha\mathbf{H} | \alpha \in \mathbb{R}\}$ with the "Lebesgue" measure on V, λ_V, given by $\lambda_V(A) = \mu_1\{\alpha | \alpha\mathbf{H} \in A\}$, where μ_1 denotes the Lebesgue measure on the real line and A denotes any Borel subset of V. I show below that $\lambda_V(\mathcal{C} - \mathbf{F}) > 0$ and that $\lambda_V(E_{a,b,\phi,i} - x) = 0$ for all $x \in \mathcal{X}$.

We have

$$\begin{aligned} \lambda_V(\mathcal{C} - \mathbf{F}) &= \mu_1(\{\alpha | \alpha\mathbf{H} \in \mathcal{C} - \mathbf{F}\}) \\ &= \mu_1(\{\alpha | \mathbf{F} + \alpha\mathbf{H} \in \mathcal{C}\}) > 0. \end{aligned} \tag{7.20}$$

The latter inequality holds because for all values of α bounded in absolute value by some $\alpha* > 0$, $\mathbf{F} + \alpha\mathbf{H} \in \mathcal{C}$.[30]

I now show that $\lambda_V(E_{a,b,\phi,i} - x) = 0$ for all $x \in \mathcal{X}$. Let us assume for now that the sets $E_{a,b,\phi,i}$ are Borel measurable; I will show further below that they are indeed closed. I establish the claim by showing that for each $x \in \mathcal{X}$, the set $\{\alpha | \alpha\mathbf{H} \in E_{a,b,\phi,i} - x\}$ has finitely many elements. We have $\{\alpha | \alpha\mathbf{H} \in E_{a,b,\phi,i} - x\} = \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C} \text{ and } rank(G_{S_i}) < N\}$. Suppose that $i = 1$ and $S_1 = [0, b]$ with $b > 0$ (the other cases are similar). The set $A_x := \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C} \text{ and } rank(G_{S_1}) < N\}$ can be decomposed as $A_x = A_{x,1} \cup A_{x,2}$, where

$$A_{x,1} = \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C},\ \mathbf{G}_{S_1} \notin \mathcal{C}\}$$

and

$$A_{x,2} = \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C},\ \mathbf{G}_{S_1} \in \mathcal{C} \text{ and } rank(G_{S_1}) < N\}.$$

Given the definition of the renormalization $\mathbf{G}_{S_1}$ of $\mathbf{G}$ on S_1, α belongs to the first set $A_{x,1}$ if and only if one component of the corresponding $\mathbf{G}_{S_i}$ is the zero function. This occurs when a component of $\mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C}$ does not vary on S_1. Let $x = (x_1, \cdots, x_N)$, I now show that for each $j = 1, \cdots, N$ there is at most one value of α such that $\mathbf{G} = x + \alpha\mathbf{H} \in \mathcal{C}$ and the j^{th} component of $\mathbf{G}$ does not vary on S_1. Suppose, for a contradiction, that there are two such distinct values α_1 and α_2, and that $x_j(t) + \alpha_1 H_j(t) = c_1$ and $x_j(t) + \alpha_2 H_j(t) = c_2$ for all $t \in S_1$ and for some constants c_1 and c_2. Taking the difference of the two quantities and using the fact that $\alpha_1 - \alpha_2 \neq 0$ yields that the component H_j is constant on the interval S_1. This contradicts the fact that by construction H_j is differentiable with a non-zero derivative on S_1. This shows that the set $A_{x,1}$ is finite (at most N elements) and has μ_1 measure zero. It now remains to show that $A_{x,2}$ is finite. This will be achieved by a discretization of

[29] Let h be, for instance, equal to the (component-wise) difference of the re-normalizations (to densities) of $(t^2, t^3, \cdots, t^{N+1})$ and $(t, t^2, \cdots, t^N)$ on the set S_i.

[30] Recall that (by construction) the elements of h are bounded and supported on S_i, and the elements of f are bounded away from zero on S_i. Hence for all α sufficiently small, the elements of $f + \alpha h$ are densities.

the elements of $\mathcal{C}$. Since the components of $\mathbf{H}$ are linearly independent on S_1 (by construction, the derivatives of the components are linearly independent on S_1), there exists a sequence of N distinct points $0 < t_1 < \cdots < t_N < b$ such that the vectors $\{\mathbf{H}(t_i)\}_{i=1}^N$ (with $\mathbf{H}(t_i) := (H_1(t_i), \cdots, H_N(t_i))$) span $\mathbb{R}^N$ [31]. Set $\Delta = \{t_1, \cdots, t_N\}$. Given $\mathbf{O} = (O_1, \cdots, O_N) \in \mathcal{X}$ (not necessarily a vector of CDFs), let $M(\mathbf{O}, \Delta)$ denote the $N+1 \times N$ matrix with $(i,j)^{th}$ element given by

$$[M(\mathbf{O}, \Delta)]_{i,j} = O_j(t_{i-1}) - O_j(t_i), \tag{7.21}$$

where we define $t_0 = 0$ and $t_{N+1} = b$. Note that the matrix $M(\mathbf{H}, \Delta)$ has full column rank, since the span of its rows is equal to the span of the vectors $\{\mathbf{H}(t_i)\}_{i=1}^N$, and that the column rank of $M(\mathbf{O}, \Delta)$ is less than N whenever the components of $\mathbf{O}$, restricted to the set $[0, b]$, are linearly dependent. If $\alpha \in A_{x,2}$, it must be the case that the components of the function $\mathbf{G} = x + \alpha\mathbf{H}$ are linearly *dependent* on S_1. Indeed, because $\mathbf{G}_{S_1}$ belongs to $\mathcal{C}$, it necessarily holds that each component G_j of $\mathbf{G}$ satisfies $G_j(b) \neq 0$, and because $rank(\mathbf{G}_{S_1}) < N$, it must be the case that the components of $\mathbf{G}$ are linearly *dependent* on S_1.[32] Therefore, we have

$$A_{x,2} \subset A_{x,3} := \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \text{ satisfies } rank(M(\mathbf{G}, \Delta)) < N\}. \tag{7.22}$$

It thus remains to show that $A_{x,3}$ is finite. Below, for notational simplicity, I omit the Δ in the notation for $M(\cdot, \Delta)$ and simply write $M(\cdot)$. Let $P(\alpha)$ denote the polynomial function of α defined by $P(\alpha) = det(M(\mathbf{G}(\alpha))^T M(\mathbf{G}(\alpha)))$, where $\mathbf{G}(\alpha) := x + \alpha\mathbf{H}$. I prove the claim by showing that all elements of $A_{x,3}$ must be roots of the polynomial P and I show that the polynomial P is non zero. Indeed, since $M(\cdot)$ is linear in its argument, we have

$$P(\alpha) = det\left(M(x)^T M(x) + \alpha\{M(x)^T M(\mathbf{H}) + M(\mathbf{H})^T M(x)\} + \alpha^2 M(\mathbf{H})^T M(\mathbf{H})\right).$$

Hence P has degree at most $2N$, and $P(\alpha) = \alpha^{2N} Q(1/\alpha)$ (for $\alpha \neq 0$) where Q is the polynomial given by

$$Q(\alpha) = det\left(\alpha^2 M(x)^T M(x) + \alpha\{M(x)^T M(\mathbf{H}) + M(\mathbf{H})^T M(x)\} + M(\mathbf{H})^T M(\mathbf{H})\right).$$

Since $Q(0) = det(M(\mathbf{H})^T M(\mathbf{H})) \neq 0$ (by construction), we conclude that P is non-zero with at most $2N$ roots. Therefore $A_{x,2}$ is a finite set.

I now show, as claimed above, that the set $E_{a,b,\phi,i}$ is closed, by showing that its complement is relatively open in $\mathcal{C}$ (recall that $\mathcal{C}$ is closed in $\mathcal{X}$). Given an element $\mathbf{F} \in \mathcal{C}$ in its complement, it must be the case that the components of $\mathbf{F}$ are linearly independent on S_i, and by using an argument similar to the one above, a set of points Δ in S_i can be found such that the corresponding matrix $M(\mathbf{F}, \Delta)$ (defined as above) has full column rank N. Using the continuity of

[31] By construction $\mathbf{H}(0) = \mathbf{H}(b) = 0$, hence the points t_i must all belong to $(0, b)$

[32] Note that the restriction of the j^{th} components of the latter to the set S_1 is obtained by multiplying the j^{th} component of $\mathbf{G}_{S_1}$ by the value $G_j(b)$.

the function $det(M(\mathbf{F},\Delta)^T M(\mathbf{F},\Delta))$ with respect to $\mathbf{F} \in \mathcal{C}$ (as a function on $\mathcal{X}$), and the fact that $det(M(\mathbf{O},\Delta)^T M(\mathbf{O},\Delta)) > 0$ only if the components of $\mathbf{O}$ are linearly independent, we conclude that a neighborhood of $\mathbf{F}$ in $\mathcal{C}$ belongs to the complement of $E_{a,b,\phi,i}$, and the complement of the set $E_{a,b,\phi,i}$ is relatively open in $\mathcal{C}$.

General case. Let the function ϕ now be a general element of $\mathcal{A}$. Since the elements of $\mathcal{A}$ are polynomials, ϕ has the representation

$$\phi(t) = \sum_{k=1}^{d} a_k t^k, \tag{7.23}$$

for all $t \in [0,1]$,[33] for some finite d (the degree of ϕ) such that $a_d \neq 0$. Let $\mathbf{F} = (F_1, \cdots, F_N) \in \mathcal{C}$ and $\mathbf{H} = (H_1, \cdots, H_N) \in \mathcal{X}$ be defined as in 7.19, with the particular choice of the functions f and h given by 28 and 29. Also, let the subspace $V \subset \mathcal{X}$ and the measure λ_V be the same as above. The same argument used in 7.20 yields $\lambda_V(\mathcal{C} - \mathbf{F}) > 0$. Thus it only remains to prove that $\lambda_V(E_{a,b,\phi,i} - x) = 0$ for all $x \in \mathcal{X}$. Fix $x \in \mathcal{X}$. We have

$$\lambda_V(E_{a,b,\phi,i} - x) = \mu_1(\{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C} \text{ and } rank(\phi \circ G_{S_i}) < N\}).$$

I show below that the set $B_x := \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C} \text{ and } rank(\phi \circ G_{S_i}) < N\}$ has finitely many elements, hence μ_1 measure zero. As above, suppose (without loss of generality) that $i = 1$ and $S_1 = [0, b]$ with $b > 0$ (the other cases are similar). Decompose B_x as $B_x = B_{x,1} \cup B_{x,2}$ where

$$B_{x,1} = \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C},\ \mathbf{G}_{S_1} \notin \mathcal{C}\}$$

and

$$B_{x,2} = \{\alpha | \mathbf{G} := x + \alpha\mathbf{H} \in \mathcal{C},\ \mathbf{G}_{S_1} \in \mathcal{C} \text{ and } rank(\phi \circ G_{S_1}) < N\}.$$

That the set $B_{x,1}$ is finite, follows from the same argument used to show that $A_{x,1}$ is finite. Suppose $B_{x,2}$ is non-empty. Then for $\alpha \in B_{x,2}$ and $\mathbf{G} := x + \alpha\mathbf{H}$, it necessarily holds that all components of $\mathbf{G}$ do not vanish at the point b, and that $G_j(b) = x_j(b) + \alpha H_j(b) = x_j(b) > 0$[34] for $j = 1, \cdots, N$. Also, since each component of $\mathbf{G}$ and $\mathbf{H}$ vanishes at 0, it follows that each component of x also vanishes at 0. Hence for $B_{x,2}$ to be non-empty, it is necessary for all components of x to be strictly positive at b and to all vanish at 0. Let us assume that this is the case. Then for $\alpha \in B_{x,2}$ and $\mathbf{G} := x + \alpha\mathbf{H}$ $\mathbf{G}_{S_1}$ can be written as

$$\mathbf{G}_{S_1} = x' + \alpha\mathbf{H'}$$

where the elements x' and $\mathbf{H'}$ are defined in accordance with the definition of $\mathbf{G}_{S_1}$. That is, the j^{th} component of x' is given by $x'_j(t) = x_j(t)/G_j(b) = x_j(t)/x_j(b)$ for $t \in [0, b]$ and $x'_j(t) =$

[33] Note that the polynomial ϕ has no constant term since $\phi(0) = 0$ (ϕ is a distribution).

[34] Recall that by construction, all components of $\mathbf{H}$ vanish at b (see 7.19).

$x_j(b)/x_j(b) = 1$ for all $t \in [b, 1]$. Similarly, the j^{th} component of $\mathbf{H'}$ is given by $H'_j(t) = H_j(t)/x_j(b)$ for $t \in [0, b]$ and $H'_j(t) = H_j(b)/x_j(b) = 0$ (see 7.19) for all $t \in [b, 1]$. Note that both x' and $\mathbf{H'}$ do not depend on α, and that by construction (see 29 and 7.19), the components of $\mathbf{H'}$ (restricted to $[0, b]$) are equal to polynomials of different degree. Hence given any integer $k \geq 1$, the components of $(\mathbf{H'})^k$ are linearly independent,[35] as their restriction on $[0, b]$ is also given by polynomials of different degrees. In particular, when the exponent is equal to the degree d of ϕ, (see 7.23) the components of $(\mathbf{H'})^d$ are linearly independent. A similar argument to the one preceding 7.22 implies that there exists a set $\Delta = \{t_1, \cdots, t_N\}$ of points in $[0, b]$ (not necessarily the same points as before) such that the matrix $M((\mathbf{H'})^d, \Delta)$ has full column rank, with the transformation $M(\cdot, \Delta)$ defined as in 7.21, and an argument similar to the one used to establish 7.22 implies that

$$B_{x,2} \subset B_{x,3} := \{\alpha | \mathbf{G} := x' + \alpha \mathbf{H'} \text{ satisfies } \ rank(M(\phi \circ \mathbf{G}, \Delta)) < N\}.$$

Using the linearity of the transformation $M(\cdot, \Delta)$ and the representation 7.23 of ϕ, it is easy to show that

$$M(\phi \circ \mathbf{G}, \Delta)) = a_d \alpha^d M((\mathbf{H'})^d, \Delta) + R(\alpha),$$

where $R(\alpha)$ is a polynomial (with $N + 1 \times N$ matrix coefficients) of degree at most $d - 1$ in α; that is:

$$R(\alpha) = \sum_{k=0}^{d-1} \alpha^k M_k$$

where M_k, for $k = 0, \cdots, d - 1$, are $N + 1 \times N$ matrices. Below, for notational simplicity, I write $M(\cdot)$ for $M(\cdot, \Delta)$. Let the polynomial P be defined by

$$\begin{aligned} P(\alpha) &= det\left(\{a_d \alpha^d M((\mathbf{H'})^d) + R(\alpha)\}^T \{a_d \alpha^d M((\mathbf{H'})^d) + R(\alpha)\}\right) \\ &= det\left(a_d^2 \alpha^{2d} M((\mathbf{H'})^d)^T M((\mathbf{H'})^d) + \tilde{R}(\alpha)\right) \end{aligned}$$

for a polynomial $\tilde{R}$ with $N \times N$ matrix coefficients of degree at most $2d - 1$ in α. The polynomial P has degree at most $2dN$, and since $det\left(M((\mathbf{H'})^d)^T M((\mathbf{H'})^d)\right) \neq 0$, an argument similar to the one used in the *simple case* implies that P has at most $2dN$ roots. Hence the set $B_{x,3}$ is finite, and this concludes the argument. □

Proof of 3.23. Corollary 3.23 is an easy consequence of 3.22, 7.1 and 7.2. □

7.2.5 Proof of Theorem 3.29

Proof. By considering the rank of matrices formed by partitioning the domain of $[B^{(i_1)} | B^{(i_2)} = a]$ and Z an argument similar to that used in the proof of Theorem 3.10 shows that N (the number of support points of the unobserved heterogeneity) is identified. As in the proof of 3.20, let $\Phi^1 =$

[35] Given $\mathbf{F} = (F_1, \cdots, F_N) \in \mathcal{X}$ and an integer $k \geq 1$, the k^{th} power of $\mathbf{F}$ is given by $\mathbf{F}^k = (F_1^k, \cdots, F_N^k)$.

$(\phi^1_1, \cdots, \phi^1_N)$, $\Phi^z = (\phi^z_1, \cdots, \phi^z_N)$ and $\Phi^3 = (\phi^3_1, \cdots, \phi^3_M)$ (with $M \geq N$) be functions defined respectively on the support of $[B^{(i_1)}|B^{(i_2)} = a]$, Z and $[B^{(i_3)}|B^{(i_2)} = a]$, and let Q^1, Q^z and Q^3 denote the associated matrices (where we now exclude I from the conditioning variables). As in the proof of 3.20, by Assumption 3.26 the functions Φ^1, Φ^3 and Φ^z can be chosen such that the matrices Q^1 and Q^z are invertible and the matrix Q^3 has distinct columns. Let the matrices $\{A^j\}_{j=0}^M$ be defined as in equations 7.14 and 7.15 (without conditioning on I). Using the law of iterated expectation (conditioning on U and I), equations 3.19 and 3.20, and the Markov property of order statistics, equations 7.16 and 7.17 can be shown to hold. A similar argument to the one preceding equation 7.17 then shows that Q^z is identified. An argument similar to the one used in the paragraph following 7.10 in the proof of Theorem 3.9 shows that the distribution of U is identified. By considering the identified expectations $E\mathbf{1}\{B^{(i_1)} \leq s_1, B^{(i_2)} \leq s_2, B^{(i_3)} \leq s_3\}\Phi^z(Z)$ for different values of s_1, s_2 and s_3 in $\mathbb{R}$, an argument similar to 7.12 shows that the distributions $\{F_{B^{(i_1)},B^{(i_2)},B^{(i_3)}|U=n}\}_{n=1}^N$ are identified. The observation of Song (2004) (see 3.19) can now be used to identify the distribution $F_{B|U=n}$ by considering the distribution of $[B^{(i_1)}|B^{(i_2)} = a, U = n]$ for values of a near the lower bound of the support of $B^{(i_2)}$. $\square$

7.3 Omitted proofs of section 4

I begin by establishing the large sample theory for the estimators $\{\hat{A}^j\}_{j=0}^N$ defined in 4.1 and 4.2. Let v^0 be defined by

$$v^0(b_1, b_3) = vec(\Phi^1(b_1) \otimes \Phi^3(b_3)),$$

and for $j = 1, \cdots, N$, define v^j by

$$v^j(b_1, b_3, s) = vec(\Phi^1(b_1) \otimes \Phi^3(b_3)\Phi^z_j(s)).$$

Also, let A denote the $N \times N^2$ matrix obtained by horizontally concatenating the matrices $\{A^j\}_{j=0}^N$, i.e $A = (A^0, A^1, \cdots, A^N)$, and let $\hat{A}$ be defined analogously.

By the continuity of the distributions of $[B^{(i_1)}|B^{(i_2)} = a]$ and $[B^{(i_3)}|B^{(i_2)} = a]$, and that of the determinant functional (on $N \times N$ matrices), for any sufficiently small $c > 0$, the components of Φ^1 and Φ^3 can be (respectively) chosen as indicator functions of partitions of the sets

$$supp\left(B^{(i_1)}|B^{(i_2)} = a\right) \cap \{B^{i_1} \geq a + c\} \text{ and } supp\left(B^{(i_3)}|B^{(i_2)} = a\right) \cap \{B^{i_3} \leq a - c\}, \tag{7.24}$$

in such a way that the matrices Q^1 and Q^3 that appear in the proof of 3.9 remain invertible. The reason for "trimming" the supports of $[B^{(i_1)}|B^{(i_2)} = a]$ and $[B^{(i_3)}|B^{(i_2)} = a]$ will be made clear in the proof of Lemma 7.5 below, where it will be used to restrict the Taylor expansion of the density $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)}}$ to a region where the latter is twice continuously differentiable. This trimming is not needed if the order statistics are well separated, i.e, $\min\{i_2 - i_1, i_3 - i_2\} \geq 3$, as this will imply that the density $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)}}$ is differentiable at the boundary of the region $\{(b_1, b_3)|b_3 < a < b_1\}$, where $[B^{(i_1)}, B^{(i_3)}|B^{(i_2)} = a]$ is supported.

The following lemma establishes the asymptotic distribution of $vec(\hat{A}^0)$ and $vec(\hat{A})$. For notational convenience, I will sometimes omit the subscripts that appear in the expressions for the joint densities $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)}}$ and $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)},Z}$, and I use the symbol ∂_x to denote the partial derivative operator with respect to the variable x.

Lemma 7.5. *Suppose that $\{f_{B|U=k}\}_{k=1}^N$ are twice continuously differentiable on their supports, and let $K(\cdot)$ be a compactly supported bounded symmetric kernel of order 2. Then provided that $Th \to \infty$ and $Th^5 \to c$ ($c \geq 0$), we have*

$$\sqrt{Th}\{vec(\hat{A}^0) - vec(A^0)\} \xrightarrow{d} N(\mu^0, \Sigma^0)$$

and

$$\sqrt{Th}\{vec(\hat{A}) - vec(A)\} \xrightarrow{d} N(\mu, \Sigma)$$

where

$$\mu^0 = \frac{c^{1/2}}{2} \int u^2 K(u)du \int v^0(b_1, b_3)\partial^2_{b_2} f(b_1, a, b_3)db_1 db_3,$$
$$\Sigma^0 = \int K(u)^2 du \int v^0(b_1, b_3) \otimes v^0(b_1, b_3) f(b_1, a, b_3)db_1 db_3,$$

and

$$\mu = \frac{c^{1/2}}{2} \int u^2 K(u)du \int v(b_1, b_3, s)\partial^2_{b_2} f(b_1, a, b_3, s)db_1 db_3 ds,$$
$$\Sigma = \int K(u)^2 du \int v(b_1, b_3, s) \otimes v(b_1, b_3, s) f(b_1, a, b_3, s)db_1 db_3 ds$$

where v is the vector obtained by vertically concatenating the vectors $\{v^j\}_{j=0}^N$, i.e,
$v = ((v^0)^T, \cdots, (v^N)^T)^T$.

Proof. I give below the derivation for the asymptotic distribution of $\sqrt{Th}\{vec(\hat{A}^0) - vec(A^0)\}$; a similar argument can be applied to derive the asymptotic distribution of $vec(\hat{A})$. By a standard argument, the bias of the estimator $vec(\hat{A}^0)$ and the variance of $\sqrt{Th}vec(\hat{A}^0)$ can be shown to respectively satisfy

$$\lim_{h\to 0} \frac{Evec(\hat{A}^0 - A^0)}{h^2} = \frac{1}{2} \int u^2 K(u)du \int v^0(b_1, b_3)\partial^2_{b_2} f(b_1, a, b_3)db_1 db_3,$$

and

$$\lim_{h\to 0} var(\sqrt{Th}vec(\hat{A}^0)) = \int K(u)^2 du \int v^0(b_1, b_3) \otimes v^0(b_1, b_3) f(b_1, a, b_3)db_1 db_3$$

where I have used the fact that the (twice) differentiability of $\{f_{B|U=k}\}_{k=1}^N$ implies that of $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)}}$ with respect to its second argument [36] in the region $\{(b_1, b_3)|b_1 < a < b_3\}$, and a second order

[36]Recall that if $B^{(i_1)}$, $B^{(i_2)}$ and $B^{(i_3)}$ are three order statistics out of I (i.i.d) draws from a distribution F with density f, then the joint density of $\left(B^{(i_1)}, B^{(i_2)}, B^{(i_3)}\right)$ is given by: $f_{B^{(i_1)},B^{(i_2)},B^{(i_3)}}(b_1, b_2, b_3) \propto \mathbf{1}\{b_1 > b_2 > b_3\}[1 - F(b_1)]^{i_1 - 1}[F(b_1) - F(b_2)]^{i_2 - i_1 - 1}[F(b_2) - F(b_3)]^{i_3 - i_2 - 1}F(b_3)^{I - i_3} f(b_1)f(b_2)f(b_3)$ (see Aron and Navada (2003)).

Taylor series expansion of $f(b_1, a+uh, b_3)$ is valid for all u in the support of $K(\cdot)$, for all h sufficiently small, and for all (b_1, b_3) in the region $\{(b_1, b_3)|b_3 + c < a < b_1 - c\}$ (see 7.24) where the function $v^0(b_1, b_3)$ is non-zero. Since $\sqrt{Th}\{vec(\hat{A}^0) - vec(A^0)\} = \sqrt{Th}\{vec(\hat{A}^0) - Evec(\hat{A}^0)\} + \sqrt{Th}\{Evec(\hat{A}^0) - vec(A^0)\}$, and

$$\sqrt{Th}\{vec(\hat{A}^0) - Evec(\hat{A}^0)\} = \sum_{t=1}^{T} \frac{X_{iT} - EX_{iT}}{\sqrt{hT}}$$

where $X_{iT} = v^0(B_t^{(i_1)}, B_t^{(i_3)})K\left(\frac{a - B_t^{(i_2)}}{h}\right)$ (note the dependence of h on T), it remains to show that the Lyapunov condition holds to obtain the conclusion of the lemma from the Lindeberg-Feller version of the central limit theorem. The latter follows if I show that

$$\lim_{T\to\infty} \sum_{t=1}^{T} E\|\frac{X_{iT} - EX_{iT}}{\sqrt{hT}}\|^4 = 0$$

where $\|\cdot\|$ denotes the Euclidian norm. By a simple argument (that relies on the boundedness of the kernel $K(\cdot)$ and of the components of the functions Φ^1 and Φ^3) one can show that

$$\begin{aligned}\sum_{t=1}^{T} E\|\frac{X_{iT} - EX_{iT}}{\sqrt{hT}}\|^4 &\leq 8\{\sum_{t=1}^{T} E\|\frac{X_{iT}}{\sqrt{hT}}\|^4 + \sum_{t=1}^{T} E\|\frac{EX_{iT}}{\sqrt{hT}}\|^4\} \\ &= 8\{E\frac{\|X_{1T}\|^4}{h^2T} + \frac{\|EX_{iT}\|^4}{h^2T}\} \\ &= O(\frac{1}{hT}) + O(\frac{h^2}{T})\end{aligned}$$

and the latter expression converges to zero if $hT \to \infty$. $\square$

From the identification argument in the proof of Theorem 3.9, the matrices D^j, $j = 1, \cdots, N$ are identified by $D^j = (Q^1)^{-1}C^jQ^1$, and Q^1 is the unique (up to permutation of its columns) probability matrix that simultaneously diagonalizes the matrices $\{C^j\}_{j=1}^N$. Analogously to Bonhomme, Jochmans, and Robin (2016), this suggests the estimator

$$\hat{D}^j = diag(\hat{Q}^{-1}\hat{C}^j\hat{Q})$$

where $\hat{C}^j = \hat{A}^j(\hat{A}^0)^{-1}$ are sample analogues of the matrices C^j, and $\hat{Q}$ is a probability matrix that solves

$$\min_{Q} \sum_{j=1}^{N} \|off(Q^{-1}\hat{C}^jQ)\|_F^2,$$

In the last expression, the minimization is done over the set of all invertible probability matrices. Although the restriction of the minimization problem to the set of probability matrices resolves some of the indeterminacies that are present in Bonhomme, Jochmans, and Robin (2016)[37], I will

[37] In Bonhomme, Jochmans, and Robin (2016), the minimization is done over the set of all invertible matrices with all columns having norm equal to one. The minimizing matrix is then unique up to a permutation and direction change (multiplication of a column by -1 still produces a "minimal" matrix) of its columns. The restriction to probability matrices resolves the indeterminacy with respect to direction change.

ignore that restriction when I estimate Q, and consider instead the whole set of all invertible matrices with all columns having unit norm. This will allow me to directly apply the results in Bonhomme, Jochmans, and Robin (2016) in deriving the asymptotic distribution of $\hat{Q}$. As is shown below, the asymptotic distribution of the estimators $\hat{D} = (\hat{D}^1, \cdots, \hat{D}^N)$ and $\hat{Q}$ depends on that of $\hat{C}^j$. Let $C = (C^1, \cdots, C^N)$ and $\hat{C} = (\hat{C}^1, \cdots, \hat{C}^N)$ denote respectively the horizontal concatenations of the matrices C^j and $\hat{C}^j$.

The following lemma establishes the asymptotic distribution of $vec(\hat{C})$.

Lemma 7.6. *Under the assumptions of Lemma 7.5 , the estimator $\hat{C}$ has the asymptotic distribution*

$$\sqrt{hT}vec(\hat{C} - C) \xrightarrow{d} N(F\mu, \Sigma_C)$$

where $\Sigma_C = F\Sigma F^T$ and F is given by

$$F = \sum_{j=1}^{N} -\{e_j^N \otimes e_1^{N+1}\} \otimes_K \{(A^0)^{-T} \otimes_K C^j\} + \{e_j^N \otimes e_{j+1}^{N+1}\} \otimes_K \{(A^0)^{-T} \otimes_K I_N\}.$$

Proof. Since $vec(C)$ is differentiable function of the elements of $vec(A)$, the asymptotic result follows from an application of the delta method to $vec(\hat{C} - C)$. Recall that given a $N \times N$ matrix B with operator norm less than 1, the inverse of $I_N - B$ has the series expansion

$$(I_N - B)^{-1} = \sum_{k=0}^{\infty} B^k.$$

The last relation and Lemma 7.5 imply that

$$(\hat{A}^0)^{-1} - (A^0)^{-1} = -(A^0)^{-1}(\hat{A}^0 - A^0)(A^0)^{-1} + O_p(\frac{1}{Th}).$$

It thus follows that

$$\begin{aligned} \hat{C}^j - C^j &= \hat{A}^j(\hat{A}^0)^{-1} - A^j(A^0)^{-1} \\ &= A^j\{(\hat{A}^0)^{-1} - (A^0)^{-1}\} + \{\hat{A}^j - A^j\}(A^0)^{-1} + O_p(\frac{1}{Th}) \\ &= -A^j(A^0)^{-1}(\hat{A}^0 - A^0)(A^0)^{-1} + \{\hat{A}^j - A^j\}(A^0)^{-1} + O_p(\frac{1}{Th}). \end{aligned} \tag{7.25}$$

Using the relation $vec(USW) = (W^T \otimes_K U)vec(S)$, which holds for any three matrices U, S and W such that the product USW is well-defined (see equation 2.1 in Magnus and Neudecker (1979)), equation 7.25 becomes

$$vec(\hat{C}^j - C^j) = -\{(A^0)^{-T} \otimes_K C^j\}vec(\hat{A}^0 - A^0) + \{(A^0)^{-T} \otimes_K I_N\}vec(\hat{A}^j - A^j) + O_p(\frac{1}{Th})$$

which yields

$$vec(\hat{C} - C) = Fvec(\hat{A} - A) + O_p(\frac{1}{Th})$$

where H is as in the statement of the lemma, and the result follows from the asymptotic normality of $\hat{A}$ given in Lemma 7.5. $\square$

I now provide the asymptotic distribution of $\hat{Q}$ and $\hat{D}$. The lemma below follows directly from Theorem 5 and 6 in Bonhomme, Jochmans, and Robin (2016) (only a slight modification is needed to adjust for the nonparametric rates that arise from Lemma 7.5) so I omit the proof. Below, given two square matrices A and B with respective dimensions m and n, I use the notation $\ominus$ to denote the Kronecker difference defined by

$$A \ominus B = A \otimes_K I_n - I_m \otimes_K B.$$

I denote by T the $N^2 \times N^3$ matrix defined by

$$T = ((D^1 \ominus D^1), \cdots, (D^N \ominus D^N)).$$

I use the matrix $S_N \equiv diag(vecI_N)$ to denote the $N^2 \times N^2$ selection matrix; note that $S_N vec(Q) = vec(diagQ)$.

Lemma 7.7. *Under the assumptions of Lemma 7.5, the estimators $\hat{Q}$ and $\hat{D}$ have the asymptotic distributions*

$$\sqrt{hT} vec(\hat{Q} - Q) \xrightarrow{d} N(GF\mu, \Sigma_Q) \tag{7.26}$$

and

$$\sqrt{hT} vec(\hat{D} - D) \xrightarrow{d} N(HF\mu, \Sigma_D) \tag{7.27}$$

where $\Sigma_Q = G\Sigma_C G^T$, $\Sigma_D = H\Sigma_C H^T$, Σ_C is as in Lemma 7.6, and the matrices G and H are given by

$$G = (I_N \otimes_N Q) \left(\sum_{j=1}^{N} (D^j \ominus D^j)^2 \right)^+ T(I_N \otimes_K Q^T \otimes_K Q^{-1})$$

and

$$H = (I_N \otimes_K S_N)(I_N \otimes_K Q^T \otimes_K Q^{-1}).$$

Proof. See Theorem 5 and 6 in Bonhomme, Jochmans, and Robin (2016). □

Recall from the proof of Theorem 3.9 that the matrix M is defined as the matrix with k^{th} row given by the diagonal elements of D^k. I derive below the asymptotic distribution of $\hat{M}$ the natural estimator for M, obtained by collecting the diagonal entries of $\hat{D}^j$. Let the $N^2 \times N^3$ matrix L be defined by

$$L = I_N \otimes_K \left(\sum_{j=1}^{N} e_j^N \otimes_K (e_j^N \otimes_K e_j^N)^T \right)$$

and note that $Lvec(D) = vec(M)$. The following corollary is an easy consequence of Lemma 7.7

Corollary 7.8. *Under the assumptions of Lemma 7.5, the asymptotic distribution of $\hat{M}$ is given by*

$$\sqrt{hT} vec(\hat{M} - M) \xrightarrow{d} N(LHF\mu, \Sigma_M) \tag{7.28}$$

where $\Sigma_M = L\Sigma_D L^T$, and all other matrices are as in Lemma 7.7

I now turn to the derivation of the asymptotic distribution of the mixture weights.

Proof of Theorem 4.2. A linearization of the expression $\hat{\delta} - \delta$ yields

$$\begin{aligned}
\hat{\delta} - \delta &= \hat{M}^{-1}\hat{d} - M^{-1}d \\
&= (\hat{M}^{-1} - M^{-1})d + M^{-1}(\hat{d} - d) + o_p\left(\frac{1}{\sqrt{hT}}\right) \\
&= -M^{-1}(\hat{M} - M)M^{-1}d + M^{-1}(\hat{d} - d) + o_p\left(\frac{1}{\sqrt{hT}}\right)
\end{aligned}$$

where I have used the fact that $\hat{d}$ is root-n consistent and that (see proof of Lemma 7.6)

$$\hat{M}^{-1} - M^{-1} = -M^{-1}(\hat{M} - M)M^{-1}d + o_p\left(\frac{1}{\sqrt{hT}}\right).$$

It thus follows that

$$\sqrt{hT}(\hat{\delta} - \delta) = -\sqrt{hT}\left(M^{-1}(\hat{M} - M)M^{-1}d\right) + o_p(1) \tag{7.29}$$

and the conclusion of the theorem follows from relation 2.1 in Magnus and Neudecker (1979). □

I now turn to the derivation of the asymptotic distribution of the estimated conditional distributions of order statistics $\hat{x}^p$. I begin by establishing the weak convergence of the process $\hat{y}^p(s)$, $s \in \mathbb{R}$.

Lemma 7.9. *For a fixed $a \in \mathbb{R}^N$, the process $\{a^T\hat{y}^p(s) | s \in \mathbb{R}\}$ is distributed asymptotically as*

$$\sqrt{T}\left(a^T\hat{y}^p(s) - a^Ty^p(s)\right) \rightsquigarrow \boldsymbol{G}_{y^p} \tag{7.30}$$

where $\boldsymbol{G}_{y^p}$ is a tight centered Gaussian process with covariance function given by

$$cov(\boldsymbol{G}_{y^p}(s), \boldsymbol{G}_{y^p}(s')) = cov(\mathbf{1}\{B^{(i_p)} \leq s\}a^T\Phi^z(Z), \mathbf{1}\{B^{(i_p)} \leq s'\}a^T\Phi^z(Z))$$

for all s and $s' \in \mathbb{R}$.

Proof. For $s \in \mathbb{R}$, let g_s be the function defined by

$$g_s(b,z) = \mathbf{1}\{b \leq s\}a^T\Phi^z(z),$$

and let $\mathcal{F} = \{g_s | s \in \mathbb{R}\}$ denote the class of all such functions. Since the collection of all unbounded and connected intervals on $\mathbb{R}$ is a VC-class, one can easily show that $\mathcal{F}$ is a VC-subgraph class. Moreover, all the functions in $\mathcal{F}$ are dominated by the largest element (in absolute value) of the vector a (recall that all elements of Φ^z are indicator functions of sets, and thus less than one in absolute value). The class $\mathcal{F}$ is thus $P-$Donsker (see Theorem 2.6.8 in Vaart and Wellner (1996)) and the conclusion of the lemma follows. □

I now turn to the proof of Theorem 4.4

Proof of Theorem 4.4. It follows from Lemma 7.9 (and a linearization similar to the one in the proof of Theorem 4.2) that

$$\begin{aligned}\sqrt{hT}(\hat{x}^{(p)}(s) - x^{(p)}(s)) = &-\sqrt{hT}\Delta^{-1}(\hat{\Delta} - \Delta)\Delta^{-1}M^{-1}y^p(s) \\ &- \sqrt{hT}\Delta^{-1}M^{-1}(\hat{M} - M)M^{-1}y^p(s) + o_p(1).\end{aligned}$$

where the error in approximation is uniform in $s \in \mathbb{R}$. The claim of the theorem then follows from relation 2.1 in Magnus and Neudecker (1979), and the fact that

$$\begin{aligned}\sqrt{hT}vec(\hat{\Delta} - \Delta) &= \left(\sum_{j=1}^{N}(e_j^N \otimes_K e_j^N) \otimes_K (e_j^N)^T\right)\sqrt{hT}(\hat{\delta} - \delta) \\ &= \left(\sum_{j=1}^{N}(e_j^N \otimes_K e_j^N) \otimes_K (e_j^N)^T\right)R\sqrt{hT}vec(\hat{M} - M) + o_p(1)\end{aligned}$$

where the last equality follows from equation 7.29 and the definition of R. □

7.4 Expression for covariance matrix in Theorem 4.2

This section contains all the key matrices involved in the computation of the covariance matrix Σ_M appearing in 4.2 (and 7.8). The matrices are listed in the order in which they are needed for the computation of Σ_M, with each matrix being only a function of prior matrices in the list. The other matrices (A^j, C^j and Q) involved in these expressions are as defined in the preceding section. Let the covariance matrix Σ be as in 7.5

$$F = \sum_{j=1}^{N} -\{e_j^N \otimes e_1^{N+1}\} \otimes_K \{(A^0)^{-T} \otimes_K C^j\} + \{e_j^N \otimes e_{j+1}^{N+1}\} \otimes_K \{(A^0)^{-T} \otimes_K I_N\}$$

$$\Sigma_C = F\Sigma F^T$$

$$S_N \equiv diag(vecI_N)$$

$$H = (I_N \otimes_K S_N)(I_N \otimes_K Q^T \otimes_K Q^{-1})$$

$$\Sigma_D = H\Sigma_C H^T$$

$$L = I_N \otimes_K \left(\sum_{j=1}^{N} e_j^N \otimes_K (e_j^N \otimes_K e_j^N)^T\right)$$

and

$$\Sigma_M = L\Sigma_D L^T.$$