

Nonparametric Estimation and Testing of the Symmetric IPV Framework with Unknown Number of Bidders

Kyoo il Kim*

Joonsuk Lee

Sungkyunkwan University

University of Virginia School of Law

May 2014

Abstract

Exploiting rich data on ascending auctions we estimate the distribution of bidders' valuations nonparametrically under the symmetric independent private values (IPV) framework and test the validity of the IPV assumption. We show the testability of the symmetric IPV when the number of potential bidders is unknown. We then develop a nonparametric test of such valuation paradigm. Unlike previous work on ascending auctions, our estimation and testing methods use more information from observed losing bids by virtue of the rich structure of our data. We find that the hypothesis of IPV is supported with our data after controlling for observed auction heterogeneity.

Keywords: Used-car auctions, Ascending auctions, Unknown number of bidders, Nonparametric testing of valuations, SNP density estimation

JEL Classification: L1, L8, C1

*We are grateful to Dan Akerberg, Pat Bajari, Sushil Bikhchandani, Jinyong Hahn, Joe Ostroy, John Riley, Jean-Laurent Rosenthal, Unjy Song, Mo Xiao and Bill Zame for helpful comments. Special thanks to Ho-Soon Hwang and Hyun-Do Shin for allowing us to access the invaluable data. All remaining errors are ours. This paper was supported by Center for Economic Research of Korea (CERK) by Sungkyunkwan University (SKKU) and Korea Economic Research Institute (KERI). Please note that an earlier version of this paper had been circulated with the title "A Structural Analysis of Wholesale Used-car Auctions: Nonparametric Estimation and Testing of Dealers' Valuations with Unknown Number of Bidders". We changed it to the current title following a referee's suggestion. Corresponding author: Kyoo il Kim, Department of Economics, University of Minnesota, 4-129 Hanson Hall, 1925 Fourth Street South, MN 55455, Tel: +1 612 625 6793, E-mail: kyookim@umn.edu.

1 Introduction

Ascending-price auctions, or English auctions, are one of the oldest trading mechanisms that have been used for selling a variety of goods ranging from fish to airwaves. Wholesale used-car markets are among those that have utilized ascending auctions for trading inventories among dealers. In this paper, we exploit a new, rich data set on a wholesale used-car auction to study the latent demand structure of the auction with a structural econometric model. The structural approach to study auctions assumes that bidders use equilibrium bidding strategies, which are predicted from game-theoretic models, and tries to recover bidders' private information from observables. The most fundamental issue of structural auction models is estimating the unobserved distribution of bidders' willingness to pay from the observed bidding data.¹

We treat our data as a collection of independent, single-object, ascending auctions. So we are not studying any feature of multi-object or multi-unit auctions in this paper though there might be some substitutability and path-dependency in the auctions we study. To reduce the complexity of our analysis, we defer those aspects to future study. Also, to deal with the heterogeneity in the objects of our auctions, we first assume there is no unobserved heterogeneity in our data and then control for the observed heterogeneity in the estimation stage (See e.g. Krasnokutskaya (2011) and Hu, McAdams, and Shum (2013) for a novel nonparametric approach regarding unobserved auction heterogeneity).

In our model, there are symmetric, risk-neutral bidders while the number of potential bidders is unknown. Game-theoretic models of bidders' valuations can be classified according to informational structures.² We first assume that the information structure of our model follows the symmetric independent private-values (hereafter IPV) paradigm and then we develop a new nonparametric test with unknown number of potential bidders to check the validity of the symmetric IPV assumption

¹Paarsch (1992) first conducted a structural analysis of auction models. He adopted a parametric approach to distinguish between independent private-values and pure common-values models. For the first-price sealed-bid auctions, Guerre, Perrigne, Vuong (2000) provided an influential nonparametric identification and estimation results, which was extended by Li, Perrigne, and Vuong (2000, 2002), Campo, Perrigne, and Vuong (2003), and Guerre, Perrigne, Vuong (2009).

See Hendricks and Paarsch (1995) and Laffont (1997) for surveys of the early empirical works. For a survey of more recent works see Athey and Haile (2005).

²Valuations can be either from the private-values paradigm or the common-values paradigm. Private-values are the cases where a bidder's valuation depends only on her own private information while common-values refer to all the other general cases. IPV is a special case of private-values where all bidders have independent private information.

after we obtain our estimates of the distribution under the IPV paradigm. This extends the testability of the symmetric IPV of Athey and Haile (2002) that assumes the number of potential bidders is known to the case when the number is unknown. We find that the null hypothesis of symmetric IPV is supported in our sample and therefore our estimation result under the assumption remains a valid approximation of the underlying distribution of dealers' valuations.

To deal with the problem of unknown number of potential bidders, we build on the methodology provided in Song (2005) for the nonparametric identification and estimation of the distribution when the number of potential bidders of ascending auctions is unknown in an IPV setting. Unlike previous work on ascending auctions, the rich structure of our data, especially the absence of jump bidding, enables us to utilize more information from observed losing bids for estimation and testing. This is an important advantage of our data compared to any other existing studies of ascending auctions. Our key idea is that because any pair of order statistics can identify the value distribution, three or more order statistics can identify two or more value distributions and under the symmetric IPV, these value distributions should be identical. Therefore, testing the symmetric IPV is equivalent to testing the similarity of value distributions obtained from different pairs of order statistics.

There has been considerable research on the first-price sealed-bid auctions in the structural analysis literature; however, there are not as many on ascending auctions. One of the possible reasons for this scarcity is the discrepancy between the theoretical model of ascending auctions, especially the model of Milgrom and Weber (1982), and the way real-world ascending auctions are conducted. Another important reason preventing structural analysis of ascending auctions is the difficulty of getting a rich and complete data set that records ascending auctions. Given these difficulties, our paper contributes to the empirical auctions literature by providing a new data set that is free from both problems and by conducting a sound structural analysis.

Milgrom and Weber (1982) (hereafter MW) described ascending auctions as a button auction, an auction with full observability of bidders' actions and, more importantly, with the irrevocable exits assumption, an assumption that a bidder is not allowed to place a bid at any higher price once she drops out at a lower price. This assumption significantly restricts each bidder's strategy space and makes the auction game easier to analyze. Without this assumption for modeling ascending auctions one would have to consider dynamic features allowing bidders to update their information

and therefore valuations continuously and to re-optimize during an auction. After MW, the button auction assumption was widely accepted by almost all the following studies on ascending auctions, both theoretical and empirical, because of its elegant simplicity.³

However, in almost all the real-world ascending auctions, we do not observe irrevocable exits. Haile and Tamer (2003) (hereafter HT) conducted an empirical study of ascending auctions without specifying such details as irrevocable exits. In their nonparametric analysis of ascending auctions with IPV, HT adopted an incomplete modelling approach with relaxing the button auction assumption and imposing only two axiomatic assumptions on bidding behavior. The first assumption of HT is that bidders do not bid more than they are willing to pay and their second assumption is that bidders do not allow an opponent to win at a price they are able to beat. With these two assumptions and known statistical properties of order statistics, HT estimated upper and lower bounds of the underlying distribution function.

The reason HT could only estimate bounds and could not make an exact interpretation of losing bids is because they relaxed the button auction assumption and allowed free forms of ascending auctions. Athey and Haile (2002) noted the difficulty of interpreting losing bids saying “*In oral ‘open outcry’ auctions we may lack confidence in the interpretation of losing bids below the transaction price even when they are observed.*”⁴ Our main difference from HT is that we are able to provide a stronger interpretation of observed losing bids. Within ascending auctions, there are a few variants that differ in the exact way the auction is conducted. Among those variants, the distinction between one in which bidders call prices and one in which an auctioneer raises prices has very important theoretical and empirical implications. The former is what Athey and Haile called as “open outcry” auctions and the latter is what we are going to exploit with our data. The main difference is that the former allows jumps in prices but the latter does not allow those jumps.

It is well known in the literature that it is empirically difficult to distinguish between private-values and common-values with actual auction data. However, there have been a few attempts to develop these tests. Hong and Shum (2003) estimated and tested general private-values and

³There are few exceptions, e.g. Harstad and Rothkopf (2000) and Izmalkov (2003) in the theoretical literature and Haile and Tamer (2003) in the empirical literature. Also see Bikhchandani and Riley (1991, 1993) for extensive discussions of modeling ascending auctions.

⁴Bikhchandani, Haile, and Riley (2002) also show there are generally multiple equilibria even with symmetry in ascending auctions. Therefore one should be careful in interpreting observed bidding data; however, they show that with private values and weak dominance, there exists uniqueness.

common-values models in ascending auctions with a certain parametric modeling assumption using a quantile estimation method. Hendricks, Pinkse, and Porter (2003) developed a test based on data of the winning bids and the ex-post values in first-price sealed-bid auctions. Athey and Levin (2001) also used an ex-post data to test the existence of common values in first-price sealed-bid auctions. Haile, Hong, and Shum (2003) developed a new nonparametric test for common-values in first-price sealed-bid auctions using Guerre, Perrigne, Vuong (2000)'s two-stage estimation method but their test requires exogenous variations in number of potential bidders.

Our paper contributes to this literature by showing the testability of the symmetric IPV even when the numbers of potential bidders are unknown and they can also vary across different auctions. Then, we develop a formal nonparametric test of the symmetric IPV by comparing distributions of valuations implied by two or more pairs of order statistics.

In a closely related paper to ours, Roberts (2013) proposes a novel approach of using variation in reserve prices to control for the unobserved heterogeneity. He uses a control variate method as Matzkin (2003) and recovers the unobserved heterogeneity in an auxiliary step under a set of identification conditions including (i) the unobserved heterogeneity is uni-variate and independent of observed heterogeneity, (ii) the reserve pricing function is strictly increasing in the unobserved heterogeneity, and (iii) the same reserve pricing function is used for all auctions. These assumptions, however, can be arguably restrictive. He then recovers the distribution of valuations via the inverse mapping of the distribution of order statistics, being conditioned on both the observed and unobserved heterogeneity (which is recovered in the auxiliary step, so is observed). In this step he assumes the number of observed bidders is identical to the number of potential bidders. As comparison to Roberts (2013), in our approach we allow that the number of observed bidders can differ from that of potential bidders. But compared to his we make a potentially restrictive assumption that the reserve prices can be approximated by some deterministic functions of the observed auction heterogeneity (while the functions can differ across auctions). This assumption allows us to sufficiently control the auction heterogeneity only by the observed one. Lastly our focus is more on testing of the symmetric IPV framework without parametric assumptions.

The remainder of this paper is organized as follows. In the next section we provide some background on the wholesale used-car auctions for which we develop our empirical model and

nonparametric testing. We then describe the data and give summary statistics in Section 3. Section 4 and 5 present the empirical model and our empirical strategy that includes identification and estimation. We develop our new nonparametric test in Section 6. We provide the empirical results in Section 7 and Section 8 discusses the implications and extensions of the results. Section 9 concludes. Mathematical proofs and technical details are gathered in Appendix.

2 Wholesale Used-car Auctions

Any trade in wholesale used-car auctions is intrinsically intended for a future resale of the car. Used-car dealers participate in wholesale auctions to manage their inventories in response to either immediate demand on hand or anticipated demand in the future. The supply of used-car is generally considered not so elastic but the demand side is at least as elastic as the demand for new cars. In the wholesale used-car auctions, some bidders are general dealers who do business in almost every model; however, others may specialize in some specific types of cars. Also, some bidders are from large dealers, while many others are from small dealerships.⁵

For dealers, there exists at least two different situations about the latent demand structure. First, a dealer may have a specific demand, i.e. a pre-order, from a final buyer on hand. This can happen when a dealer gets an order for a specific used-car but does not have one in stock or when a consumer looks up the item list of an auction and asks a dealer to buy a specific car for her. The item list is publicly available on the auction house’s website 3-4 days prior to each auction day and these kinds of delegated biddings seem to be not uncommon from casual observations. We call this as “demand-on-hand” situation. In this case, we make a conjecture that a dealer already knows how much money she can make if she gets the car and sells it to the consumer. Second, though a dealer does not have any specific demand on hand, but she anticipates some demand for a car in the auction in near future. We call this as “anticipated-demand” situation. In this case, we make a conjecture that a dealer is not so certain and less confident about her anticipation of future market condition and has some incentives to find out other dealers’ opinions about it.

The auction data comes from an offline auction house located in Suwon, Korea. Suwon is located within an hour drive south of Seoul, the capital of Korea. It opened in May 2000 and it is the first

⁵This suggests that there may exist asymmetry among bidders, which we do not study in this paper.

fully-computerized wholesale used-car auction house in Korea. The auction house engages only in the wholesale auctions and it has held wholesale used-car auctions once a week since its opening.

The auction house mainly plays the role of an intermediary as many auction houses do. While sellers can be anyone, both individuals and firms, who wants to sell her own car through the auction, only a used-car dealer who is registered as a member of the auction house can participate in the auction as a buyer. At the beginning, the number of total memberships was around 250, and now it has grown to about 350 and the set of the members is a relatively stable group of dealers, which makes the data from this auction more reliable to conduct meaningful analyses than those from online auctions in general.

Roughly, about a half of the members come to the auction each week. 600-1000 cars are auctioned on a single auction day and 40-50 percent of those cars are actually sold through the auctions, which implies that a typical dealer who comes to the auction house on an auction day gets 2-4 cars/week on average. Bidders, i.e. dealers, in the auction have resale markets and, therefore, we can view this as if they try to win these used-cars only to resell them to the final consumers or, rarely, to other dealers.

The auction house generates its revenue by collecting fees from both a seller and a buyer when a car is sold through the auction. The fee is 2.2% of the transaction price, both for the seller and the buyer. The auction house also collects a fixed fee, about 50 US dollars, from a seller when the seller consigns her car to the auction house. The auction house's objective should be long-term profit maximization. Since there exists repeated relationship between dealers and the house, it may be important for the house to build favorable reputation. There also exists a competition among three similar wholesale used-car auction houses in Korea. In this paper, we ignore any effect from the competition and consider the auction house as a single monopolist for simplicity.

3 Data

3.1 Description of the Data

The auction design itself is an interesting and unique variant of ascending auctions. There is a reserve price set by an original seller with consultations from the auction house. An auction starts

from an opening bid, which is somewhere below the reserve price. The opening bids are made public on the auction house's website two or three days before an auction day. After an auction starts, the current price of the auction increases by a small, fixed increment, about 30 US dollars for all price ranges, whenever there are two or more bidders at the price who press their buttons beneath their desks in the auction hall. When the current price is below the reserve price, there only need to be one bidder to press the button for an auction to continue.

In the auction hall, there are big screens in front that display pictures and key descriptions of a car being auctioned. The current price is also displayed and updated real-time. Next to the screens, there are two important indicators for information disclosure. One of them resembles a traffic light with green, amber, and red lights, and the other is a sign that turns on when the current price rises above the reserve price, which means that the reserve price is made public once the current price reaches the level.

The three-colored lights indicate the number of bids at the current price. The green light signals three or more bids, yellow means two bids, and red indicates that there is only one bid at the current price while the current price is above the reserve price. When the current price is below the reserve price, they are indicating two or more, one, and zero bids respectively. This indicator is needed because, unlike the usual open out-cry ascending auctions, the bidders in this auction do not see who are pressing their buttons and therefore do not know how many bidders are placing their bids at the current price. With the colored lights, bidders only get somewhat incomplete information on the number of current bidders and they never observe the identities of the current bidders.

There is a short length of a single time interval, such that all the bids made in the same interval are considered as made at the same price. A bidder can indicate her willingness to buy at the current price by pressing her button at any time she wants, i.e. exit and reentry are freely allowed. The auction ends when three seconds have passed after there remains only one active bid at the current price. When an auction ends at the price above the reserve price, the item goes to the bidder who presses last, but when it ends at the price below the reserve price, the item is not sold.

Our data set consists of 59 weekly auctions from a period between October 2001 and December 2002. While all kinds of used-cars including passenger cars, buses, trucks, and other special-purpose vehicles, are auctioned in the auction house during the period, we select only passenger cars to

ensure a minimum heterogeneity of the objects in the auction. For those 59 weekly auctions, there were 28334 passenger-car auctions. However, we use only data from those auctions where a car is sold and there exist at least four bidders above the reserve price in an auction. After we apply this criteria, we have 5184 cars, 18.3% of all the passenger-cars auctioned. Although this step restricts our sample from the original data set, we do this because we need at least four meaningful bids, which means they should be above the reserve price, to conduct a nonparametric test we develop later and we think our analysis remains meaningful since what are important in generating most of the revenue for the auction house are those cars for which there are at least four bidders above the reserve price. In the estimation stage of our econometric analysis of the sample, we are forced to forgo additional cars out of those 5184 auctions because of the ties among the second, third, and fourth highest bids. We never observe the first highest bids in ascending auctions.

We remove the total of 1358 cars and we use the final sample of 3826 auctions.⁶ Among those, 1676 cars are from the maker Hyundai, the biggest Korean carmaker, 1186 cars are from the maker Daewoo, now owned by General Motors, and the remaining 964 cars are from the other makers like Kia, Ssangyong, and Samsung while most of them are from Kia, which merged with Hyundai in 1999. The market shares of major car makers in the sample are presented in Table 1 and some important summary statistics of the sample are provided in Table 2.

Available data includes the detailed bid-level, button-pressing log data for every auction. Auction covariates, very detailed characteristics of cars auctioned, are also available. The covariates include each car's make, model, production date, engine-size, mileage, rating - the auction house inspects each car and gives a 10-0 scaled rating to each car, which can be viewed as a summary statistic of the car's exterior and mechanical condition, transmission-type, fuel-type, color, options, purpose, body-type etc. We also observe the opening bids and the reserve prices of all auctions. Some bidder-specific covariates such as identities, locations, 'members since' date are also available. The date of title change is available for each car sold, which may be used as a proxy for the resale date in a future study. Last, we only observe the information on 'who' come to the auction house at 'what time' of an auction day for a very rough estimate on the potential set of bidders.

Here is a descriptive snapshot of a typical auction day, September 4th, 2002. A total of 567

⁶The second-highest bid and the third-highest bid are tied in 842 auctions. In 624 auctions, the third-highest bid and the fourth-highest bid are tied, and for 108 auctions the second-, third-, and fourth-highest bids are all tied.

cars were auctioned on that day, 386 cars of which were passenger cars and the remaining 181 were full-size vans, trucks, buses, etc. Since this auction day was the first week of the month, there was relatively a small number of cars (the number of cars auctioned was greatest in the last auctions of the month) 248 cars (43 percent) were sold through the auction and, among those unsold, at least 72 cars sold afterwards through post-auction bargaining,⁷ or re-auctioning next week, etc. The log data shows that the first auction of the day started at 13:19 PM and the last auction of the day ended at 16:19 PM. It only took 19 seconds per auction and 43 seconds per transaction. 152 ID cards (132 dealers since some dealers have multiple IDs) were recorded as entered the house. On average each ID placed bids for 7.82 auctions during the day. There were 98 bidders who won at least one car but 40 bidders did not win a single car. On average, each bidder won 1.8 cars and there are three bidders who won more than 10 cars. Among 386 passenger cars, there were at least one bid in 218 auctions. Among those 218 auctions, 170 cars were successfully sold through auctions.

3.2 Interpretation of Losing Bids

As we briefly mentioned in the introduction, a strong interpretation of losing bids in our data is one of the main strong points of this study that distinguishes it from other work on ascending auctions. Our basic idea is that with the assumption of IPV and with some institutional features of this auction such as a fixed discrete increments of prices raised by an auctioneer, we are able to make strong interpretations of losing bids above the reserve price and therefore identify the corresponding order statistics of valuations from observed bids.

We do this based on the equilibrium implications of the auction game in the spirit of Haile and Tamer (2003). So, in the end, without imposing the button auction model explicitly, we can treat some observed bids as if they come from a button auction. Then, we use this information from observed bids to estimate the exact underlying distribution of valuations. Within the private values paradigm, in ascending auctions without irrevocable exits, the last price at which each bidder shows her willingness to win (i.e. each bidder's final exit price) can be directly interpreted as her private value, and with symmetry, as relevant order statistics because it is a weakly dominant strategy

⁷We do not study this post-auction bargaining. Larsen (2012) studies the efficiency of the post-auction bargaining in the wholesale used-car auctions in the US.

(also a strictly dominant strategy with high probabilities) for a bidder to place a bid at the highest price she can afford.⁸ Here, we assume there is no ‘real’ cost associated with each bidding action, i.e. an additional button pressing requires a bidder to consume zero or negligible additional amount of energy, given that the bidder is interested in a car being auctioned.

The basic setup in our auction game model is that we have a collection of single-object ascending auctions with all symmetric, risk-neutral bidders. The number of potential bidders is never observed to any bidders nor an econometrician. For the information structure, we assume independent private values (IPV), which means a bidder’s bidding strategy does not depend on any other bidder’s action or private information, i.e. an ascending auction with exogenous rise of prices with fixed discrete increments with no jump. Therefore, with IPV, the fact that any bidder has very limited observability of other current bidders’ identities or bidding activities does not complicate a bidder’s bidding strategy.

Within this setting, while it is obviously a weakly dominant strategy for a bidder to press her bidding button until the last moment she can afford to do so, it does not necessarily guarantee that a bidder’s actual observed last press corresponds to the bidder’s true maximum willingness to pay. A strictly positive potential benefit from pressing at the last price a bidder can afford will ensure that our strong interpretation of losing bids is a close approximation to the truth. Such a chance for a strictly positive benefits can be present when there exists a possibility that all the other remaining bidders simultaneously exit at the price, at which a bidder is considering to press her button. When the auction price rises continuously as in Milgrom and Weber (1982)’s button auction, with continuous distribution of valuations, the probability of such event is zero at any price; however, when price rises in a discrete manner with a minimum increment as in the auction we study, this is not a measure-zero event, especially for the higher bidders like the second-, third-, and fourth-highest bidders as well as the winner.⁹

⁸In common values model, this is not the case and the analysis is much more complicated because a bidder may try not to press her button unless it is absolutely necessary because of strategic consideration to conceal her information. See Riley (1988) and Bikhchandani and Riley (1991, 1993).

⁹More precisely, for this argument, we need to model the auction game such that each bidder can only press the button at any price simultaneously and at most once and that the auction stops when only one bidder presses her button at a price. Although this is not an exact design of the real auction we study, we think this modeling is a close approximation of the real auction.

4 Empirical Model and Identification

This section describes the basic set-up of an IPV model we analyze. Consider a Wholesale Used-Car Auction (hereafter, WUCA) of a single object with the number of risk-neutral potential bidders, $N \geq 2$, drawn from $p_n = Pr(N = n)$. Each potential bidder i has the valuation V^i , which is independently drawn from the absolutely continuous distribution $F(\cdot)$ with support $\mathcal{V} = [\underline{v}, \bar{v}]$. Each bidder knows only his valuation but the distribution $F(\cdot)$ and the distribution p_n are common knowledge. By the design of WUCA, we can treat it as a button auction if we disregard the minimum increment. The minimum increment (about 30 dollars) in WUCA is small relative to the average car value (around 3,000 dollars) sold in WUCA, which is about one percent of the average car value.

Hence, in what follows, we simply disregard the existence of the minimum increment in WUCA to make our discussion simple and we consider a bound estimation approach that is robust to the minimum increment in Section 8.2. Suppose we observe the number of potential bidders and any i^{th} order statistic of the valuation (identical to i^{th} order statistic of the bids). Then we can identify the distribution of valuations from the cumulative density function (CDF) of the i^{th} order statistic as done in the previous literature (See e.g. Athey and Haile (2002) for this result. Also see Arnold et al. (1992) and David (1981) for extensive statistical treatments on order statistics).

Define the CDF of the i^{th} order statistic of the sample size n as

$$G^{(i:n)}(v) \equiv H(F(v); i : n) = \frac{n!}{(i-1)!(n-i)!} \int_0^{F(v)} t^{i-1} (1-t)^{n-i} dt$$

Then, we obtain the distribution of the valuations $F(\cdot)$ from

$$F(v) = H^{-1}(G^{(i:n)}(v); i : n).$$

However, in the auction we consider, we do not know the exact number of potential bidders in a given auction and the number of potential bidders varies over different auctions. Nonetheless we can still identify the distribution of valuations $F(\cdot)$ following the methodology proposed by Song (2005), since we observe several order statistics in a given auction. Song (2005) showed that an arbitrary absolutely continuous distribution $F(\cdot)$ is nonparametrically identified from observations

of any pair of order statistics from an i.i.d. sample, even when the sample size n is unknown and stochastic.

The key idea is that we can interpret the density of the k_1^{th} highest value \tilde{V} conditional on the k_2^{th} highest value V as the density of the $(k_2 - k_1)^{th}$ order statistic from a sample of size $(k_2 - 1)$ that follows $F(\cdot)$. To see this denote the probability density function (PDF) of the i^{th} order statistic of the n sample as $g^{(i:n)}(v)$

$$g^{(i:n)}(v) = \frac{n!}{(i-1)!(n-i)!} [F(v)]^{i-1} [1-F(v)]^{n-i} f(v) \quad (1)$$

and the joint density of the i^{th} and the j^{th} order statistics of the n sample for $n \geq j > i \geq 1$ as $g^{(i,j:n)}(\tilde{v}, v)$

$$g^{(i,j:n)}(\tilde{v}, v) = \frac{n! [F(v)]^{i-1} [F(\tilde{v}) - F(v)]^{j-i-1} [1-F(\tilde{v})]^{n-j} f(v) f(\tilde{v})}{(i-1)!(j-i-1)!(n-j)!} I_{\{\tilde{v} > v\}}.$$

Then, the density of \tilde{V} conditional on V , denoted by $p_{k_1|k_2}(\tilde{v}|V = v)$ can be written

$$\begin{aligned} p_{k_1|k_2}(\tilde{v}|v) &= \frac{g^{(n-k_2+1, n-k_1+1:n)}(\tilde{v}, v)}{g^{(n-k_2+1:n)}(v)} \quad (2) \\ &= \frac{(k_2 - 1)!}{(k_2 - k_1 - 1)!(k_1 - 1)!} \frac{(F(\tilde{v}) - F(v))^{k_2 - k_1 - 1} (1 - F(\tilde{v}))^{k_1 - 1} f(\tilde{v})}{(1 - F(v))^{k_2 - 1}} I_{\{\tilde{v} \geq v\}} \\ &= \frac{(k_2 - 1)!}{(k_2 - k_1 - 1)!(k_1 - 1)!} \frac{((1 - F(v))F(\tilde{v}|v))^{k_2 - k_1 - 1}}{((1 - F(v))(1 - F(\tilde{v}|v)))^{k_1 - 1} f(\tilde{v}|v)(1 - F(v))} I_{\{\tilde{v} \geq v\}} \\ &= \frac{(k_2 - 1)!}{(k_2 - 1 - k_1)!(k_2 - 1 - (k_2 - k_1))!} F(\tilde{v}|v)^{k_2 - 1 - k_1} (1 - F(\tilde{v}|v))^{k_1 - 1} f(\tilde{v}|v) I_{\{\tilde{v} \geq v\}} \\ &\equiv g^{(k_2 - k_1 : k_2 - 1)}(\tilde{v}|v), \end{aligned}$$

where $f(\tilde{v}|v)(g^{(\cdot)}(\tilde{v}|v))$ denotes the truncated density of $f(\cdot)(g^{(\cdot)}(\cdot))$ truncated at v from below and $F(\tilde{v}|v)$ denotes the truncated distribution of $F(\cdot)$ truncated at v from below.¹⁰ In the above $p_{k_1|k_2}(\tilde{v}|v)$ is nonparametrically identified directly from the bidding data and thus $g^{(k_2 - k_1 : k_2 - 1)}(\tilde{v}|v)$ is identified. Then we can interpret $g^{(k_2 - k_1 : k_2 - 1)}(\tilde{v}|v)$ as the probability density function $g^{(i:n)}(\cdot)$ of the i^{th} order statistic of the n sample with $n = k_2 - 1$ and $i = k_2 - k_1$, further noting that

¹⁰To be precise, $f(\tilde{v}|v) = \frac{f(\tilde{v})}{1-F(v)}$, $F(\tilde{v}|v) = \frac{F(\tilde{v})-F(v)}{1-F(v)}$, and $g^{(k_2 - k_1 : k_2 - 1)}(\tilde{v}|v) = \frac{g^{(k_2 - k_1 : k_2 - 1)}(\tilde{v})}{1-G^{(k_2 - k_1 : k_2 - 1)}(v)}$.

$$\lim_{v \rightarrow \underline{v}} p_{k_1|k_2}(\tilde{v}|v) = \lim_{v \rightarrow \underline{v}} g^{(k_2-k_1:k_2-1)}(\tilde{v}|v) = g^{(k_2-k_1:k_2-1)}(\tilde{v}).$$

Then, identification of the distribution of valuations $F(\cdot)$ is straightforward by Theorem 1 in Athey and Haile (2002) stating that the parent distribution is identified whenever the distribution of any order statistic (here $k_2 - k_1$) with a known sample size (here $k_2 - 1$) is identified.

4.1 Observed Auction Heterogeneity

In practice, the valuation of objects sold in WUCA (as in other auctions) varies according to several observed characteristics such as car type, make, mileage, year, and etc. We want to control the effect of these observables on the individual valuation. For this purpose, we assume the following form of the valuation $\mathbb{V}_{ti} \equiv \mathbb{V}_i(X_t)$ as

$$\ln \mathbb{V}_{ti} = l(X_t) + V_{ti}, \tag{3}$$

where $l(\cdot)$ is a nonparametric or parametric function of X_t , X_t is a vector of observable characteristics of the auction $t = 1, \dots, T$, and i denotes the bidder i . We assume that V_{ti} is independent of X_t . Thus, we assume the multiplicatively separable structure of the value function in level, which is preserved by equilibrium bidding. Ignoring the minimal increment, we then have

$$\ln B(\mathbb{V}_{ti}) = l(X_t) + b(V_{ti}),$$

where $B(\mathbb{V}_{ti})$ is a bidding function of a bidder i with observed heterogeneity X_t of the auction t and $b(V_{ti})$ is a bidding function of a *pseudo* auction with the *homogeneous* objects. Thus, under the IPV assumption, we have $B(\mathbb{V}) = \mathbb{V}$ and $b(V) = V$ which we will name as a *pseudo* bid or valuation.

In our empirical approach we use a parametric specification of $l(\cdot)$ and focus on the flexible nonparametric estimation of the distribution of V denoted by $F(\cdot)$. We let

$$\ln \mathbb{V}_{ti} = X_t' \beta + V_{ti} \tag{4}$$

for a $\dim(X_t) \times 1$ parameter β . In Appendix B, we consider a nonparametric extension of $l(\cdot)$.

5 Estimation

To estimate the distribution of valuations we build on the semi-nonparametric (SNP) estimation procedure developed by Gallant and Nychka (1987) and Coppejans and Gallant (2002). In particular, we implement a particular sieve estimation of the unknown density function of the valuations using a Hermite series. First, we approximate the function space, \mathcal{F} containing the true distribution of valuations with a sieve space of the Hermite series, denoted by \mathcal{F}_T . Once we set up the objective function based on a Hermite series approximation of the unknown density function, then the estimation procedure becomes a finite dimensional parametric problem. In particular, we use the (pseudo) maximum likelihood methods. What remains is to specify the particular rate in which a sieve space \mathcal{F}_T - defined below - gets closer to \mathcal{F} achieving the consistency of the estimator. We will specify several regularity conditions for this in Appendix E.

Since we observe at least the second-, third- and fourth-highest bids in each auction of WUCA. We can estimate several different versions of the distributions of valuations $F(\cdot)$, because each pair of order statistics can identify the parent distribution as we discuss in the previous section. Here, we use two pairs of order statistics (second-, fourth-) and (third-, fourth-) highest bids and obtain two different estimates of $F(\cdot)$, which provides us an opportunity to test the hypothesis that WUCA is the IPV. This testable implication comes from the fact that under the IPV, value of $F(\cdot)$ implied by the distributions of different order statistics must be identical for all valuations (For detailed discussion, see Athey and Haile 2002).

Once we do not reject the assumption of the IPV, then we can combine several order statistics to recover $F(\cdot)$. This version of estimator can be more efficient than the version that uses a pair of order statistics only in the sense that we are using more information.

5.1 Estimation of the Distribution of Valuations

In this section, to simplify the discussion, we assume that there is no observed heterogeneity in the auction. In other words, we impose $\ln \mathbb{V}_{ti} = V_{ti}$, i.e. $l(\cdot) \equiv 0$. We also let the mean of V_{ti} equal to zero and the variance equal to one. We can do this without loss of generality because we can always standardize the data before the estimation and also the SNP estimator is the scale- and location-

invariant. In the actual estimation, we estimate the mean and the variance with other parameters.

We focus on the estimation of the distribution of valuations using the second- and fourth-highest bids. Let (\tilde{V}_t, V_t) denote the second- and fourth-highest pseudo-bids (or valuations) for each auction t and (\tilde{v}_t, v_t) denote their realizations. We assume T observations of the auctions are available as the data. Let $v_m = \min_t v_t$ and $v^M = \max_t \tilde{v}_t$.¹¹ Noting that $F(v)$ for $v < v_m$ or $v > v^M$ cannot be recovered solely from the data, we impose $\underline{v} = v_m - \epsilon$ and $\bar{v} = v^M + \epsilon$ for some small positive ϵ ¹² and let $F^*(\cdot) \equiv F(\cdot|\underline{v}, \bar{v})$ (i.e. we let $\mathcal{V} = [\underline{v}, \bar{v}]$) as the model primitive of interest, where $F(\cdot|\underline{v}, \bar{v})$ denotes the truncated distribution of $F(\cdot)$ from below at \underline{v} and from above at \bar{v} as

$$F^*(v) \equiv F(v|\underline{v}, \bar{v}) = \frac{F(v) - F(\underline{v})}{F(\bar{v}) - F(\underline{v})} \text{ and hence } f^*(v) \equiv f(v|\underline{v}, \bar{v}) = \frac{f(v)}{F(\bar{v}) - F(\underline{v})}. \quad (5)$$

Then, we obtain the density of \tilde{V}_i conditional on V_i denoted by $p_{2|4}(\tilde{v}_i|V_i = v_i)$ from (2) as

$$p_{2|4}(\tilde{v}|V = v) = \frac{6(F^*(\tilde{v}) - F^*(v))(1 - F^*(\tilde{v}))f^*(\tilde{v})}{(1 - F^*(v))^3} \text{ for } \bar{v} > \tilde{v} > v > \underline{v}. \quad (6)$$

To estimate the unknown function $f^*(z)$ (hence, $F^*(z) = \int_{\underline{v}}^z f^*(v)dv$), we first approximate $f(z)$ with a member of \mathcal{F}_T , denoted by $f_{(K)}$, up to the order $K(T)$ - we will suppress the argument T in $K(T)$ unless noted otherwise:

$$\begin{aligned} \mathcal{F}_T &= \{f_{(K)} : f(z, \theta) = \left(\sum_{j=1}^K \vartheta_j H_j(z)\right)^2 + \epsilon_0 \phi(z), \theta \in \Theta_T\} \\ \Theta_T &= \left\{ \theta = (\vartheta_1, \dots, \vartheta_{K(T)}) : \sum_{j=1}^{K(T)} \vartheta_j^2 + \epsilon_0 = 1 \right\} \end{aligned} \quad (7)$$

for a small positive constant ϵ_0 and the standard normal pdf $\phi(z)$ where the Hermite series $H_j(z)$

¹¹Note that v_m is a consistent estimator of the lower bound of \mathcal{V} under no binding reserve price and a consistent estimator of the reserve price under the binding case. Similarly v^M is a consistent estimator of the upper bound of \mathcal{V} .

¹²This is useful for implementing our estimation. For example, in (9), we need this restriction. Otherwise, for $v_i = \bar{v} = v^M$ with a non-negligible measure, (9) is not defined by construction. As an alternative, we may let $\mathcal{V} = [v_m, v^M]$ and consider a trimmed (pseudo) MLE by trimming out those observations of $v < v_m + \epsilon$ and $v > v^M - \epsilon$. In Appendix we derive our asymptotic results with a trimming device in case one may want to use this trimming in the estimation, although it is not necessary. Note that this is a separate issue from handling of outliers. Outliers should be eliminated before the estimation if it is necessary.

is defined recursively as

$$\begin{aligned}
H_1(z) &= (\sqrt{2\pi})^{-1/2} e^{-z^2/4}, \\
H_2(z) &= (\sqrt{2\pi})^{-1/2} z e^{-z^2/4}, \\
H_j(z) &= [zH_{j-1}(z) - \sqrt{j-1}H_{j-2}(z)]/\sqrt{j}, \text{ for } j \geq 3.
\end{aligned} \tag{8}$$

This \mathcal{F}_T is the sieve space considered originally in Gallant and Nychka (1987), which can approximate space of continuous density functions when $K(T) \rightarrow \infty$ as $T \rightarrow \infty$. Then, we construct the single observation likelihood function based on $f_{(K)}(\cdot)$ instead of the true $f(\cdot)$ using (6):

$$L(f_{(K)}; \tilde{v}_t, v_t) = \frac{6(F_{(K)}^*(\tilde{v}_t) - F_{(K)}^*(v_t))(1 - F_{(K)}^*(\tilde{v}_t))f_{(K)}^*(\tilde{v}_t)}{(1 - F_{(K)}^*(v_t))^3}, \tag{9}$$

where $f_{(K)}^*(\cdot) = \frac{f_{(K)}(v)}{F_{(K)}(\tilde{v}) - F_{(K)}(v)}$, $F_{(K)}(v) = \int_{-\infty}^v f_{(K)}(z)dz$, and $F_{(K)}^*(v) = \int_{-\infty}^v f_{(K)}^*(z)dz$. Noting that (9) is a parametric likelihood function for a given value of K , one can estimate $f_{(K)}(\cdot)$ with $\hat{f}(\cdot)$ as the maximum likelihood estimator:

$$\begin{aligned}
\hat{f}(z) &= \left(\sum_{j=1}^K \hat{\vartheta}_j H_j(z) \right)^2 + \epsilon_0 \phi(z), \text{ where} \\
\hat{\theta} &= \left(\hat{\vartheta}_1, \dots, \hat{\vartheta}_K \right) = \operatorname{argmax}_{\theta \in \Theta_T} \frac{1}{T} \sum_{t=1}^T \ln L(f_{(K)}; \tilde{v}_t, v_t) \\
\text{or equivalently } \hat{f} &= \operatorname{argmax}_{f_{(K)} \in \mathcal{F}_T} \frac{1}{T} \sum_{t=1}^T \ln L(f_{(K)}; \tilde{v}_t, v_t).
\end{aligned} \tag{10}$$

Note that actually a pseudo-valuation \tilde{v}_t or v_t is defined as the residual in (4) - or approximated in the nonparametric case as the residual in Appendix B (29). Thus, we have additional set of parameters β in (4) to estimate together with $\hat{\theta}$ and our SNP density estimator of the valuations is defined by

$$\hat{f}(z) = \left(\sum_{j=1}^K \hat{\vartheta}_j H_j(z) \right)^2 + \epsilon_0 \phi(z), \tag{11}$$

where

$$(\hat{\beta}, \hat{\theta}) = \operatorname{argmax}_{\beta, \theta \in \Theta_T} \ln L(f_{(K)}; \tilde{v}_t = \ln \tilde{\mathbb{V}}_t - X_t' \beta, v_t = \ln \mathbb{V}_t - X_t' \beta)$$

and $(\tilde{\mathbb{V}}_t, \mathbb{V}_t)$ denote the second- and fourth-highest bids or valuations.

One may be concerned about the fact that we implicitly assume the support of the values is

$(-\infty, \infty)$ in the SNP density where the true support is $[\underline{v}, \bar{v}]$. This turns out to be not a concern according to Kim (2007). He uses a truncated version of the Hermite polynomials to resolve this issue and shows that there exists an one-to-one mapping between the function space of the original Hermite polynomials and its truncated version. These issues are handled in Appendix D and E in details. In Appendix E we study the consistency and the convergence rate of the SNP estimator. Finally, from (5), we obtain the consistent estimators of $f^*(\cdot)$ and $F^*(\cdot)$, respectively as

$$\hat{f}^*(v) = \frac{\hat{f}(v)}{\hat{F}(\bar{v}) - \hat{F}(\underline{v})} \text{ and } \hat{F}^*(v) = \frac{\hat{F}(v) - \hat{F}(\underline{v})}{\hat{F}(\bar{v}) - \hat{F}(\underline{v})}$$

where $\hat{F}(v) = \int_{-\infty}^v \hat{f}(z) dz$.

In what follows we will denote the conditional density identified from the second- and fourth-highest bids by $f_{2|4}$ and its estimator by $\hat{f}_{2|4}$ for comparison with other density functions obtained from other pairs of order statistics. Similarly we can also identify and estimate $f^*(\cdot)$ using the pair of third- and fourth-highest bids. First, denote (\tilde{V}_t, V_t) to be the third- and fourth-highest pseudo-bids for each auction t and let (\tilde{v}_t, v_t) denote their realizations, respectively. Then, we obtain the conditional density

$$p_{3|4}(\tilde{v}|V = v) = \frac{3(1 - F^*(\tilde{v}))^2 f^*(\tilde{v})}{(1 - F^*(v))^3} \text{ for } \bar{v} > \tilde{v} > v > \underline{v} \quad (12)$$

from (2). Denote the estimate of $f^*(\cdot)$ based on (12) as $\hat{f}_{3|4}(\cdot)$:

$$\hat{f}_{3|4} = \frac{\hat{f}(v)}{\hat{F}(\bar{v}) - \hat{F}(\underline{v})} \text{ for } \hat{f} = \operatorname{argmax}_{f_{(K)} \in \mathcal{F}_T} \frac{1}{T} \sum_{t=1}^T \ln \frac{3(1 - F_{(K)}^*(\tilde{v}_t))^2 f_{(K)}^*(\tilde{v}_t)}{(1 - F_{(K)}^*(v_t))^3}. \quad (13)$$

The SNP density estimators using other pairs of order statistics (e.g. the second highest and the third highest bids) can be similarly obtained by constructing required likelihood functions from (2).

Note that the approximation precision of the SNP density depends on the choice of smoothing parameter K . We can pick the optimal length of series following the Coppejans and Gallant (2002)'s method, which is a cross-validation strategy as often used in Kernel density estimations. See Appendix C for detailed discussion on how to choose the optimal K in our estimation.

5.2 Two-Step Estimation

Though the estimation procedure considered in the previous section can be implemented in one-step, one may prefer a two-step estimation method as follows, so that one can avoid the computational burden involved in estimating $l(\cdot)$ and $f(\cdot)$ simultaneously. We also consider nonparametric estimation of $l(\cdot)$. In this case, a two-step procedure will be more useful.

In the two-step approach, first we estimate the following first-stage regression:

$$\ln \mathbb{V}_{ti} = X_t' \beta + V_{ti}$$

to obtain the pseudo-valuations. Then, we construct the residuals for each order statistics as

$$\hat{v}_{ti} = \ln \mathbb{V}_{ti} - X_t' \hat{\beta}.$$

In the second step, based on the estimated pseudo values \hat{v}_{ti} , we estimate $f(\cdot)$ as (10) or (13) in the previous section. Note that the convergence rate of $\hat{\beta}$ to β (including mean and variance estimates of the valuations), which is typically the parametric rate, will be faster than that of \hat{f} to f . Even for the fully nonparametric case, this is true if a consistent estimator of $l(\cdot)$ converges to $l(\cdot)$ faster than $\hat{f}(\cdot)$ to $f(\cdot)$. Thus, this two-step method will not affect the convergence rate of the SNP density estimator. This further justifies the use of the two-step estimation.

The consistency and the convergence rates of this two-step estimator are provided in Appendix E (Theorem E.1).

6 Nonparametric Testing

Here we consider a novel and practical testing framework that can test the IPV assumption. Our main idea is that when several versions of the distribution of valuations can be recovered from the observed bids, they should be identical under the IPV assumption. For our testing we compare two versions of the distribution of valuations. In particular, we focus on two densities: one is from the pair of the second- and fourth- highest bids and the other is from the pair of the third- and fourth-order statistics. Therefore, by comparing $\hat{f}_{2|4}(\cdot)$ and $\hat{f}_{3|4}(\cdot)$, we can test the following hypothesis

H_0

$$H_0 : \text{WUCA is an IPV auction} \quad (14)$$

$$H_A : \text{WUCA is not an IPV auction}$$

because under H_0 , there should be no significant difference between $f_{2|4}(\cdot)$ and $f_{3|4}(\cdot)$. Therefore under the null we have

$$H_0 : f_{2|4}(\cdot) = f_{3|4}(\cdot) \quad (15)$$

against $H_A : f_{2|4}(\cdot) \neq f_{3|4}(\cdot)$. This testability result extends Theorem 1 of Athey and Haile (2002) that assume the known number of potential bidders to the case when the number is unknown. We summarize this result as a theorem:

Theorem 6.1 *In the symmetric IPV model with the unknown number of potential bidders, the model is testable if we have three or more order statistics of bids from the auction.*

6.1 Tests based on Means or Higher Moments

One can test a weak version of (14) based on the means or higher moments implied by $f_{2|4}$ and $f_{3|4}$ as

$$\begin{aligned} H'_0 & : \mu_{2|4}^j = \mu_{3|4}^j \text{ for all } j = 1, 2, \dots, J \\ H'_A & : \mu_{2|4}^j \neq \mu_{3|4}^j \text{ for some } j = 1, 2, \dots, J \end{aligned} \quad (16)$$

where $\mu_{2|4}^j = \int_{\underline{v}}^{\bar{v}} v^j f_{2|4}(v) dv$ and $\mu_{3|4}^j = \int_{\underline{v}}^{\bar{v}} v^j f_{3|4}(v) dv$, since (14) implies (16). One can compare estimates of several moments implied by $\hat{f}_{2|4}$ and $\hat{f}_{3|4}$ and test for the significance of difference in each pair of moments by constructing standardized test statistics. A difficulty of implementing this approach will be to reflect into the test statistics the fact that one uses pre-estimated density functions $\hat{f}_{2|4}$ and $\hat{f}_{3|4}$ to estimate the moments.

6.2 Comparison of Densities Using the Pseudo Kullback-Leibler Divergence Measure

We are interested in testing the equivalence of two densities $f_{2|4}$ and $f_{3|4}$ where these densities are estimated from (11) and (13), respectively using the SNP density estimation. One measure to compare $f_{2|4}$ and $f_{3|4}$ will be the integrated squared error given by $I_s(f_{2|4}(z), f_{3|4}(z)) = \int_{\mathcal{V}} (f_{2|4}(z) - f_{3|4}(z))^2 dz$. Under the null we have $I_s(f_{2|4}(z), f_{3|4}(z)) = 0$. Li (1996) develops a test statistic of this sort when both densities are estimated using a kernel method. Other useful measures for the distance of two density functions are the Kullback-Leibler (KL) information distance and the Hellinger metric. For testing of (e.g.) serial independence using densities, it is well known that test statistics based on these two measures have better small sample properties than those based on a squared distance in terms of a second order asymptotic theory. The KL measure is entertained in Ullah and Singh (1989), Robinson (1991), and Hong and White (2005) when they test the affinity of two densities. The KL measure is defined by

$$I_{KL} = \int_{\mathcal{V}} (\ln f_{2|4}(z) - \ln f_{3|4}(z)) f_{2|4}(z) dz \quad (17)$$

or $I_{KL} = \int_{\mathcal{V}} (\ln f_{3|4}(z) - \ln f_{2|4}(z)) f_{3|4}(z) dz$, which is equally zero under the null and has positive values under the alternative (see Kullback and Leibler 1951). One could construct a test statistic using a sample analogue of (17):

$$I_{KL}(\hat{f}_{2|4}, \hat{f}_{3|4}) = \int_{\mathcal{V}} (\ln \hat{f}_{2|4}(z) - \ln \hat{f}_{3|4}(z)) d\hat{F}_{2|4}(z).$$

Now suppose $\{\tilde{v}_t\}_{t=1}^T$ in the data are the second-highest order statistics. Then, one may expect $\int_{\mathcal{V}} (\ln \hat{f}_{2|4}(z) - \ln \hat{f}_{3|4}(z)) d\hat{F}_{2|4}(z) \approx \frac{1}{T} \sum_{t=1}^T (\ln \hat{f}_{2|4}(\tilde{v}_t) - \ln \hat{f}_{3|4}(\tilde{v}_t))$ but this is not true because \tilde{v}_t follows the distribution of the second-highest order statistic, not that of valuations $F_{2|4}$. We can, however, argue that

$$\frac{1}{T} \sum_{t=1}^T (\ln \hat{f}_{2|4}(\tilde{v}_t) - \ln \hat{f}_{3|4}(\tilde{v}_t))$$

is still useful in terms of comparing two densities because under the null the following object equals to zero

$$\int_{\mathcal{V}} (\ln f_{2|4}(z) - \ln f_{3|4}(z)) g^{(n-1:n)}(z) dz \quad (18)$$

where $g^{(n-1:n)}(z)$ is the density of the second-highest order statistic of the n sample.¹³ Note that (18) is not a distance measure since it can take negative values but still can serve as a divergence measure.

6.3 Test Statistic

We will use a modification of (18) as a divergence measure of two densities

$$I^g(f_{2|4}, f_{3|4}) = \left(A \int_{\mathcal{V}} (\ln f_{2|4}(z) - \ln f_{3|4}(z)) g^{(n-1:n)}(z) dz \right)^2 \quad (19)$$

for a positive constant A and develop a practical test statistic similar to Robinson (1991).¹⁴ Using a sample analogue of (19) with \hat{A} being a consistent estimator of A , one may propose a test statistic of the following form using the second-highest order statistics¹⁵

$$\hat{I}^g(\hat{f}_{2|4}, \hat{f}_{3|4}) = \left(\hat{A} \frac{1}{T} \sum_{t \in \mathcal{T}_2} (\ln \hat{f}_{2|4}(\tilde{v}_t) - \ln \hat{f}_{3|4}(\tilde{v}_{t+1})) \right)^2 \quad (20)$$

where \mathcal{T}_2 is a subset of $\{1, 2, \dots, T-1\}$, which trims out those observations of $\hat{f}_{2|4}(\cdot) < \delta_1(T)$ or $\hat{f}_{3|4}(\cdot) < \delta_2(T)$ for chosen positive values of $\delta_1(T)$ and $\delta_2(T)$ that tend to zero as $T \rightarrow \infty$. To be precise, \mathcal{T}_2 is defined by

$$\mathcal{T}_2 = \left\{ t : 1 \leq t \leq T-1 \text{ such that } \hat{f}_{2|4}(\tilde{v}_t) > \delta_1(T) \text{ and } \hat{f}_{3|4}(\tilde{v}_{t+1}) > \delta_2(T) \right\}.$$

Trimming is often used in an inference procedure of nonparametric estimations. Even though the SNP density estimator is always positive by construction, we still introduce this trimming device to avoid the excess influence of one or several summands when $\hat{f}_{2|4}(\cdot)$ or $\hat{f}_{3|4}(\cdot)$ are arbitrary small. (20)

¹³This also holds for other order statistics of valuations. We will be confident with the testing results if the testing is consistent regardless of choice of the distribution which we are performing the test based on.

¹⁴Robinson (1991) uses a kernel density estimator while we use the SNP density estimator. We cannot use a kernel estimation in our problem because our objective function is nonlinear in the density function of interest.

¹⁵Instead, we can use other order statistics. In that case simply replace $g(n-1:n)$ with $g(n-2:n)$ or $g(n-3:n)$.

will converge to $I^g(f_{2|4}, f_{3|4}) = (A \int_{\mathcal{Y}} (\ln f_{2|4}(z) - \ln f_{3|4}(z)) g^{(n-1:n)}(z) dz)^2$ under certain regularity conditions that will be discussed later. However, $\sqrt{T} \cdot \widehat{I}^g(\widehat{f}_{2|4}, \widehat{f}_{3|4})$ will have degenerate distributions under the null by a similar reason discussed in Robinson (1991) and cannot be used as reasonable statistic. To resolve this problem, we entertain modification of (20) in the spirit of Robinson (1991) as

$$\widehat{I}_\gamma^g(\widehat{f}_{2|4}, \widehat{f}_{3|4}) = \left(\widehat{A} \frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \left(\ln \widehat{f}_{2|4}(\tilde{v}_t) - \ln \widehat{f}_{3|4}(\tilde{v}_{t+1}) \right) \right)^2 \quad (21)$$

where for a nonnegative constant γ we define

$$c_t(\gamma) = 1 + \gamma \text{ if } t \text{ is odd and } c_t(\gamma) = 1 - \gamma \text{ if } t \text{ is even}$$

and T_γ is defined as¹⁶

$$T_\gamma = T + \gamma \text{ if } T \text{ is odd and } T_\gamma = T \text{ if } T \text{ is even.} \quad (22)$$

We let

$$I^g = I^g(f_{2|4}, f_{3|4}), \quad \widetilde{I}_\gamma^g = \widehat{I}_\gamma^g(f_{2|4}, f_{3|4}), \text{ and } \widehat{I}_\gamma^g = \widehat{I}_\gamma^g(\widehat{f}_{2|4}, \widehat{f}_{3|4})$$

for notational simplicity. Now note that for any increasing sequence $d(T)$, any positive $C < \infty$, and γ ,

$$\Pr \left[d(T) \widehat{I}_\gamma^g < C \right] \leq \Pr \left[d(T) \left| \widehat{I}_\gamma^g - I^g \right| > d(T) I^g - C \right] \leq \Pr \left[\left| \widehat{I}_\gamma^g - I^g \right| > I^g / 2 \right]$$

holds when T is sufficiently large and (15) is not true under the reference density g (i.e. $I^g > 0$). Since the probability $\Pr \left[\left| \widehat{I}_\gamma^g - I^g \right| > I^g / 2 \right]$ goes to zero under the alternative as long as $\widehat{I}_\gamma^g \xrightarrow{p} I^g$, one can construct a test statistic of the form

$$\text{Reject (15) when } d(T) \widehat{I}_\gamma^g > C. \quad (23)$$

¹⁶Consider $s(T) \equiv \frac{1}{T_\gamma} \sum_{t=1}^T c_t(\gamma)$ when $T = 2m$ and $T = 2m + 1$, respectively. It follows that $s(2m) = \frac{((1+\gamma)m + (1-\gamma)m)}{T_\gamma} = \frac{2m}{T_\gamma} = \frac{T}{T_\gamma}$ and $s(2m + 1) = \frac{1}{T_\gamma} ((1 + \gamma)m + (1 - \gamma)m + (1 + \gamma)) = \frac{2m + 1 + \gamma}{T_\gamma} = \frac{T + \gamma}{T_\gamma}$. Thus, by constructing T_γ as (22), we have $s(T) = 1$.

Therefore, as long as $\widehat{I}_\gamma^g \xrightarrow[p]{p} I^g$, (23) is a valid test consistent against departures from (15) in the sense of $I^g > 0$. The test statistic follows an asymptotic Chi-square distribution, so the test is easy to implement.

We summarize the properties of our proposed test statistic in the following theorem. See Appendix F.1 for the proof. We assume that

Assumption 6.1 *The observed data of two pairs of order statistics $\{\tilde{v}_t, v_t\}$ and $\{\tilde{\tilde{v}}_t, v_t\}$ are randomly drawn from the continuous density of the parent distribution of valuations.*

Theorem 6.2 *Suppose Assumption 6.1 hold. Provided that Conditions 2-4 in Appendix hold under (15), we have $\widehat{\tau}_\gamma \equiv T\widehat{I}_\gamma^g \xrightarrow[d]{d} \chi^2(1)$ for any $\gamma > 0$ with $\widehat{A} = 1/(\sqrt{2\gamma}\widehat{\Sigma}^{\frac{1}{2}})$ and $\widehat{\Sigma} \xrightarrow[p]{p} \Sigma = 2\text{Var}_g[\ln f_{2|4}(\cdot)]$. Thus we reject (15) if $\widehat{\tau}_\gamma > C_\alpha$ where C_α is the critical value of the Chi-square distribution with the size equal to α .*

As a consistent estimator of the variance Σ one can use

$$\begin{aligned} \widehat{\Sigma}_1 &= 2 \left(\frac{1}{T} \sum_{t=1}^T (\ln \widehat{f}_{2|4}(\tilde{v}_t))^2 - \left\{ \frac{1}{T} \sum_{t=1}^T \ln \widehat{f}_{2|4}(\tilde{v}_t) \right\}^2 \right) \text{ or} \\ \widehat{\Sigma}_2 &= 2 \left(\frac{1}{T} \sum_{t=1}^T (\ln \widehat{f}_{3|4}(\tilde{v}_t))^2 - \left\{ \frac{1}{T} \sum_{t=1}^T \ln \widehat{f}_{3|4}(\tilde{v}_t) \right\}^2 \right) \end{aligned}$$

or its average $\widehat{\Sigma}_3 = (\widehat{\Sigma}_1 + \widehat{\Sigma}_2)/2$ because under the null $f_{2|4} = f_{3|4}$.

To implement the proposed test we need to pick a value for γ . Even though the test is valid for any choice of $\gamma > 0$ asymptotically, it is not necessarily true for finite samples (see Robinson 1991). The testing result could be sensitive to the choice of γ . Robinson (1991) suggests taking γ between 0 and 1. He argues that the power of the test is inversely related with the value of γ but one should not take γ too small for which the asymptotic normal approximation breaks down (i.e. the statistic does not have a standard asymptotic distribution).

7 Empirical Results

7.1 Monte Carlo Experiments

In this section, we perform several Monte Carlo experiments to illustrate the validity of our estimation strategy. We generate artificial data of $T = 1000$ auctions using the following design. The

number of potential bidders, $\{N_i\}$, are drawn from a Binomial distribution with $(n, p) = (50, 0.1)$ for each auction ($t = 1, \dots, T$). N_t potential bidders are assumed to value the object according to:

$$\ln \mathbb{V}_{ti} = \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + V_{ti}, \quad (24)$$

where $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = 0.5$, $X_{1t} \sim N(0, 1)$, $X_{2t} \sim Exp(1)$, $X_{3t} = X_{1t} \cdot X_{2t} + 1$, and $V_{ti} \sim Gamma(9, 3)$.¹⁷ $X_{.t}$'s represent the observed auction heterogeneity and V_{ti} is bidder i 's private information in auction t , whose distribution is of our primary interest. To consider the case of binding reserve prices, we also generate the reserve prices as

$$\ln R_t = \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \eta_t,$$

where $\eta_t \sim Gamma(9, 3) - 2$. Note that by construction \mathbb{V}_{ti} and R_t are independent conditional on $X_{.t}$'s. An artificial bidder bids only when \mathbb{V}_{ti} is greater than R_t . Here we assume our imaginary researcher does not know the presence of potential bidders with valuations below R_t . Thus, in each experiment, she has a data set of $X_{.t}$'s, and the second-, the third-, and the fourth-highest bids among observed bids. Auctions with fewer than four actual bidders are dropped. Hence, our researcher has the sample size less than $T = 1000$ (on average $T = 680$ with 50 repetitions). Our researcher estimates α_1 , α_2 , α_3 and $f(\cdot)$ (pdf of V_{ti}) by varying the approximation order (K) of the SNP estimator, from 0 to 7 without knowing the specification of the distribution of \mathbb{V}_{ti} in (24). We used the two-step estimation approach to reduce the computational burden. Among K between 0 to 7, we report estimation results with $K = 6$ because it performs best in this Monte Carlo experiments.

Figure 1 illustrates a performance of the estimator. By construction of our MC design, the three versions of density function estimates for valuations should not be statistically different from each other (one is based on the pair of $(2^{nd}, 4^{th})$ order statistics and others are based on the pair of $(3^{rd}, 4^{th})$ and $(2^{nd}, 3^{rd})$.)

¹⁷Note that for $V \sim Gamma(9, 3)$, $E(V) = 3$ and $Var(V) = 1$.

7.2 Estimation Results

We discuss estimation and testing results using the real auction data. In the first stage regression obtaining the function of the observed heterogeneity $l(x)$, we used a linear specification to minimize the computational burden of the SNP density estimation. We use the following covariates: X_1 is the vector of dummy variables indicating the car make such as Hyundai, Daewoo, Kia or Others; X_2 is the age; X_3 is the engine size; X_4 is the mileage; X_5 is the rating by the auction house; X_6 is the dummy for the transmission type; X_7 is the dummy for the fuel type; X_8 is the dummy for the popular colors; X_9 and X_{10} are the dummy variables for the options. Table 2 presents summary statistics for some of these covariates and several order statistics over the reserve price.

We estimate $l(x)$ separately for each car make such that $l(x) = l_m(x) = x' \beta^{(m)}$ where m denotes the car make and obtain the estimate of pseudo valuations as residuals. Table A1 provides the first-stage estimates with the linear specification. The signs for the coefficient of age, transmission, engine size, and rating variables look reasonable and the coefficients are significant, while the interpretation of coefficient signs for the fuel type, mileage, colors, options, and title remaining is not so clear.

Based on the estimated pseudo valuations, in the second step, we estimate the distribution of valuations using three different pairs of order statistics $(2^{nd}, 4^{th})$, $(3^{rd}, 4^{th})$, and $(2^{nd}, 3^{rd})$. Figure 2 illustrates the estimated density function of valuations using the linear estimation in the first stage. With these three nonparametric estimates of the distribution, we conduct our nonparametric test of the IPV. Testing results using our test statistics defined in (21) are rather mixed depending on which distributions we compare but overall it does support the IPV. When we compare the estimated distribution of valuations using $(2^{nd}, 4^{th})$ vs. $(3^{rd}, 4^{th})$ order statistics, clearly we do not reject the IPV even under the 10% significance level. However, when the estimated distributions using $(2^{nd}, 4^{th})$ vs. $(2^{nd}, 3^{rd})$ order statistics are compared, we reject the IPV under the 5% significance while do not reject under the 1% level. Values for the test statistic are provided in Table 3.

As we note in Figure 2 and Table 3, the estimated distribution using the pair of the 2^{nd} - and the 3^{rd} -highest order statistics is quite different from the others. Moreover, we note that the estimation results using the pair of $(2^{nd}, 3^{rd})$ order statistics are substantially varying over different settings (e.g. using different approximation order K and different initial values for estimation). We find that in the data, the differences between 2^{nd} - and the 3^{rd} -highest order statistics are quite small in many

observations, which caused the numerical instability. Note that by construction of the conditional density $p_{k_1|k_2}(\tilde{v}|v)$ in (2), identification of the distribution of valuations becomes stronger when the differences between order statistics are larger. From this observation we conclude that the testing results based on the estimated distributions using $(2^{nd}, 4^{th})$ vs. $(3^{rd}, 4^{th})$ order statistics are more reliable, which clearly support the IPV. Note, however, that a high value of the test statistic comparing the estimated distributions from $(2^{nd}, 4^{th})$ vs. $(2^{nd}, 3^{rd})$ order statistics illustrates the power of our proposed test. Table 4 reports additional testing results by varying the approximation order K in the SNP density and also the testing parameter γ .¹⁸

8 Discussions and Implications

8.1 Discussion of the Testing Result

In this section, we discuss the result of our IPV test. With our estimates, we do not reject the null hypothesis of IPV in our auction. We can interpret this result as an evidence of no informational dependency among bidders about the valuations of observed characteristics of a used-car and no effect of any unobserved (to an econometrician) characteristics on valuations. Our conjecture is that this situation may arise because each dealer operates in her own local market and there is no interdependency among those markets, or when a dealer has a specific demand (order) from a final buyer on hand. This can happen when a dealer gets an order for a specific used-car but does not have one in stock or when a consumer looks up the item list of an auction and asks a dealer to buy a specific car for her.

If the assumption of IPV were rejected, it could have been due to a violation of the independence assumption or a violation of the private value assumption. This could happen when a dealer does not have a specific demand on hand, but she anticipates some demand in near future from the analysis of the overall performance of the national market since every used-car won in the auction is to be resold to a final consumer. In this case, a dealer may have some incentives to find out other dealers' opinions about the prospect of the national market. With our data, we find no evidence to support this hypothesis.

¹⁸Table 3 is the case with $\gamma = 1/2$. We found similar results with different values of γ unless it is too small (see Table 4).

8.2 Bounds Estimation

We have disregarded the minimum increment of around 30 dollars in WUCA. In this section, we discuss how to obtain the bounds of the distribution of valuations incorporating the minimum increment. The bounds considered here is much simpler than those considered in Haile and Tamer (2003), since in WUCA any order statistic of valuations other than the first highest one is bounded as

$$b_{(i:n)} \leq v_{(i:n)} \leq b_{(i:n)} + \Delta, \text{ for all } i = 1, \dots, n-1, \quad (25)$$

where $(i:n)$ denotes the i^{th} order statistic out of the n sample and Δ depicts the minimum increment. By the first-order stochastic dominance, noting $G_{b_{(i:n)}+\Delta}(v) = G_{b_{(i:n)}}(v - \Delta)$, (25) implies

$$G_{b_{(i:n)}}(v) \geq G_{v_{(i:n)}}(v) \geq G_{b_{(i:n)}}(v - \Delta),$$

where $G(\cdot)$ denotes the distribution of the order statistics. Then, using the identification method discussed in Section 4, we have

$$F_b(v) \geq F_v(v) \geq F_b(v - \Delta) \cong F_b(v) - f_b(v)\Delta,$$

where $F_b(\cdot)$ and $f_b(\cdot)$ are the CDF and PDF, respectively, of valuations based on observed bids. The last weak equality comes from the first-order Taylor series expansion. Therefore, we can estimate the bounds of $F_v(v)$ as

$$\hat{F}_b(v) \geq F_v(v) \geq \hat{F}_b(v) - \hat{f}_b(v)\Delta,$$

where $\hat{f}_b(\cdot)$ the SNP estimator based on the certain observed order statistics of bids, $\hat{F}_b(x) = \int_{\min(b)}^x \hat{f}_b(v)dv$ and $\min(b)$ is the minimum among the observed bids considered.

8.3 Combining Several Order Statistics

Once we show the several versions of estimates for the distribution of valuations are statistically not different each other, we may obtain a more efficient estimate by combining them. One way to do this is to consider the joint density function of two or more order statistics conditional on a certain order statistic. Suppose we have the k_1^{th} , k_2^{th} , and k_3^{th} -highest order statistics, which are the

$(n - k_1 - 1)^{th}$, $(n - k_2 - 1)^{th}$, $(n - k_3 - 1)^{th}$ order statistics respectively ($1 \leq k_1 < k_2 < k_3 \leq n$). Denote the joint density of these three order statistics as $\tilde{g}^{(k_1, k_2, k_3; n)}(\cdot)$ - we use this notation $\tilde{g}^{(\cdot)}$ to distinguish it from $g^{(\cdot)}$ so that k_i denotes the k_i^{th} highest order statistics:

$$\begin{aligned} \tilde{g}^{(k_1, k_2, k_3; n)}(\tilde{v}, \tilde{v}, v) &= \frac{n!}{(n - k_3)!(k_3 - k_2 - 1)!(k_2 - k_1 - 1)!(k_1 - 1)!} \\ &\times F(v)^{n - k_3} f(v) [F(\tilde{v}) - F(v)]^{k_3 - k_2 - 1} f(\tilde{v}) [F(v) - F(\tilde{v})]^{k_2 - k_1 - 1} f(\tilde{v}) [1 - F(\tilde{v})]^{k_1 - 1}, \end{aligned}$$

where \tilde{V} denotes the k_1^{th} -, \tilde{V} denotes the k_2^{th} -, and V denotes the k_3^{th} -highest order statistics. Using this joint density function together with the marginal density (1), we obtain the conditional joint density of the k_1^{th} and the k_2^{th} -highest order statistics conditional on the k_3^{th} -highest statistics as

$$\begin{aligned} &P_{(k_1, k_2) | k_3}(\tilde{v}, \tilde{v} | v) \\ &= \frac{(k_3 - 1)!}{(k_3 - k_2 - 1)!(k_2 - k_1 - 1)!(k_1 - 1)!} \frac{[F(\tilde{v}) - F(v)]^{k_3 - k_2 - 1} f(\tilde{v}) [F(v) - F(\tilde{v})]^{k_2 - k_1 - 1} f(\tilde{v}) [1 - F(\tilde{v})]^{k_1 - 1}}{[1 - F(v)]^{k_3 - 1}} \\ &= \frac{(k_3 - 1)!}{(k_3 - k_2 - 1)!(k_2 - k_1 - 1)!(k_1 - 1)!} [(1 - F(v)) F(\tilde{v} | v)]^{k_3 - k_2 - 1} \\ &\quad \cdot f(\tilde{v} | v) (1 - F(v)) [(F(\tilde{v} | v) - F(\tilde{v} | v)) (1 - F(v))]^{k_2 - k_1 - 1} \cdot \frac{f(\tilde{v} | v) (1 - F(v)) [(1 - F(\tilde{v} | v)) (1 - F(v))]^{k_1 - 1}}{[1 - F(v)]^{k_3 - 1}} \\ &= \frac{(k_3 - 1)!}{(k_3 - k_2 - 1)!(k_2 - k_1 - 1)!(k_1 - 1)!} F(\tilde{v} | v)^{k_3 - k_2 - 1} f(\tilde{v} | v) \\ &\quad \cdot [F(\tilde{v} | v) - F(\tilde{v} | v)]^{(k_3 - k_1) - (k_3 - k_2) - 1} f(\tilde{v} | v) [1 - F(\tilde{v} | v)]^{(k_3 - 1) - (k_3 - k_1)} \\ &= g^{(k_3 - k_1, k_3 - k_2; k_3 - 1)}(\tilde{v}, \tilde{v} | v), \end{aligned} \tag{26}$$

where $F(\cdot | v)$ and $f(\cdot | v)$ ($g(\cdot | v)$) are the truncated CDF and PDF's truncated at v respectively. The last equality comes from the joint density of the j -th and i -th order statistics ($n \geq j > i \geq 1$),

$$g^{(j, i; n)}(a, b) = \frac{n! [F(b)]^{i-1} [F(a) - F(b)]^{j-i-1} [1 - F(a)]^{n-j} f(b) f(a)}{(i-1)!(j-i-1)!(n-j)!} I_{\{a > b\}}$$

where a and b are the j -th and i -th order statistics, respectively by letting $j = k_3 - k_1$, $i = k_3 - k_2$, and $n = k_3 - 1$. Therefore we can interpret $p_{(k_1, k_2) | k_3}(\cdot)$ as the joint density of $(k_3 - k_1)^{th}$ and $(k_3 - k_2)^{th}$ order statistics from a sample of size equal to $(k_3 - 1)$. When $(k_1, k_2, k_3) = (2, 3, 4)$, (26) becomes

$$p_{(2,3) | 4}(\tilde{v}, \tilde{v} | v) = 6 f(\tilde{v}) f(\tilde{v}) [1 - F(\tilde{v})] [1 - F(v)]^{-3}. \tag{27}$$

Based on (27), one can estimate the distribution of valuations $f(\cdot)$ similarly with the method proposed in Section 5.1. The resulting estimator will be more *efficient* than $\hat{f}_{2|4}(\cdot)$ or $\hat{f}_{3|4}(\cdot)$ in the sense that it uses more information.

9 Conclusions

In this paper we developed a practical nonparametric test of symmetric IPV in ascending auctions when the number of potential bidders is unknown. Our key idea for testing is that because any pair of order statistics can identify the value distribution, three or more order statistics can identify two or more value distributions and under the symmetric IPV, those value distributions should be identical. Therefore, testing the symmetric IPV is equivalent to testing the similarity of value distributions obtained from different pairs of order statistics. This extends the testability of the symmetric IPV of Athey and Haile (2002) that assume the known number of potential bidders to the case when the number is unknown.

Using the proposed methods we conducted a structural analysis of ascending-price auctions using a new data set on a wholesale used-car auction, in which there is no jump bidding. Exploiting the data, we estimated the distribution of bidders' valuations nonparametrically within the IPV paradigm. We can implement our test because the data enables us to exploit information from observed losing bids. We find that the null hypothesis of symmetric IPV is supported with our data after controlling for observed auction heterogeneity, and therefore our estimation result remains a valid approximation of the distribution of dealers' valuations. The richness of our data has allowed us to conduct a structural analysis that bridges the gap between theoretical models based on Milgrom and Weber (1982) and real-world ascending-price auctions.

In this paper, we have considered the auction as a collection of isolated single-object auctions. In future work, we will look at the data more closely in the alternative environments. For example, we will examine intra-day dynamics of auctions with daily budget constraints for bidders, or possible complementarity and substitutability in a multi-object auction environment. Another possibility is to consider an asymmetric IPV paradigm noting that information of winning bidders' identities is available in the data.

Tables and Figures

TABLE 1: Market Shares in the Sample

	Hyundai	Daewoo	Kia	Ssangyong	Others
Share (%)	44.72	30.86	21.05	2.66	0.71

TABLE 2: Summary Statistics (Sample Size: 5184)

	Age (years)	Engine Size (ℓ)	Mileage (km)	Rating (0-10)	2nd highest bid	3rd -	4th -	Reserve price	Opening bid
Mean	5.9278	1.7174	96888	3.1182	3812.1	3758.8	3690.5	3402.7	3102.5
S.D.	2.3689	0.3604	49414	1.2127	3004.8	2999.7	2992.1	2905.4	2866.1
Median	6.1667	1.4980	94697	3.5	3200	3150	3085	2800	2500
Max	12.75	3.4960	426970	7	29640	29580	29490	27800	27000
Min	0.1667	1.1390	69	0	100	70	70	50	0

Note: All prices are in 1000 Korean Won (1000 Korean Won \approx 1 US Dollar).

TABLE 3: Test Statistics

Specification	(2nd-4th) vs. (3rd-4th)	(2nd-4th) vs. (2nd-3rd)
Values	0.7381	4.8835

Note: Cutoff points for $\chi^2(1)$: 6.635 (1% level), 3.841 (5% level), 2.706 (10% level)

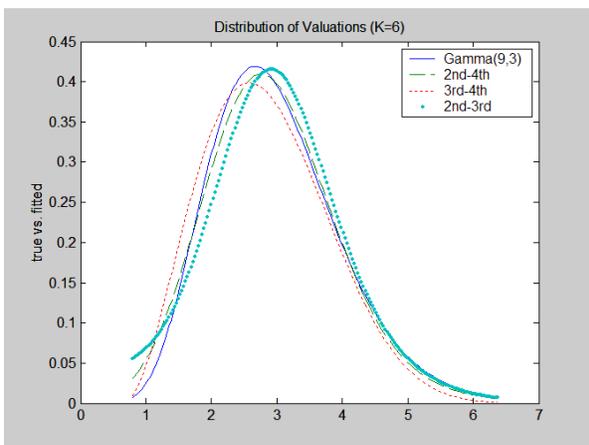


FIGURE 1: Estimated density function with $K = 6$ for a Monte Carlo simulated data

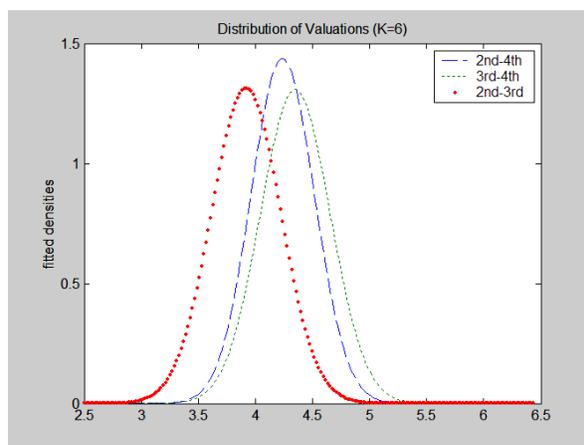


FIGURE 2: Estimated density function with $K = 6$ for the whole-sale used-car auction data

TABLE 4: Test Statistics Comparing Distributions of Valuations from (2nd-4th) vs. (3rd-4th)

Specification of the SNP	K=0	K=2	K=4	K=6
Testing para (γ)				
$\gamma = 0.2$	1.5971	1.7245	1.6610	1.3494
$\gamma = 0.4$	0.9703	1.0098	0.9845	0.8273
$\gamma = 0.5$	0.8635	0.8897	0.8703	0.7381
$\gamma = 0.6$	0.7958	0.8138	0.7981	0.6814
$\gamma = 0.8$	0.7151	0.7237	0.7122	0.6138

Note: Cutoff points for $\chi^2(1)$: 6.635 (1% level), 3.841 (5% level), 2.706 (10% level)

References

- [1] Ai, C. and X. Chen (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, v71, No.6, 1795-1843.
- [2] Arnold, B., Balakrishnan, N., and Nagaraja, H. (1992), *A First Course in Order Statistics*. (New York: John Wiley & Sons.)
- [3] Athey, S. and Haile, P. (2002), "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107-2140.
- [4] Athey, S. and Haile, P. (2005), "Nonparametric Approaches to Auctions," forthcoming in J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Elsevier.
- [5] Athey, S. and Levin, J. (2001), "Information and Competition in U.S. Forest Service Timber Auctions," *Journal of Political Economy*, 109, 375-417.
- [6] Bikhchandani, S., Haile, P. and Riley, J. (2002), "Symmetric Separating Equilibria in English Auctions," *Games and Economic Behavior*, 38, 19-27.
- [7] Bikhchandani, S. and Riley, J. (1991), "Equilibria in Open Common Value Auctions," *Journal of Economic Theory*, 53, 101-130.
- [8] Bikhchandani, S. and Riley, J. (1993), "Equilibria in Open Auctions," mimeo, UCLA.
- [9] Campo, S., Perrigne, I., and Vuong, Q. (2003) "Asymmetry in First-Price Auctions With Affiliated Private Values," *Journal of Applied Econometrics*, 18, 179-207.
- [10] Chen, X. (2007), "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, V.6.
- [11] Coppejans, M. and Gallant, A. R. (2002), "Cross-validated SNP density estimates," *Journal of Econometrics*, 110, 27-65.
- [12] David, H. A. (1981), *Order Statistics*. Second Edition. (New York: John Wiley & Sons.)

- [13] Fenton V. M. and Gallant, A. R. (1996), “Convergence Rates of SNP Density Estimators,” *Econometrica*, 64, 719-727.
- [14] Gallant, A. R. and Nychka, D. (1987), “Semi-nonparametric Maximum Likelihood Estimation,” *Econometrica*, 55, 363-390.
- [15] Gallant, A. R. and G. Souza (1991), “On the Asymptotic Normality of Fourier Flexible Form Estimates,” *Journal of Econometrics*, 50, 329-353.
- [16] Guerre, E., Perrigne, I., and Vuong, Q. (2000), “Optimal Nonparametric Estimation of First-Price Auctions,” *Econometrica*, 68, 525-574.
- [17] Guerre, E., Perrigne, I., and Vuong, Q. (2009), “Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions,” *Econometrica*, 77(4), 1193-1228.
- [18] Haile, P., Hong, H., and Shum, M. (2003), “Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions,” NBER Working Paper No. 10105.
- [19] Haile, P. and Tamer, E. (2003), “Inference with an Incomplete Model of English Auctions,” *Journal of Political Economy*, 111(1), 1-51.
- [20] Hansen, B. (2004), “Nonparametric Conditional Density Estimation,” Working paper, University of Wisconsin-Madison.
- [21] Harstad, R. and Rothkopf, M. (2000), “An “Alternating Recognition” Model of English Auctions,” *Management Science*, 46, 1-12.
- [22] Hendricks, K. and Paarsch, H. (1995), “A Survey of Recent Empirical Work concerning Auctions,” *Canadian Journal of Economics*, 28, 403-426.
- [23] Hendricks, K., Pinkse, J., and Porter, R. (2003), “Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions,” *Review of Economic Studies*, 70(1), 115-146.
- [24] Hoeffding, W. (1963), “Probability Inequalities for Sums of Bounded Random Variables,” *Journal of the American Statistical Association*, 58, 13-30.
- [25] Hong, H., and Shum, M. (2003), “Econometric Models of Asymmetric Ascending Auctions,” *Journal of Econometrics*, 112, 327-358.
- [26] Hong, Y.-M. and White, H. (2005), “Asymptotic Distribution Theory for an Entropy-Based Measure of Serial Dependence,” *Econometrica*, 73, 837-902.
- [27] Hu, Y., D. McAdams, and M. Shum (2013), “Identification of First-Price Auction Models with Non-Separable Unobserved Heterogeneity,” *Journal of Econometrics*, 174(2), 186-193.
- [28] Izmalkov, S. (2003), “English Auctions with Reentry,” mimeo, MIT.
- [29] Krasnokutskaya, E. (2011), “Identification and Estimation of Auction Models with Unobserved Heterogeneity,” *Review of Economic Studies*, 78(1), 293-327.
- [30] Kim, K. (2007), “Uniform Convergence Rate of the SNP Density Estimator and Testing for Similarity of Two Unknown Densities,” *Econometrics Journal*, 10(1), 1-34.

- [31] Kullback, S. and Leibler, R.A. (1951), "On Information and Sufficiency," *Ann. Math. Stat.*, 22, 79-86.
- [32] Laffont, J.-J. (1997), "Game Theory and Empirical Economics: The Case of Auction Data," *European Economic Review*, 41, 1-35.
- [33] Larsen, B. (2012), "The Efficiency of Dynamic, Post-Auction Bargaining: Evidence from Wholesale Used-Auto Auctions," MIT job market paper.
- [34] Li, Q. (1996), "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15, 261-274.
- [35] Li, T., Perrigne, I., and Vuong, Q. (2000), "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129-161.
- [36] Li, T., Perrigne, I., and Vuong, Q. (2002), "Structural Estimation of the Affiliated Private Value Auction Model," *RAND Journal of Economics*, 33, 171-193.
- [37] Lorentz, G. (1986), *Approximation of Functions*, New York: Chelsea Publishing Company.
- [38] Matzkin, R. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339-1375.
- [39] Milgrom, P. and Weber, R. (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089-1122.
- [40] Myerson, R. (1981), "Optimal Auction Design," *Math. Operations Res.*, 6, 58-73.
- [41] Newey, W. (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.
- [42] Newey, W., J. Powell and F. Vella (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-603.
- [43] Paarsch, H. (1992), "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191-215.
- [44] Riley, J. (1988), "Ex Post Information in Auctions," *Review of Economic Studies*, 55, 409-429.
- [45] Roberts, J (2013), "Unobserved Heterogeneity and Reserve Prices in Auctions," forthcoming in *RAND Journal of Economics*.
- [46] Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437-453.
- [47] Song, U. (2005), "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders," Economics Department Discussion Paper 05-14, University of British Columbia.
- [48] Ullah, A. and Singh, R.S. (1989), "Estimation of a Probability Density Function with Applications to Nonparametric Inference in Econometrics," in B. Raj (ed.) *Advances in Econometrics and Modelling*, Kluwer Academic Press, Holland.

Appendix

FIGURE A1: Sample Bidding Log

(Opening bid: 1700, Reserve price: 2000, Transaction price: 2210, Total bidders logged: 5)

Price \ ID	025110	062101	069702	075401	076101	Total
1700					1	1
1730					1	1
1760					1	1
1790	1				1	2
1820	1				1	2
1850					1	1
1880					1	1
1910					1	1
1940			1		1	2
1970		1	1			2
2000		1			1	2
2030		1	1	1	1	4
2060	1			1	1	3
2090		1			1	2
2120	1	1				2
2150	1				1	2
2180				1	1	2
2210					1	1
Total	5	5	3	3	17	

TABLE A1: First-Stage Estimation Results (Linear model)

Maker	Covariate	Estimate	Standard Error
Hyundai (T=1676)	Intercept	5.2247	0.0945
	Age	-0.2439	0.0058
	Engine Size (l)	0.7760	0.0367
	Mileage ($10^4 km$)	0.0004	0.0022
	Rating	0.1241	0.0082
	Automatic	0.2874	0.0192
	Gasoline	0.0155	0.0381
	Popular Colors	0.0567	0.0184
	Best Options	0.2441	0.0316
	Better Options	0.1178	0.0210
Daewoo (T=1186)	Intercept	5.2821	0.0991
	Age	-0.3369	0.0062
	Engine Size (l)	0.9470	0.0479
	Mileage ($10^4 km$)	-0.0108	0.0027
	Rating	0.0508	0.0101
	Automatic	0.2430	0.0215
	Gasoline	0.3306	0.0374
	Popular Colors	-0.0551	0.0191
	Best Options	0.1344	0.0299
	Better Options	0.0244	0.0218
Kia & etc. (T=964)	Intercept	5.2698	0.1082
	Age	-0.2234	0.0070
	Engine Size (l)	0.5731	0.0390
	Mileage ($10^4 km$)	0.0054	0.0026
	Rating	0.1004	0.0111
	Automatic	0.3347	0.0247
	Gasoline	-0.1943	0.0414
	Popular Colors	0.0278	0.0225
	Best Options	0.3947	0.0424
	Better Options	0.2062	0.0273

A Normalized Hermite Polynomials

Here are examples of the normalized Hermite polynomials that we use to approximate the density of valuations. These are generated by the recursive system in (8). Note that $\int_{-\infty}^{\infty} H_j^2 dz = 1$ and $\int_{-\infty}^{\infty} H_j H_k dz = 0$, $j \neq k$ by construction.

$$\begin{aligned}
 H_1 &= \frac{1}{\sqrt[4]{2\pi}} e^{-\frac{1}{4}z^2}, \quad H_2 = \frac{z}{\sqrt[4]{2\pi}} e^{-\frac{1}{4}z^2} \\
 H_3 &= \frac{1}{2\sqrt[4]{2\pi}} (z^2\sqrt{2} - \sqrt{2}) e^{-\frac{1}{4}z^2}, \quad H_4 = \frac{1}{6\sqrt[4]{2\pi}} (z^3\sqrt{6} - 3z\sqrt{6}) e^{-\frac{1}{4}z^2} \\
 H_5 &= \frac{1}{12\sqrt[4]{2\pi}} (3\sqrt{6} - 6z^2\sqrt{6} + z^4\sqrt{6}) e^{-\frac{1}{4}z^2} \\
 H_6 &= \frac{1}{60\sqrt[4]{2\pi}} (15z\sqrt{30} - 10z^3\sqrt{30} + z^5\sqrt{30}) e^{-\frac{1}{4}z^2} \\
 H_7 &= \frac{1}{60\sqrt[4]{2\pi}} (45z^2\sqrt{5} - 15\sqrt{5} - 15z^4\sqrt{5} + z^6\sqrt{5}) e^{-\frac{1}{4}z^2}.
 \end{aligned}$$

B Nonparametric Extension: the Observed Heterogeneity

In the main text we have assumed that $l(\cdot)$ belongs to a parametric family. Here we consider a nonparametric specification of $l(\cdot)$. We can approximate the unknown function $l(X_t)$ in (3) using a sieve such as power series or splines (see e.g. Chen 2007). We approximate the function space \mathcal{L} containing the true $l(\cdot)$ with the following power series sieve space \mathcal{L}_T

$$\mathcal{L}_T = \{l(X)|l(X) = R^{k_1(T)}(X)' \pi \text{ for all } \pi \text{ satisfying } \|l\|_{\Lambda^{\gamma_1}} \leq c_1\},$$

where $R^{k_1}(X)$ is a triangular array of some basis polynomials with the length of k_1 . Here $\|\cdot\|_{\Lambda^\gamma}$ denotes the Hölder norm:

$$\|g\|_{\Lambda^\gamma} = \sup_x |g(x)| + \max_{a_1+a_2+\dots+a_{d_x}=\gamma} \sup_{x \neq x'} \frac{|\nabla^a g(x) - \nabla^a g(x')|}{(\|x - x'\|_E)^{\gamma-\underline{\gamma}}} < \infty,$$

where $\nabla^a g(x) \equiv \frac{\partial^{a_1+a_2+\dots+a_{d_x}} g(x)}{\partial x_1^{a_1} \dots \partial x_{d_x}^{a_{d_x}}}$ with $\underline{\gamma}$ the largest integer smaller than γ . The Hölder ball (with radius c) $\Lambda_c^\gamma(\mathcal{X})$ is defined accordingly as

$$\Lambda_c^\gamma(\mathcal{X}) \equiv \{g \in \Lambda^\gamma(\mathcal{X}) : \|g\|_{\Lambda^\gamma} \leq c < \infty\}$$

and we take the function space $\mathcal{L} \equiv \Lambda_{c_1}^{\gamma_1}(\mathcal{X})$. Functions in $\Lambda_c^\gamma(\mathcal{X})$ can be approximated well by various sieves such as power series, Fourier series, splines, and wavelets.

The functions in \mathcal{L}_T is getting dense as $T \rightarrow \infty$ but not that fast, i.e. $k_1 \rightarrow \infty$ as $T \rightarrow \infty$ but $k_1/T \rightarrow 0$. Then, according to Theorem 8, p.90 in Lorentz (1986), there exists a π_{k_1} such that for $R^{k_1}(x)$ on the compact set \mathcal{X} (the support of X) and a constant c_1 ¹⁹

$$\sup_{x \in \mathcal{X}} |l(x) - R^{k_1}(x)' \pi_{k_1}| < c_1 k_1^{-\frac{[\gamma_1]}{d_x}}, \quad (28)$$

where $[s]$ is the largest integer less than s and d_x is the dimension of X . Thus, we approximate the pseudo-value V_{ti} in (3) as

$$V_{ti} = \ln \mathbb{V}_{ti} - l_{k_1}(X_t), \quad (29)$$

where $l_{k_1}(x) = R^{k_1}(x)' \pi_{k_1}$.

Specifically, we consider the following polynomial basis considered by Newey, Powell, and Vella (1999). First let $\mu = (\mu_1, \dots, \mu_{d_x})'$ denote a vector of nonnegative integers with the norm $|\mu| = \sum_{j=1}^{d_x} \mu_j$, and let $x^\mu \equiv \prod_{j=1}^{d_x} (x_j)^{\mu_j}$. For a sequence $\{\mu(k)\}_{k=1}^\infty$ of distinct vectors, we construct a tensor-product power series sieve as $R^{k_1}(x) = (x^{\mu(1)}, \dots, x^{\mu(k_1)})'$. Then, replacing each power x^μ by the product of orthonormal univariate polynomials of the same order, we may reduce collinearity. Note that the approximation precision depends on the choice of smoothing parameter k_1 . We discuss how to pick the optimal length of series k_1^* in Appendix C.

C Choosing the optimal approximation parameters

Instead of using the Leave-one-out method for cross-validation, we will partition the data into P groups, making the size of each group as equal as possible and use the *Leave-one partition-out* method. We do this because it will be computationally too expensive to use the Leave-one-out

¹⁹We often suppress the argument T in $k_1(T)$, unless otherwise noted.

method, when the sample size is large. We let T_p denote the set of the data indices that belongs to the p^{th} group such that $T_p \cap T_{p'} = \emptyset$ for $p \neq p'$ and $\cup_{p=1}^P T_p = \{1, 2, \dots, T\}$.

C.1 Choosing K^*

Coppejans and Gallant (2002) employ a cross-validation method based on the ISE (Integrated Squared Error) criteria. Let $h(x)$ be a density and $\hat{h}(x)$ be its estimator. Then the ISE is defined by

$$\mathbf{ISE}(\hat{h}) = \int \hat{h}^2(x)dx - 2 \int \hat{h}(x)h(x)dx + \int h(x)^2dx = M_{(1)} - 2M_{(2)} + M_{(3)}.$$

We present our approach in terms of fitting $p_{2|4}(\tilde{v}|v)$. We find the optimal K^* by minimizing an approximation of the ISE. To approximate the ISE in terms of $p_{2|4}(\tilde{v}|v)$, we use the cross-validation strategy with the data partitioned into P groups. We first approximate $M_{(1)}$ with

$$\widehat{M}_{(1)}(K) = \int (\hat{p}_{2|4}^K(\tilde{v}|v))^2 d(\tilde{v}, v) \equiv \int \left(\frac{6(\hat{F}_{(K)}(\tilde{v}) - \hat{F}_{(K)}(v))(\hat{F}_{(K)}(\bar{v}) - \hat{F}_{(K)}(\tilde{v}))\hat{f}_{(K)}(\tilde{v})}{(\hat{F}_{(K)}(\bar{v}) - \hat{F}_{(K)}(v))^3} \right)^2 d(\tilde{v}, v),$$

where $\hat{f}_{(K)}(\cdot)$ denotes the SNP estimate with the length of the series equal to K and $\hat{F}_{(K)}(z) = \int_{-\infty}^z \hat{f}_{(K)}(t)dt$. For $M_{(2)}$, we consider

$$\begin{aligned} \widehat{M}_{(2)}(K) &= \frac{1}{T} \sum_{p=1}^P \sum_{t \in T_p} \hat{p}_{2|4}^{p,K}(\tilde{v}_t|v_t) \\ &\equiv \frac{1}{T} \sum_{p=1}^P \sum_{t \in T_p} \frac{6(\hat{F}_{p,(K)}(\tilde{v}_t) - \hat{F}_{p,(K)}(v_t))(\hat{F}_{p,(K)}(\bar{v}) - \hat{F}_{p,(K)}(\tilde{v}_t))\hat{f}_{p,(K)}(\tilde{v}_t)}{(\hat{F}_{p,(K)}(\bar{v}) - \hat{F}_{p,(K)}(v_t))^3}, \end{aligned}$$

where $\hat{f}_{p,(K)}(\cdot)$ denotes the SNP estimate obtained from the sample excluding p^{th} group with the length of the series K and $\hat{F}_{p,(K)}(z) = \int_{-\infty}^z \hat{f}_{p,(K)}(v)dv$. Noting $M_{(3)}$ is not a function of K , we pick K^* by minimizing the approximated ISE such that

$$K^* = \underset{K}{\operatorname{argmin}} \widehat{M}_{(1)}(K) - 2\widehat{M}_{(2)}(K).$$

C.2 Choosing k_1^*

We can use a sample version of the Mean Squared Error criterion for the cross-validation as

$$\text{SMSE}(\hat{l}) = \frac{1}{T} \sum_{t=1}^T [\hat{l}(X_t) - l(X_t)]^2,$$

where $\hat{l}(\cdot) = R^{k_1}(\cdot)' \hat{\pi}$. We use the *Leave-one partition-out* method to reduce the computational burden. Namely, we estimate the function $l(\cdot)$ from the sample after deleting the p^{th} group with the length of the series equal to k_1 and denote this as $\hat{l}_{p,k_1}(\cdot)$. In the next step, we choose k_1^* by minimizing the SMSE such that²⁰

$$k_1^* = \underset{k_1}{\operatorname{argmin}} \frac{1}{T} \sum_{p=1}^P \sum_{t \in T_p} [\ln \mathbb{V}_t - \hat{l}_{p,k_1}(X_t)]^2 \quad (30)$$

²⁰Here we can find k_1^* for particular order statistics or find one for over all observed bids. In the latter case, we take the sum of the minimand in (30) over all observed bids.

noting that

$$\begin{aligned} \text{SMSE}(\hat{l}) &= \frac{1}{T} \sum_{t=1}^T [\hat{l}(X_t) - l(X_t)]^2 \\ &= \frac{1}{T} \sum_{t=1}^T [\hat{l}(X_t) - \ln \mathbb{V}_t]^2 + 2 \frac{1}{T} \sum_{t=1}^T [\hat{l}(X_t) - \ln \mathbb{V}_t][\ln \mathbb{V}_t - l(X_t)] + \frac{1}{T} \sum_{t=1}^T [\ln \mathbb{V}_t - l(X_t)]^2 \end{aligned} \quad (31)$$

where the second term of (31) has the mean of zero (assuming $E[V_t|X_t] = 0$) and the third term of (31) does not depend on k_1 .

D Alternative Characterization of the SNP Estimator

We consider an alternative characterization of the SNP density using the specification proposed in Kim (2007) where a truncated version of the SNP density estimator with a compact support is developed, which turns out to be convenient for deriving the large sample theories. Instead of defining the true density as (5), we follow Kim (2007)'s specification by denoting the true density as

$$f(z) = h_j^2(z)e^{-z^2/2} + \epsilon_0 \phi(z) / \int_{\mathcal{V}} \phi(z) dz$$

where \mathcal{V} denotes the support of z . Kim (2007) uses a truncated version of Hermite polynomials to approximate a density function f with a compact support as

$$\mathcal{F}_T = \left\{ f : f(z, \theta) = \left(\sum_{j=1}^K \vartheta_j w_{jK}(z) \right)^2 + \epsilon_0 \phi(z) / \int_{\mathcal{V}} \phi(z) dz, \theta \in \Theta_T \right\}, \quad (32)$$

where $\Theta_T = \left\{ \theta = (\vartheta_1, \dots, \vartheta_{K(T)}) : \sum_{j=1}^{K(T)} \vartheta_j^2 + \epsilon_0 = 1 \right\}$ and $\{w_{jK}(z)\}$ are defined below following Kim (2007). First we define $\bar{w}_{jK}(z) = H_j(z) / \sqrt{\int_{\mathcal{V}} H_j^2(z) dz}$ that is bounded by

$$\sup_{z \in \mathcal{V}, j \leq K} |\bar{w}_{jK}(z)| \leq \sup_{z \in \mathcal{V}} |H_j(z)| / \sqrt{\min_{j \leq K} \int_{\mathcal{V}} H_j^2(v) dv} < C\tilde{H}$$

for some constant $C < \infty$, because $\int_{\mathcal{V}} H_j^2(z) dz$ is bounded away from zero for all j and $|H_j(z)| < \tilde{H}$ uniformly over z and j . Denoting $\bar{W}^K(z) = (\bar{w}_{1K}(z), \dots, \bar{w}_{KK}(z))'$, further define $Q_{\bar{W}} = \int_{\mathcal{V}} \bar{W}^K(z) \bar{W}^K(z)' dz$ and its symmetric matrix square root as $Q_{\bar{W}}^{-1/2}$. Now let

$$W^K(z) \equiv (w_{1K}(z), \dots, w_{KK}(z))' \equiv Q_{\bar{W}}^{-1/2} \bar{W}^K(z). \quad (33)$$

Then by construction, we have $\int_{\mathcal{V}} W^K(z) W^K(z)' dz = I_K$. These truncated and transformed Hermite polynomials are orthonormal

$$\int_{\mathcal{V}} w_{jK}^2(z) dz = 1, \int_{\mathcal{V}} w_{jK}(z) w_{kK}(z) dz = 0, j \neq k$$

from which the condition $\sum_{j=1}^{K(n)} \vartheta_j^2 + \epsilon_0 = 1$ follows since for any f in \mathcal{F}_T , we have $\int_{\mathcal{V}} f dz = 1$. Now define $\zeta(K) = \sup_{z \in \mathcal{V}} \|W^K(z)\|$ using a matrix norm $\|A\| = \sqrt{\text{tr}(A'A)}$ for a matrix A , which is the Euclidian norm for a vector. Then, we have $\zeta(K) = O(\sqrt{K})$ as shown in Lemma G.1. If the

range of \mathcal{V} are sufficiently large, then $\int_{\mathcal{V}} H_j^2(z) dz \approx 1$ and $Q_{\overline{W}} \approx I_K$ and hence $w_{jK} \approx H_j$ which implies

$$\sup_{z \in \mathcal{V}} \|W^K(z)\| \approx \sup_{z \in \mathcal{V}} \sqrt{\sum_{j=1}^K H_j^2} \leq \sqrt{K \tilde{H}^2} = O(\sqrt{K}).$$

The SNP density estimator is now written as (e.g.)

$$\begin{aligned} \hat{f}_{2|4} &= \operatorname{argmax}_{f \in \underline{\mathcal{F}}_T} \frac{1}{T} \sum_{t=1}^T \ln L(f; \tilde{v}_t, v_t) \\ &\equiv \operatorname{argmax}_{f \in \underline{\mathcal{F}}_T} \frac{1}{T} \sum_{t=1}^T \ln \frac{6(F(\tilde{v}_t) - F(v_t))(1 - F(\tilde{v}_t))f(\tilde{v}_t)}{(1 - F(v_t))^3} \end{aligned} \quad (34)$$

where we define $F(z) = \int_{\underline{v}}^z f(z) dz$. The above maximization problem can be written equivalently

$$\hat{f}_{2|4} = f(\cdot, \hat{\theta}_K) \in \underline{\mathcal{F}}_T, \hat{\theta}_K = \operatorname{argmax}_{\theta \in \Theta_T} \frac{1}{T} \sum_{t=1}^T \ln L(f(\cdot, \theta_K); \tilde{v}_t, v_t).$$

E Large Sample Theory of the SNP Density Estimator

First we show the one-to-one mapping between functions in \mathcal{F}_T and $\underline{\mathcal{F}}_T$ defined by (7) and (32), respectively. Based on this equivalence, our asymptotic analyses will be on density functions in $\underline{\mathcal{F}}_T$.

E.1 Relationship between \mathcal{F}_T and $\underline{\mathcal{F}}_T$

Consider a specification of the SNP density estimator with unbounded support as in \mathcal{F}_T ,

$$f(z, \theta) = (H^{(K)}(z)' \theta)^2 + \epsilon_0 \phi(z)$$

where $H^{(K)}(z) = (H_1(z), H_2(z), \dots, H_K(z))$ and $H_j(z)$'s are the Hermite polynomials constructed recursively. Now consider a truncated version of the density on a truncated compact support \mathcal{V} ,

$$\bar{f}(z, \theta) = \frac{(H^{(K)}(z)' \theta)^2 + \epsilon_0 \phi(z)}{\int_{\mathcal{V}} (H^{(K)}(z)' \theta)^2 dz + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz}.$$

Let B be a $K \times K$ matrix whose elements are b_{ij} 's where $b_{ii} = \sqrt{\int_{\mathcal{V}} H_i(z)^2 dz}$ for $i = 1, \dots, K$ and $b_{ij} = 0$ for $i \neq j$. Recall that $\overline{W}^J(z) = B^{-1} H^{(K)}(z)$, $Q_{\overline{W}} = \int_{\mathcal{V}} \overline{W}^K(z) \overline{W}^K(z)' dz$, and $W^K(z) = Q_{\overline{W}}^{-1/2} \overline{W}^K(z) = Q_{\overline{W}}^{-1/2} B^{-1} H^{(K)}(z)$. Using those notations, we obtain

$$\begin{aligned} \bar{f}(z, \theta) &= \frac{(H^{(K)}(z)' \theta)^2 + \epsilon_0 \phi(z)}{\int_{\mathcal{V}} (H^{(K)}(z)' \theta)^2 dz + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz} = \frac{\theta' H^{(K)}(z) H^{(K)}(z)' \theta + \epsilon_0 \phi(z)}{\theta' \int_{\mathcal{V}} H^{(K)}(z) H^{(K)}(z)' dz \theta + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz} \\ &= \frac{\theta' B Q_{\overline{W}}^{1/2} Q_{\overline{W}}^{-1/2} B^{-1} H^{(K)}(z) H^{(K)}(z)' B^{-1} Q_{\overline{W}}^{-1/2} Q_{\overline{W}}^{1/2} B \theta + \epsilon_0 \phi(z)}{\theta' B \int_{\mathcal{V}} B^{-1} H^{(K)}(z) H^{(K)}(z)' B^{-1} dz B \theta + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz} \\ &= \frac{\theta' B Q_{\overline{W}}^{1/2} W^K(z) W^K(z)' Q_{\overline{W}}^{1/2} B \theta + \epsilon_0 \phi(z)}{\theta' B \int_{\mathcal{V}} \overline{W}^K(z) \overline{W}^K(z)' dz B \theta + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz} = \frac{\theta' B Q_{\overline{W}}^{1/2} W^K(z) W^K(z)' Q_{\overline{W}}^{1/2} B \theta + \epsilon_0 \phi(z)}{\theta' B Q_{\overline{W}}^{1/2} Q_{\overline{W}}^{1/2} B \theta + \epsilon_0 \int_{\mathcal{V}} \phi(z) dz}. \end{aligned}$$

Now let $\tilde{\theta} = Q_{\overline{W}}^{1/2} B\theta$ and $\epsilon_0 = \tilde{\epsilon}_0 / \int_{\mathcal{Y}} \phi(z) dz$. Then, we obtain

$$\bar{f}(z, \theta) = \frac{\tilde{\theta}' W^K(z) W^K(z)' \tilde{\theta} + \tilde{\epsilon}_0 \phi(z) / \int_{\mathcal{Y}} \phi(z) dz}{\tilde{\theta}' \tilde{\theta} + \tilde{\epsilon}_0} = \left(W^J(z)' \tilde{\theta} \right)^2 + \tilde{\epsilon}_0 \phi(z) / \int_{\mathcal{Y}} \phi(z) dz \quad (35)$$

by restricting $\tilde{\theta}' \tilde{\theta} + \tilde{\epsilon}_0 = 1$ such that $\tilde{\theta} \in \Theta_T$. Note that (35) coincides with the specification we consider in \mathcal{F}_n of (32). This illustrates why the proposed SNP estimator is a truncated version of the original SNP estimator with unbounded support. We also note that the relationship between the (truncated) parameters of the original density and the (truncated) parameters of the proposed specification is explicit as $\tilde{\theta} = Q_{\overline{W}}^{1/2} B\theta$.

E.2 Convergence Rate of the First Stage

We develop the convergence rate result when $l(\cdot)$ is nonparametrically specified, which nests the parametric case. Here we impose the following regularity conditions. Define the matrix norm $\|A\| = \sqrt{\text{trace}(A'A)}$ for a matrix A .

Assumption E.1 $\{(\mathbb{V}_t, X_t), \dots, (\mathbb{V}_T, X_T)\}$ are i.i.d. for all observed bids and $\text{Var}(\mathbb{V}_t|X)$ is bounded for all observed bids.

Assumption E.2 (i) the smallest and the largest eigenvalue of $E[R^{k_1}(X)R^{k_1}(X)']$ are bounded away from zero uniformly in k_1 ; (ii) there is a sequence of constants $\zeta_0(k_1)$ s.t. $\sup_{x \in \mathcal{X}} \|R^{k_1}(x)\| \leq \zeta_0(k_1)$ and $k_1 = k_1(T)$ such that $\zeta_0(k_1)^2 k_1 / T \rightarrow 0$ as $T \rightarrow \infty$.

Under Assumption E.1 and E.2 (which are variations of Assumption 1 and 2 in Newey (1997)), we obtain the convergence rate of $\hat{l}(\cdot) = R^{k_1}(\cdot)' \hat{\pi}$ to $l(\cdot)$ in the sup-norm by Theorem 1 of Newey (1997), because (28) implies Assumption 3 in Newey (1997) for the polynomial series approximation. Theorem 1 in Newey (1997) states that

$$\sup_{x \in \mathcal{X}} \left| \hat{l}(x) - l(x) \right| = O_p(\zeta_0(k_1)) \left[\sqrt{k_1} / \sqrt{T} + k_1^{-\frac{[\gamma_1]}{d_x}} \right].$$

Noting $\zeta_0(k_1) \leq O(k_1)$ for power series sieves, we have

$$\sup_{x \in \mathcal{X}} \left| \hat{l}(x) - l(x) \right| = \max \left\{ O_p(k_1^{3/2} / \sqrt{T}), O_p(k_1^{1-[\gamma_1]/d_x}) \right\} = O_p \left(T^{\max\{3\kappa/2-1/2, (1-[\gamma_1]/d_x)\kappa\}} \right) \quad (36)$$

with $k_1 = O(T^\kappa)$ and $0 < \kappa < \frac{1}{3}$.

The convergence rate result of (36) implies that by choosing a proper κ , we can make sure that the convergence rate of the second step density estimator will not be affected by the first step estimation. Comparing (36) and (49) - the result we obtain later in Section E.3, we find that it is required that for any small $\delta > 0$, $\frac{\max\{T^{3\kappa/2-1/2}, T^{(1-[\gamma_1]/d_x)\kappa}\}}{\max\{T^{-1/2+\alpha/2+\delta}, K(T)^{-s/2}\}} \rightarrow 0$ to ensure that the convergence rate of the second step density estimator is not affected by the first step estimation. Moreover, the asymptotic distributions of the second step density estimator or its functional is not affected by the first estimation step, which makes our discussion much easier. This is noted also in Hansen (2004) when he derives the asymptotic distribution of the two-step density estimator which contains the estimates of the conditional mean in the first step. This result will be immediately obtained if we

consider a parametric specification of observed auction heterogeneities since we will achieve the \sqrt{T} consistency for those finite dimensional parameters.

Based on this result we can obtain the convergence rate of the SNP density estimator as being abstract from the first stage.

E.3 Consistency and Convergence Rate of the SNP Estimator

We derive the convergence rate of the SNP density estimator obtained from any pair of order statistics. Denote (q_t, r_t) to be a pair of k_1 -th and k_2 -th highest order statistics. We derive the convergence rate of the SNP density estimator when the data $\{(q_t, r_t)\}_{t=1}^T$ are from the true values (after removing the observed heterogeneity part), not estimated ones from the first step estimation due to the reason described in the previous section. First, we construct the single observation Likelihood function

$$L(f; q_t, r_t) = \frac{(k_2 - 1)!}{(k_2 - k_1 - 1)!(k_1 - 1)!} \frac{(F(q_t) - F(r_t))^{k_2 - k_1 - 1} (1 - F(q_t))^{k_1 - 1} f(q_t)}{(1 - F(r_t))^{k_2 - 1}} \quad (37)$$

where $f \in \underline{\mathcal{F}}_T$ and $F(z) = \int_{\underline{v}}^z f(v) dv$. Then the SNP density estimator \hat{f} is obtained by solving

$$\hat{f} = \operatorname{argmax}_{f \in \underline{\mathcal{F}}_T} \frac{1}{T} \sum_{t=1}^T \ln L(f(\cdot, \theta); q_t, r_t) \quad (38)$$

or equivalently $\hat{f} = f(z, \hat{\theta}_K) \in \underline{\mathcal{F}}_T$, $\hat{\theta}_K = \operatorname{argmax}_{\theta \in \Theta_T} \frac{1}{T} \sum_{t=1}^T \ln L(f(\cdot, \theta); q_t, r_t)$.

From now on, we will denote the true density function, by $f_0(\cdot)$. To establish the convergence rate, we need the following assumptions

Assumption E.3 *The observed data of a pair of order statistics $\{(q_t, r_t)\}$ are randomly drawn from the continuous density of the parent distribution $f_0(\cdot)$.*

Assumption E.4 *(i) $f_0(z)$ is s -times continuously differentiable with $s \geq 3$, (ii) $f_0(z)$ is uniformly bounded from above and bounded away from zero on its compact support \mathcal{V} , (iii) $f(z)$ has the form of $f_0(z) = h_{f_0}^2(z) e^{-z^2/2} + \epsilon_0 \phi(z) / \int_{\mathcal{V}} \phi(z) dz$ for arbitrary small positive number ϵ_0 .*

Note that differently from Fenton and Gallant (1996) and Coppejans and Gallant (2002), we do not require a tail condition since we impose the compact truncated support.

Now recall that we denote (q_t, r_t) to be a pair of k_1 -th and k_2 -th highest order statistics ($k_1 < k_2$). We derive the convergence rate of the SNP density estimator when the data $\{(q_t, r_t)\}_{t=1}^T$ are from the true residual values (after removing the observed heterogeneity part). Though we use a particular sieve here, we derive the convergence rate results for a general sieve that satisfies some conditions.

First we derive the approximation order rate of the true density using the SNP density approximation in $\underline{\mathcal{F}}_T$. According to Theorem 8, p.90, in Lorentz (1986), we can approximate a v -times continuously differentiable function h such that there exists a K -vector γ_K that satisfies

$$\sup_{z \in \mathcal{Z}} |h(z) - R^K(z)' \gamma_K| = O(K^{-\frac{v}{d_z}}) \quad (39)$$

where $\mathcal{Z} \subset \mathbb{R}^{d_z}$ is the compact support of z and $R^K(z)$ is a triangular array of polynomials. Now note $f_0(z) = h_{f_0}^2(z) e^{-z^2/2} + \epsilon_0 \frac{\phi(z)}{\int_{\mathcal{V}} \phi(z) dz}$ and assume $h_{f_0}(z)$ (and hence $f_0(z)$) is s -times continuously

differentiable. Denote a K -vector $\theta_K = (\vartheta_{1K}, \dots, \vartheta_{KK})'$. Then, there exists a θ_K such that

$$\sup_{z \in \mathcal{V}} \left| h_{f_0}(z) - e^{z^2/4} W^K(z)' \theta_K \right| = O(K^{-s}) \quad (40)$$

by (39) noting $h_{f_0}(z)$ is s -times continuously differentiable over $z \in \mathcal{V}$, \mathcal{V} is compact, and $\{e^{z^2/4} w_{jK}(z)\}$ are linear combinations of power series. (40) implies that

$$\sup_{z \in \mathcal{V}} \left| h_{f_0}(z) e^{-z^2/4} - W^K(z)' \theta_K \right| \leq \sup_{z \in \mathcal{V}} e^{-z^2/4} \sup_{z \in \mathcal{V}} \left| h_{f_0}(z) - e^{z^2/4} W^K(z)' \theta_K \right| = O(K^{-s}) \quad (41)$$

from $\sup_z e^{-z^2/4} \leq 1$. Now we show that for $f(z, \theta_K) \in \underline{\mathcal{F}}_T$, $\sup_{z \in \mathcal{V}} |f(z) - f(z, \theta_K)| = O(\zeta(K)K^{-s})$. First, note (41) implies

$$W^K(z)' \theta_K - O(K^{-s}) \leq h_{f_0}(z) e^{-z^2/4} \leq W^K(z)' \theta_K + O(K^{-s})$$

from which it follows that

$$\begin{aligned} (W^K(z)' \theta_K - O(K^{-s}))^2 - (W^K(z)' \theta_K)^2 &\leq h_{f_0}^2(z) e^{-x^2/2} - (W^K(z)' \theta_K)^2 \\ &\leq (W^K(z)' \theta_K + O(K^{-s}))^2 - (W^K(z)' \theta_K)^2 \end{aligned} \quad (42)$$

assuming $W^K(z)' \theta_K$ is positive without loss of generality. Now, note that

$$\sup_{z \in \mathcal{V}} |W^K(z)' \theta| \leq \sup_{z \in \mathcal{V}} \|W^K(z)\| \|\theta\| = O(\zeta(K)) \quad (43)$$

by the Cauchy-Schwarz inequality and from $\|\theta\|^2 < 1$ for any $\theta \in \Theta_T$ by construction. Now applying the mean value theorem to the upper bound of (42), we have

$$\begin{aligned} \sup_{z \in \mathcal{V}} \left| (W^K(z)' \theta_K + O(K^{-s}))^2 - (W^K(z)' \theta_K)^2 \right| &= \sup_{z \in \mathcal{V}} \left| (2W^K(z)' \theta_K + O(K^{-s})) O(K^{-s}) \right| \\ &\leq \sup_{z \in \mathcal{V}} \left| 2W^K(z)' \theta_K \right| O(K^{-s}) + O(K^{-2s}) = O(\zeta(K)K^{-s}) \end{aligned}$$

where the last result is from (43). Similarly for the lower bound, we obtain

$$\sup_{z \in \mathcal{V}} \left| (W^K(z)' \theta_K - O(K^{-s}))^2 - (W^K(z)' \theta_K)^2 \right| = O(\zeta(K)K^{-s}).$$

From (42), it follows that $\sup_{z \in \mathcal{V}} \left| h_{f_0}^2(z) e^{-z^2/2} - (W^K(z)' \theta_K)^2 \right| = O(\zeta(K)K^{-s})$ and hence

$$\sup_{z \in \mathcal{V}} |f_0(z) - f(z, \theta_K)| = O(\zeta(K)K^{-s}).$$

Now to establish the convergence rate of the SNP estimator we define a pseudo true density function. We also introduce a trimming device $\tau(q_t, r_t)$ (an indicator function) that excludes those observations near the lower and the upper bounds of the support $q_t, r_t \geq \bar{v} - \varepsilon$ and $q_t, r_t \leq \underline{v} + \varepsilon$ in case one may want to use the trimming in the SNP density estimation. We denote this trimmed support as $\mathcal{V}_\varepsilon = [\underline{v} + \varepsilon, \bar{v} - \varepsilon]$. The pseudo true density is given by

$$f_K^*(z) = (W^K(z)' \theta_K^*)^2 + \epsilon_0 \phi(z) / \int_{\mathcal{V}} \phi(z) dz$$

such that for $L(f(\cdot, \theta); q_t, r_t)$ defined in (37) we have

$$\theta_K^* \equiv \operatorname{argmax}_{\theta} E [\tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)] \quad (44)$$

while the estimator solves $\hat{\theta}_K = \operatorname{argmax}_{\theta} \frac{1}{T} \sum_{t=1}^T \tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)$ from (38).

Further define $Q(\theta) = E [\tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)]$. Also let $\lambda_{\min}(B)$ denote the smallest eigenvalue of a matrix B . Then, we impose the following high-level condition that requires the objective function is locally concave at least around the pseudo true value.

Condition 1 We have (i) $\lambda_{\min} \left(-\frac{1}{2} \frac{\partial^2 Q(\theta)}{\partial \theta \partial \theta'} \right) \geq C_1 > 0$ for $\theta \in \Theta_T$ if $\|\theta - \theta_K^*\| = o(\zeta(K)^{-2})$ or (ii) $\lambda_{\min} \left(-\frac{1}{2} \frac{\partial^2 Q(\theta)}{\partial \theta \partial \theta'} \right) \geq C_2 \cdot \zeta(K)^{-2} \cdot \|\theta - \theta_K^*\|^{-1}$ otherwise.

This condition needs to be verified for each example of $Q(\theta)$, which could be difficult since $Q(\theta)$ is a highly nonlinear function of θ . Kim (2007) shows this condition is satisfied for the SNP density estimation with compact support.

Lemma E.1 Suppose Assumption E.4 and Condition 1 hold. Then for the θ_K in (40), we have $\|\theta_K - \theta_K^*\| = O(K^{-s/2})$ and

$$\sup_{v \in \mathcal{V}_\varepsilon} |f_0(z) - f_K^*(z)| = O\left(\zeta(K)^2 K^{-s/2}\right). \quad (45)$$

Proofs of lemmas and technical derivations are gathered in Section G. Lemma E.1 establishes the distance between the true density and the pseudo true density.

Next we derive the stochastic order of $\|\hat{\theta}_K - \theta_K^*\|$. Define the sample objective function $\hat{Q}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)$. Then we have

$$\sup_{\theta \in \Theta_T} \left| \hat{Q}_T(\theta) - Q(\theta) \right| = o_p\left(T^{-1/2+\alpha/2+\delta}\right) \quad (46)$$

for all sufficiently small $\delta > 0$ as shown in Lemma G.6 later in Section G. Now for $\|\theta - \theta_K^*\| \leq o(\eta_T)$, we also have

$$\sup_{\|\theta - \theta_K^*\| \leq o(\eta_T), \theta \in \Theta_T} \left| \hat{Q}_T(\theta) - \hat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| = o_p\left(\eta_T T^{-1/2+\alpha/2+\delta}\right) \quad (47)$$

as shown in Lemma G.7 later. From (46) and (47), it follows that

Lemma E.2 Suppose Assumption E.4 and Condition 1 hold and suppose $\frac{\zeta(K)^2 K}{T} \rightarrow 0$. Then for $K(T) = O(T^\alpha)$, we have $\|\hat{\theta}_K - \theta_K^*\| = o_p\left(T^{-1/2+\alpha/2+\delta}\right)$.

Combining Lemma E.1 and E.2, we obtain the convergence rate of the SNP density estimator to the pseudo true density function as

$$\begin{aligned} & \sup_{z \in \mathcal{V}_\varepsilon} \left| \hat{f}(z) - f_K^*(z) \right| = \sup_{z \in \mathcal{V}_\varepsilon} \left| \left(W^K(z)' (\hat{\theta}_K - \theta_K^*) \right) \left(W^K(z)' (\hat{\theta}_K + \theta_K^*) \right) \right| \\ & \leq C_1 \left(\sup_{z \in \mathcal{V}} \|W^K(z)\| \right)^2 \|\hat{\theta}_K - \theta_K^*\| = O\left(\zeta(K)^2\right) o_p\left(T^{-1/2+\alpha/2+\delta}\right) \end{aligned} \quad (48)$$

since $\|\theta\|^2 < 1$ for any $\theta \in \Theta_T$. Finally, by (45) and (48) we obtain the convergence rate of the SNP density to the true density as

$$\begin{aligned} \sup_{z \in \mathcal{V}_\varepsilon} \left| \widehat{f}(z) - f_0(z) \right| &\leq \sup_{z \in \mathcal{V}_\varepsilon} \left| \widehat{f}(z) - f_K^*(z) \right| + \sup_{z \in \mathcal{V}_\varepsilon} |f_K^*(z) - f_0(z)| \\ &= O(\zeta(K)^2) o_p\left(T^{-1/2+\alpha/2+\delta}\right) + O\left(\zeta(K)^2 K^{-s/2}\right). \end{aligned}$$

We summarize the result in Theorem E.1.

Theorem E.1 *Suppose Assumptions E.3-E.4 and Condition 1 hold and suppose $\zeta(K)^2 K/T \rightarrow 0$. Then, for $K = O(T^\alpha)$ with $\alpha < 1/3$, we have*

$$\sup_{z \in \mathcal{V}_\varepsilon} \left| \widehat{f}(z) - f_0(z) \right| = O(\zeta(K)^2) o_p\left(T^{-1/2+\alpha/2+\delta}\right) + O\left(\zeta(K)^2 K^{-s/2}\right) \quad (49)$$

for arbitrary small positive constant δ .

F Asymptotics for Test Statistics

F.1 Proof of Theorem 6.2

Recall that $I^g = I^g(f_{2|4}, f_{3|4})$, $\widetilde{I}_\gamma^g = \widehat{I}_\gamma^g(f_{2|4}, f_{3|4})$, and $\widehat{I}_\gamma^g = \widehat{I}_\gamma^g(\widehat{f}_{2|4}, \widehat{f}_{3|4})$. To prove Theorem 6.2, first we show the following lemma that establishes the conditions for $\widehat{I}_\gamma^g \xrightarrow{p} I^g$. We let $E_g[\cdot]$ denote an expectation operator that takes expectation with respect to a density g . In the main discussion, we have let $g(\cdot) = g^{(n-1:n)}(\cdot)$ for the test but $g(\cdot)$ could be a pdf of any other distribution for the validity of the test as long as the specified conditions in lemmas are satisfied.

Lemma F.1 *Suppose Assumption E.3 is satisfied. Further suppose (i) $f_{2|4}$ and $f_{3|4}$ are continuous, (ii) $E_g[\ln f_{2|4}] < \infty$ and $E_g[\ln f_{3|4}] < \infty$, (iii) $\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) = o(1)$. Suppose (iv)*

$$\frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{2|4}(\widetilde{v}_t) / f_{2|4}(\widetilde{v}_t) \right) \xrightarrow{p} 0 \text{ and } \frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{3|4}(\widetilde{v}_{t+1}) / f_{3|4}(\widetilde{v}_{t+1}) \right) \xrightarrow{p} 0.$$

Then we have $\widehat{I}_\gamma^g \xrightarrow{p} I^g$.

Proof. Note that

$$\begin{aligned} &\frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln f_{2|4}(\widetilde{v}_t) \\ &= \frac{1}{T_\gamma - 1} \sum_{t \text{ odd}} (1 + \gamma) \ln f_{2|4}(\widetilde{v}_t) + \frac{1}{T_\gamma - 1} \sum_{t \text{ even}} (1 - \gamma) \ln f_{2|4}(\widetilde{v}_t) + O_p\left(\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2)\right) \\ &= \frac{1}{2}(1 + \gamma) E_g[\ln f_{2|4}(\widetilde{v}_t)] + o_p(1) + \frac{1}{2}(1 - \gamma) E_g[\ln f_{2|4}(\widetilde{v}_t)] + o_p(1) + o_p(1) \\ &= E_g[\ln f_{2|4}(\widetilde{v}_t)] + o_p(1) \end{aligned}$$

by the law of large numbers under the condition (i) and by the condition (ii) and since $\{\widetilde{v}_t\}_{t=1}^T$ are iid. Similarly we have

$$\frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln f_{3|4}(\widetilde{v}_{t+1}) = E_g[\ln f_{3|4}(\widetilde{v}_{t+1})] + o_p(1)$$

and thus noting $\widehat{A} \xrightarrow[p]{} A$, by the Slutsky theorem

$$\widetilde{I}_\gamma^g \xrightarrow[p]{} I^g \quad (50)$$

since $I^g(f_{2|4}, f_{3|4})$ can be expressed as $I^g(f_{2|4}, f_{3|4}) = (A \cdot E_g[\ln f_{2|4}(\widetilde{v}_t) - \ln f_{3|4}(\widetilde{v}_t)])^2$. The condition (iv) implies $\widehat{I}_\gamma^g - \widetilde{I}_\gamma^g \xrightarrow[p]{} 0$ applying the Slutsky theorem. Combining this with (50), we conclude $\widehat{I}_\gamma^g \xrightarrow[p]{} I^g$. ■

Now we prove Theorem 6.2. We impose the following three conditions similar to Kim (2007).

Condition 2 $\sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{2|4}(\widetilde{v}_t) / f_{2|4}(\widetilde{v}_t) \right) = o_p(\sqrt{T})$ and $\sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{3|4}(\widetilde{v}_{t+1}) / f_{3|4}(\widetilde{v}_{t+1}) \right) = o_p(\sqrt{T})$.

Condition 3 $\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) = o(T^{-1/2})$.

Condition 4 $E_g [|\ln f_{2|4}(\widetilde{v}_t)|^2] < \infty$ and $E_g [|\ln f_{3|4}(\widetilde{v}_t)|^2] < \infty$.

Condition 2 only requires the SNP density estimator converge to the true one and the estimation error be negligible such that we can replace the estimated density with the true one in the asymptotic expansion of the test statistic. Condition 3 imposes that \mathcal{T}_2 the set of indices after we remove observations having very small densities grows large enough such that the trimming does not affect the asymptotic results. Condition 4 imposes bounded second moments of the log densities. We further discuss and verify these conditions in Section F.2 for our SNP density estimators of valuations.

Considering the powers of the proposed test, we may want to achieve the largest possible order of $d(T)$ while letting $d(T)\widehat{I}_\gamma^g$ preserve the some limiting distributions under the null. In what follows, we show that we can achieve this with $d(T) = O(T)$. Suppose Condition 2 holds, then it follows immediately that

$$\widehat{I}_\gamma^g - \widetilde{I}_\gamma^g = o_p(1/\sqrt{T})$$

for all $\gamma \geq 0$. This implies that the asymptotic distribution of $T\widehat{I}_\gamma^g$ will be identical to that of $T\widetilde{I}_\gamma^g$ under the null, which means the effect of nonparametric estimation is negligible. Now consider, under the null $f_{2|4} = f_{3|4}$

$$\begin{aligned} & \frac{1}{T_\gamma-1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) (\ln f_{2|4}(\widetilde{v}_t) - \ln f_{3|4}(\widetilde{v}_{t+1})) \\ &= \frac{2\gamma}{T_\gamma-1} \sum_{t \in \mathcal{Q}} (\ln f_{2|4}(\widetilde{v}_t) - \ln f_{2|4}(\widetilde{v}_{t+1})) + \frac{1+\gamma}{T_\gamma-1} \ln f_{2|4}(\widetilde{v}_1) - \frac{c_{\max \mathcal{Q}+1}(\gamma)}{T_\gamma-1} \ln f_{2|4}(\widetilde{v}_{\max \mathcal{Q}+1}) \\ & \quad + O_p\left(\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2)\right) \end{aligned} \quad (51)$$

where $\mathcal{Q} = \{t : 1 \leq t \leq T-1, t \text{ even}\}$. Now suppose Conditions 4 hold. Then, under the null we have

$$\begin{aligned} & \frac{1}{\sqrt{T/2}} \sum_{t \in \mathcal{Q}} (\ln f_{2|4}(\widetilde{v}_{t+1}) - \ln f_{2|4}(\widetilde{v}_t)) \xrightarrow[d]{} N(0, \Sigma) \text{ and} \\ & \frac{1}{\sqrt{T/2}} \sum_{t \in \mathcal{Q}} (\ln f_{3|4}(\widetilde{v}_{t+1}) - \ln f_{3|4}(\widetilde{v}_t)) \xrightarrow[d]{} N(0, \Sigma) \end{aligned} \quad (52)$$

where $\Sigma \equiv E_g \left[(\ln f_{2|4}(\tilde{v}_{t+1}) - \ln f_{2|4}(\tilde{v}_t))^2 \right]$ which is equal to $2\text{Var}_g[\ln f_{2|4}(\cdot)]$ by the Lindberg-Levy central limit theorem. Under the null $E_g \left[(\ln f_{3|4}(\tilde{v}_{t+1}) - \ln f_{3|4}(\tilde{v}_t))^2 \right]$ is identical to Σ . Therefore from (51) and (52), we conclude that under Conditions 2-4,

$$\begin{aligned} & \sqrt{T} \frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \left(\ln \hat{f}_{2|4}(\tilde{v}_t) - \ln \hat{f}_{3|4}(\tilde{v}_{t+1}) \right) \\ &= \sqrt{T} \frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \left(\ln f_{2|4}(\tilde{v}_t) - \ln f_{3|4}(\tilde{v}_{t+1}) \right) \xrightarrow{d} N(0, 2\gamma^2 \Sigma) \end{aligned}$$

for any $\gamma > 0$ noting $f_{2|4} = f_{3|4}$ under the null. Finally, for any $\hat{\Sigma} = \Sigma + o_p(1)$, we conclude that

$$\sqrt{T} \frac{1}{T_\gamma - 1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \left(\ln \hat{f}_{2|4}(\tilde{v}_t) - \ln \hat{f}_{3|4}(\tilde{v}_{t+1}) \right) / \sqrt{2\gamma \hat{\Sigma}^{\frac{1}{2}}} \xrightarrow{d} N(0, 1)$$

and hence $T\hat{I}_\gamma^{g2} \xrightarrow{d} \chi^2(1)$ with $A = 1/(\sqrt{2}\gamma\Sigma^{\frac{1}{2}})$ and $\hat{A} = 1/(\sqrt{2}\gamma\hat{\Sigma}^{\frac{1}{2}})$. Possible candidates of $\hat{\Sigma}$ are

$$\begin{aligned} \hat{\Sigma}_1 &= 2 \left(\frac{1}{T} \sum_{t=1}^T (\ln \hat{f}_{2|4}(\tilde{v}_t))^2 - \left\{ \frac{1}{T} \sum_{t=1}^T \ln \hat{f}_{2|4}(\tilde{v}_t) \right\}^2 \right) \text{ or} \\ \hat{\Sigma}_2 &= 2 \left(\frac{1}{T} \sum_{t=1}^T (\ln \hat{f}_{3|4}(\tilde{v}_t))^2 - \left\{ \frac{1}{T} \sum_{t=1}^T \ln \hat{f}_{3|4}(\tilde{v}_t) \right\}^2 \right) \end{aligned}$$

or its average $\hat{\Sigma}_3 = (\hat{\Sigma}_1 + \hat{\Sigma}_2)/2$. All of these are consistent under Condition 4 and under the conditions (iii) and (iv) of Lemma F.1.

F.2 Primitive Conditions for the SNP Estimators

In this section, we show that all the conditions in Lemma F.1 and Theorem 6.2 are satisfied for the SNP density estimators of (34). Here we should note that Lemma F.1 holds whether or not the null ($f_{2|4} = f_{3|4}$) is true while Theorem 6.2 is required to hold only under the null.

To simplify notation we are being abstract from the trimming device, which does not change our fundamental argument to obtain results below. We start with conditions for Lemma F.1. First, note the condition (i) is directly assumed and the condition (ii) in Lemma F.1 immediately hold since $f_{2|4}$ and $f_{3|4}$ are continuous and \mathcal{V} is compact. The condition (iii) of Lemma F.1 is verified as follows. For $\delta_1(T)$ and $\delta_2(T)$ that are positive numbers tending to zero as $T \rightarrow \infty$, consider

$$\begin{aligned} & \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) \\ & \leq \sum_{t=1}^{T-1} \Pr \left(\hat{f}_{2|4}(\tilde{v}_t) \leq \delta_1(T) \text{ or } \hat{f}_{3|4}(\tilde{v}_{t+1}) \leq \delta_2(T) \right) \\ & \leq \sum_{t=1}^{T-1} \Pr \left(|\hat{f}_{2|4}(\tilde{v}_t) - f_{2|4}(\tilde{v}_t)| + \delta_1(T) \geq f_{2|4}(\tilde{v}_t) \text{ or } |\hat{f}_{3|4}(\tilde{v}_{t+1}) - f_{3|4}(\tilde{v}_{t+1})| + \delta_2(T) \geq f_{3|4}(\tilde{v}_{t+1}) \right) \\ & \leq \sum_{t=1}^{T-1} \Pr \left(\begin{array}{l} \sup_{z \in \mathcal{V}} \left| \hat{f}_{2|4}(z) - f_{2|4}(z) \right| + \delta_1(T) \geq f_{2|4}(\tilde{v}_t) \\ \text{or } \sup_{z \in \mathcal{V}} \left| \hat{f}_{3|4}(z) - f_{3|4}(z) \right| + \delta_2(T) \geq f_{3|4}(\tilde{v}_{t+1}) \end{array} \right) \end{aligned} \tag{53}$$

and hence as long as $\sup_{z \in \mathcal{V}} \left| \hat{f}_{2|4}(z) - f_{2|4}(z) \right| = o_p(1)$, $\sup_{z \in \mathcal{V}} \left| \hat{f}_{3|4}(z) - f_{3|4}(z) \right| = o_p(1)$, $\delta_1(T) = o(1)$, and $\delta_2(T) = o(1)$, we have $\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) = o_p(1)$ since $f_{2|4}$ and $f_{3|4}$ are bounded away

from zero. Therefore, under $\alpha \leq 1/3 - 2\delta/3$ and $s > 2$, the condition (iii) of Lemma F.1 holds from Theorem E.1.

The condition (iv) of Lemma F.1 is easily established from the uniform convergence rate result. Using $|\ln(1+a)| \leq 2|a|$ in a neighborhood of $a = 0$, consider

$$\begin{aligned} & \left| \frac{1}{T\gamma-1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{2|4}(\tilde{v}_t) / f_{2|4}(\tilde{v}_t) \right) \right| \\ & \leq (1+\gamma) \sup_{z \in \mathcal{V}} \left| \ln \widehat{f}_{2|4}(z) - \ln f_{2|4}(z) \right| \leq (1+\gamma) \sup_{z \in \mathcal{V}} 2 \left| \frac{f_{2|4}(z) - \widehat{f}_{2|4}(z)}{\widehat{f}_{2|4}(z)} \right| \\ & = O(\zeta(K)^2) o_p(T^{-1/2+\alpha/2+\delta}) + O(\zeta(K)^2 K^{-s/2}) \end{aligned}$$

from Theorem E.1 and $\widehat{f}_1(\cdot)$ is bounded away from zero. Thus, $\frac{1}{T\gamma-1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{2|4}(\tilde{v}_t) / f_{2|4}(\tilde{v}_t) \right) = o_p(1)$ under $\alpha \leq 1/3 - 2\delta/3$ and $s > 2$ with $K = O(T^\alpha)$ noting $\zeta(K) = O(\sqrt{K})$. Similarly we can show $\frac{1}{T\gamma-1} \sum_{t \in \mathcal{T}_2} c_t(\gamma) \ln \left(\widehat{f}_{3|4}(\tilde{v}_{t+1}) / f_{3|4}(\tilde{v}_{t+1}) \right) = o_p(1)$ under $\alpha \leq 1/3 - 2\delta/3$ and $s > 2$.

Now we establish the conditions for Theorem 6.2. Again Condition 4 immediately holds since $f_{2|4}$ and $f_{3|4}$ are assumed to be continuous and \mathcal{V} is compact. Next, we show Condition 3. From (53) and the Markov inequality, we have

$$\begin{aligned} & \frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) \\ & \leq \sup_{z \in \mathcal{V}} \left(\frac{1}{f_{2|4}(z)^\tau} \right) \frac{1}{T-1} \sum_{t=1}^{T-1} E_g \left[\left| \widehat{f}_{2|4}(\tilde{v}_t) - f_{2|4}(\tilde{v}_t) + \delta_1(T) \right|^\tau \right] \\ & + \sup_{z \in \mathcal{V}} \left(\frac{1}{f_{3|4}(z)^\tau} \right) \frac{1}{T-1} \sum_{t=1}^{T-1} E_g \left[\left| \widehat{f}_{3|4}(\tilde{v}_{t+1}) - f_{3|4}(\tilde{v}_{t+1}) + \delta_2(T) \right|^\tau \right] \end{aligned}$$

and hence $\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) = o_p(1/\sqrt{T})$ as long as

$$\sup_{z \in \mathcal{V}} \left| \widehat{f}_{2|4}(z) - f_{2|4}(z) \right|^\tau = o_p(T^{-1/2}), \quad \sup_{z \in \mathcal{V}} \left| \widehat{f}_{3|4}(z) - f_{3|4}(z) \right|^\tau = o_p(T^{-1/2}), \quad (54)$$

$\delta_1(T)^\tau = o(1/\sqrt{T})$, and $\delta_2(T)^\tau = o(1/\sqrt{T})$ noting $f_{2|4}(\cdot)$ and $f_{3|4}(\cdot)$ are bounded away from zero. Note

$$\begin{aligned} \sup_{z \in \mathcal{V}} \left| \widehat{f}_{2|4}(z) - f_{2|4}(z) \right|^\tau &= o_p(T^{(-1/2+3\alpha/2+\delta)\tau}) + O(T^{(1-s/2)\alpha\tau}) \quad \text{and} \\ \sup_{z \in \mathcal{V}} \left| \widehat{f}_{3|4}(z) - f_{3|4}(z) \right|^\tau &= o_p(T^{(-1/2+3\alpha/2+\delta)\tau}) + O(T^{(1-s/2)\alpha\tau}) \end{aligned}$$

from Theorem E.1 and $\zeta(K) = O(\sqrt{K})$ letting $K = O(T^\alpha)$. In particular, we choose $\tau = 4$ and hence (54) holds under $\alpha \leq 1/4 - 2\delta/3$ and $s \geq 2 + 1/4\alpha$. $\delta_1(T)^\tau = o(1/\sqrt{T})$ and $\delta_2(T)^\tau = o(1/\sqrt{T})$ hold under $\delta_1(T) = o(T^{-\frac{1}{8}})$ and $\delta_2(T) = o(T^{-\frac{1}{8}})$. Therefore, under $\alpha \leq 1/4 - 2\delta/3$, $s \geq 2 + 1/4\alpha$, $\delta_1(T) = o(T^{-\frac{1}{8}})$, and $\delta_2(T) = o(T^{-\frac{1}{8}})$, we finally have $\frac{1}{T-1} \sum_{t=1}^{T-1} \Pr(t \notin \mathcal{T}_2) = o(1/\sqrt{T})$. Condition 2 remains to be verified. One can verify this condition for specific examples similarly with Kim (2007).

G Technical Lemmas and Proofs

In this section we prove Lemma E.1 and Lemma E.2 that we used to prove the convergence rate results in Theorem E.1. We start with showing a series of lemmas that are useful to prove Lemma E.1 and Lemma E.2. Let C, C_1, C_2, \dots denote generic constants.

G.1 Bound of the Truncated Hermite Series

Lemma G.1 *Suppose $W^K(z)$ is given by (33). Then, $\sup_{v \in \mathcal{V}} \|W^K(v)\| = \zeta(K) = O(\sqrt{K})$.*

Proof. Without loss of generality, we can impose \mathcal{V} to be symmetric around zero such that $\underline{v} + \bar{v} = 0$ by a normalization. Then, the claim follows from Kim (2007). ■

G.2 Uniform Law of Large Numbers

Here we establish a uniform convergence with rate as

$$\sup_{\theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - Q(\theta) \right| = o_p \left(T^{-1/2 + \alpha/2 + \delta} \right)$$

following Lemma 2 in Fenton and Gallant (1996) where $\widehat{Q}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)$ and $Q(\theta) = E[\tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)]$.

Lemma G.2 (Lemma 2 in Fenton and Gallant (1996)) *Let $\{\Theta_T\}$ be a sequence of compact subsets of a metric space (Θ, ρ) . Let $\{s_{Tt}(\theta) : \theta \in \Theta; t = 1, \dots, T; T = 1, \dots\}$ be a sequence of real valued random variables defined over a complete probability space (Ω, \mathcal{A}, P) . Suppose that there are sequences of positive numbers $\{d_T\}$ and $\{M_T\}$ such that for each θ° in Θ_T and for all θ in $\eta_T(\theta^\circ) = \{\theta \in \Theta_T : \rho(\theta, \theta^\circ) < d_T\}$, we have $|s_{Tt}(\theta) - s_{Tt}(\theta^\circ)| \leq \frac{1}{T} M_T \rho(\theta, \theta^\circ)$. Let $G_T(\tau)$ be the smallest number of open balls of radius τ necessary to cover Θ_T . If $\sup_{\theta \in \Theta_T} P \left\{ \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \epsilon \right\} \leq \Gamma_T(\epsilon)$, then for all sufficiently small $\epsilon > 0$ and all sufficiently large T ,*

$$P \left\{ \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \epsilon M_T d_T \right\} \leq G_T(\epsilon d_T/3) \Gamma_T(\epsilon M_T d_T/3).$$

First denote a set τ^c that contains (q_t, r_t) 's that survive the trimming device $\tau(\cdot)$. Now define $s_{Tt}(\theta) = \frac{1}{T} \tau(q_t, r_t) \ln L(f(\cdot, \theta); q_t, r_t)$. Then, we have $\widehat{Q}_T(\theta) = \sum_{t=1}^T s_{Tt}(\theta)$ and $Q(\theta) = \sum_{t=1}^T E[s_{Tt}(\theta)]$. To entertain Lemma G.2 in what follows, three Lemmas G.3-G.5 are verified.

Lemma G.3 *Suppose Assumption E.4 holds. Then, $|s_{Tt}(\theta) - s_{Tt}(\theta^\circ)| \leq C \frac{1}{T} \zeta(K(T))^2 \|\theta - \theta^\circ\|$.*

Proof. We write

$$\widetilde{L}(f(\cdot, \theta); q_t, r_t) = \tau(q_t, r_t) \frac{(F(q_t, \theta) - F(r_t, \theta))^{k_2 - k_1 - 1} (1 - F(q_t, \theta))^{k_1 - 1} f(q_t, \theta)}{(1 - F(r_t, \theta))^{k_2 - 1}}$$

where we denote $F(z, \theta) = \int_{\underline{v}}^z f(v, \theta) dv$.

Now note if $0 < c \leq a \leq b$, then $|\ln a - \ln b| \leq |a - b|/c$. Since $f(z, \theta)$ is bounded away from zero for all $\theta \in \Theta_T$ and $z \in \mathcal{V}$, $F(r_t, \theta)$ and $F(q_t, \theta)$ are bounded away from one (since $\max_t r_t < \max_t q_t \leq \bar{v} - \epsilon$ for $q_t, r_t \in \tau^c$), and $q_t > r_t$ for all t , we have $0 < C \leq \widetilde{L}(f(\cdot, \theta); q_t, r_t)$ for all $q_t, r_t \in \tau^c$. It follows that

$$|s_{Tt}(\theta) - s_{Tt}(\theta^\circ)| \leq \left| \widetilde{L}(f(\cdot, \theta); q_t, r_t) - \widetilde{L}(f(\cdot, \theta^\circ); q_t, r_t) \right| / TC.$$

Consider

$$\begin{aligned} & \left| \widetilde{L}(f(\cdot, \theta); q_t, r_t) - \widetilde{L}(f(\cdot, \theta^\circ); q_t, r_t) \right| \leq C_1 |f(q_t, \theta) - f(q_t, \theta^\circ)| \\ & = C_1 |W^K(q_t)' \theta + W^K(q_t)' \theta^\circ| |W^K(q_t)' \theta - W^K(q_t)' \theta^\circ| \\ & \leq C_1 \sup_{z \in \mathcal{V}} \|W^K(z)\| (\|\theta\| + \|\theta^\circ\|) \sup_{z \in \mathcal{V}} \|W^K(z)\| \|\theta - \theta^\circ\| \leq C_2 \zeta(K)^2 \|\theta - \theta^\circ\| \end{aligned}$$

where the first inequality is obtained since $F(r_t, \theta)$ is bounded above from one, $0 < F(q_t, \theta) - F(r_t, \theta) < 1$, and $0 < F(q_t, \theta) < 1$ for all $(q_t, r_t) \in \tau^c$. The last inequality is obtained from $\|\theta\| < 1$ for all $\theta \in \Theta_T$ and $\sup_{z \in \mathcal{V}} \|W^K(z)\| = \zeta(K)$. It follows that

$$|s_{Tt}(\theta) - s_{Tt}(\theta^o)| \leq C_2 \zeta(K)^2 \|\theta - \theta^o\| / T.$$

■

Lemma G.4 *Suppose Assumption E.4 holds and $\zeta(K) = O(K^\zeta)$. Then,*

$$\Pr \left\{ \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \epsilon \right\} \leq 2 \exp \left(-2\epsilon^2 / T \left(\frac{1}{T} 2\zeta \ln K(T) + \frac{1}{T} C \right)^2 \right).$$

Proof. We have $0 < C_1 \leq f(z, \theta) < C_2 K^{2\zeta} + \epsilon_0 \frac{\phi(0)}{\int_{\mathcal{V}} \phi(z) dz}$ by construction and since $f(z, \theta)$ is bounded away from zero. Moreover $0 < F(q_t, \theta) < 1$, $0 < F(r_t, \theta) < 1$, and $0 < F(q_t, \theta) - F(r_t, \theta) < 1$ for all $(q_t, r_t) \in \tau^c$. Thus it follows that $\frac{1}{T} C_3 < s_{Tt}(\theta) < \frac{1}{T} 2\zeta \ln K + \frac{1}{T} C_4$ for sufficiently large K and for all $(q_t, r_t) \in \tau^c$. Hoeffding's (1963) inequality implies that $\Pr(|Y_1 + \dots + Y_T| \geq \epsilon) \leq 2 \exp \left(-2\epsilon^2 / \sum_{t=1}^T (b_t - a_t)^2 \right)$ for independent random variables centered zero with ranges $a_t \leq Y_t \leq b_t$. Applying this inequality, we obtain the result. ■

Lemma G.5 (Lemma 6 in Fenton and Gallant (1996)) *The number of open balls of radius δ required to cover Θ_T is bounded by $2K(T)(2/\delta + 1)^{K(T)-1}$.*

Proof. Lemma 1 of Gallant and Souza (1991) shows that the number of radius- δ balls needed to cover the surface of a unit sphere in \mathbb{R}^p is bounded by $2p(2/\delta + 1)^{p-1}$. Noting $\dim(\Theta_T) = K(T)$, the result follows immediately. ■

Applying the results of Lemma G.3-G.5, finally we obtain

Lemma G.6 *Let $K(T) = C \cdot T^\alpha$ with $0 < \alpha < 1$ and suppose Assumption E.4 holds. Then,*

$$\sup_{\theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - Q(\theta) \right| = o_p(T^{-1/2+\alpha/2+\delta}).$$

Proof. Let $M_T = C_1 O(K^{2\zeta}) = C_2 T^{2\zeta\alpha}$, $d_n = \frac{1}{C_1} T^{-(2\zeta-1)\alpha-\beta}$, and $\rho(\theta, \theta^o) = \|\theta - \theta^o\|$. Then from Lemma G.2, we have

$$\begin{aligned} & \Pr \left\{ \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \epsilon T^{\alpha-\beta} \right\} \\ & \leq 4C \cdot T^\alpha \left(\frac{6C_1}{\epsilon} T^{(2\zeta-1)\alpha+\beta} + 1 \right)^{T^\alpha-1} \exp \left(-2 \left(\frac{\epsilon T^{\alpha-\beta}}{3} \right)^2 / T \left(\frac{1}{T} 2\zeta \ln K(T) + \frac{1}{T} C_2 \right)^2 \right) \end{aligned}$$

applying Lemma G.3, Lemma G.4, and Lemma G.5.

Note $4C \cdot T^\alpha \left(\frac{6C_1}{\epsilon} T^{(2\zeta-1)\alpha+\beta} + 1 \right)^{T^\alpha-1}$ is dominated by $C_3 T^\alpha T^{((2\zeta-1)\alpha+\beta)(T^\alpha-1)}$ for sufficiently large T and note $T \left(\frac{1}{T} 2\zeta \ln K(T) + \frac{1}{T} C_2 \right)^2$ is dominated by $T \left(\frac{1}{T} 2\zeta \ln K(T) \right)^2$. Thus, we simplify

$$\begin{aligned} & \Pr \left\{ \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \epsilon T^{\alpha-\beta} \right\} \\ & \leq C_4 \exp \left(\ln \left(T^\alpha T^{((2\zeta-1)\alpha+\beta)(T^\alpha-1)} \right) - \frac{2\epsilon^2}{9} T^{2\alpha-2\beta+1} / (2\zeta \ln K(T))^2 \right) \\ & = C_4 \exp \left(\alpha \ln T + ((2\zeta-1)\alpha + \beta) (T^\alpha - 1) \ln T - \frac{2\epsilon^2}{9} T^{2\alpha-2\beta+1} / (2\zeta \ln K(T))^2 \right) \end{aligned}$$

for sufficiently large T . As long as $2\alpha - 2\beta + 1 > \alpha$, $\frac{2\varepsilon^2}{9}T^{2\alpha-2\beta+1}/(2\zeta \ln K(T))^2$ dominates $\alpha \ln T + ((2\zeta - 1)\alpha + \beta)(T^\alpha - 1) \ln T$ and hence we conclude

$$\Pr \left\{ \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| > \varepsilon T^{\alpha-\beta} \right\} = o(1)$$

provided that $\frac{\alpha+1}{2} > \beta > \alpha$. By taking $\beta = \frac{1}{2} + \frac{1}{2}\alpha - \delta$ (the best possible rate), we have

$$\sup_{\theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - Q(\theta) \right| = \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T (s_{Tt}(\theta) - E[s_{Tt}(\theta)]) \right| = o_p(T^{-\frac{1}{2} + \frac{1}{2}\alpha + \delta})$$

for all sufficiently small $\delta > 0$. ■

Lemma G.7 Suppose (i) Assumption E.4 holds and (ii) $\frac{\zeta(K)^2 K}{T} \rightarrow 0$. Let $\eta_T = T^{-\beta_\eta}$ with $\alpha < \beta_\eta \leq 1/2 - \alpha/2 - \delta$. Then

$$\sup_{\|\theta - \theta_K^*\| \leq o(\eta_T), \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| = o_p(\eta_T T^{-1/2 + \alpha/2 + \delta}).$$

Proof. In the following proof we treat $\epsilon_0 = 0$ to simplify discussions since we can pick ϵ_0 arbitrary small but still we need to impose that $f(\cdot, \theta) \in \mathcal{F}_T$ is bounded away from zero.

Now applying the mean value theorem for $\bar{\theta}$ that lies between θ and θ_K^* , we can write

$$\begin{aligned} & \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) = \widehat{Q}_T(\theta) - Q(\theta) - \left(\widehat{Q}_T(\theta_K^*) - Q(\theta_K^*) \right) \\ & = \left(\partial(\widehat{Q}_T(\bar{\theta}) - Q(\bar{\theta})) / \partial \theta' \right) (\theta - \theta_K^*). \end{aligned} \quad (55)$$

Now consider for any $\tilde{\theta}$ such that $\|\tilde{\theta} - \theta_K^*\| \leq o(\eta_T)$,

$$\begin{aligned} & \partial \widehat{Q}_T(\tilde{\theta}) / \partial \theta - \partial Q(\tilde{\theta}) / \partial \theta \\ & = -(k_1 - 1) \left(\frac{1}{T} \sum_{t=1}^T \tau(q_t, r_t) \frac{f(q_t, \tilde{\theta})}{1 - F(q_t, \tilde{\theta})} \frac{\partial f(q_t, \tilde{\theta})}{\partial \theta} - E \left[\tau(q_t, r_t) \frac{f(q_t, \tilde{\theta})}{1 - F(q_t, \tilde{\theta})} \frac{\partial f(q_t, \tilde{\theta})}{\partial \theta} \right] \right) \end{aligned} \quad (56)$$

$$+ \left(\frac{1}{T} \sum_{t=1}^T \tau(q_t, r_t) \frac{1}{f(q_t, \tilde{\theta})} \frac{\partial f(q_t, \tilde{\theta})}{\partial \theta} - E \left[\tau(q_t, r_t) \frac{1}{f(q_t, \tilde{\theta})} \frac{\partial f(q_t, \tilde{\theta})}{\partial \theta} \right] \right) \quad (57)$$

$$+ (k_2 - 1) \left(\frac{1}{T} \sum_{t=1}^T \frac{\tau(q_t, r_t) f(r_t, \tilde{\theta})}{1 - F(r_t, \tilde{\theta})} \frac{\partial f(r_t, \tilde{\theta})}{\partial \theta} - E \left[\tau(q_t, r_t) \frac{f(r_t, \tilde{\theta})}{1 - F(r_t, \tilde{\theta})} \frac{\partial f(r_t, \tilde{\theta})}{\partial \theta} \right] \right). \quad (58)$$

We first bound (56). Note $\frac{\partial f(q_t, \theta)}{\partial \theta} = 2W^K(q_t)W^K(q_t)'\theta$ and define

$$M_{Tt} = \tau(q_t, r_t) \frac{f(q_t, \tilde{\theta})}{1 - F(q_t, \tilde{\theta})} W^K(q_t)W^K(q_t)'\tilde{\theta} - E \left[\tau(q_t, r_t) \frac{f(q_t, \tilde{\theta})}{1 - F(q_t, \tilde{\theta})} W^K(q_t)W^K(q_t)'\tilde{\theta} \right].$$

Considering that M_{Tt} is a triangular array of i.i.d random variables with mean zero, we bound (56) as follows. First consider

$$\text{Var}[M_{Tt}] = E[M_{Tt}M_{Tt}'] \leq E \left[\tau(q_t, r_t) \left(\frac{f(q_t, \tilde{\theta})}{1 - F(q_t, \tilde{\theta})} \right)^2 \left(W^K(q_t)'\tilde{\theta} \right)^2 W^K(q_t)W^K(q_t)' \right]. \quad (59)$$

The right hand side of (59) is bounded by

$$\begin{aligned}
& E \left[\tau(q_t, r_t) \frac{f(q_t, \tilde{\theta})^4}{(1-F(q_t, \tilde{\theta}))^2} W^K(q_t) W^K(q_t)' \right] \\
& \leq \sup_{z \in \mathcal{V}_\varepsilon} \left(\frac{f(z, \tilde{\theta})^4}{(1-F(z, \tilde{\theta}))^2} \right) \sup_{z \in \mathcal{V}} g^{(n-k_1-1:n)}(z) \int_{\mathcal{V}} W^K(z) W^K(z)' dz \\
& \leq C_1 \cdot \left(\sup_{z \in \mathcal{V}_\varepsilon} \left| f(z, \tilde{\theta}) - f_0(z) \right|^4 + \sup_{z \in \mathcal{V}} f_0(z)^4 \right) I_K
\end{aligned}$$

since $1 - F(z, \tilde{\theta})$ is bounded away from zero uniformly over $z \in \mathcal{V}_\varepsilon$ and since $g^{(n-k_1-1:n)}(z)$ is bounded away from above uniformly over $z \in \mathcal{V}_\varepsilon$. Finally note $\sup_{z \in \mathcal{V}_\varepsilon} \left| f(z, \tilde{\theta}) - f_0(z) \right| \leq \sup_{z \in \mathcal{V}_\varepsilon} \left| f(z, \tilde{\theta}) - f(z, \theta_K^*) \right| + \sup_{z \in \mathcal{V}_\varepsilon} \left| f(z, \theta_K^*) - f_0(z) \right| = O(\zeta(K)^2) (o(\eta_T) + O(K^{-s/2}))$ by (45) and $\left\| \tilde{\theta} - \theta_K^* \right\| \leq o(\eta_T)$ and hence we have

$$\text{Var}[M_{Tt}] \leq C_2 \zeta(K)^8 (o(\eta_T^4) + O(K^{-2s})) I_K + C_3 I_K \leq C_4 I_K$$

under $\zeta(K)^8 (o(\eta_T^4) + O(K^{-2s})) = o(1)$ which holds as long as $s > 2$ and $-4\beta_\eta + 4\alpha < 0$. Now note

$$E \left(\left\| \sum_{t=1}^T M_{Tt} / \sqrt{T} \right\| \right) \leq \sqrt{\text{tr}(\text{Var}[L_{Tt}])} \leq \sqrt{C_4 \text{tr}(I_K)} = O(\sqrt{K})$$

and hence (56) is $O_p(\sqrt{K/T})$ from the Markov inequality. Similarly we can show that (58) is also $O_p(\sqrt{K/T})$.

Now we bound (57). Define $L_{Tt} = \left(\tau(q_t, r_t) \frac{W^K(q_t) W^K(q_t)' \tilde{\theta}}{f(q_t, \tilde{\theta})} - E \left[\tau(q_t, r_t) \frac{W^K(q_t) W^K(q_t)' \tilde{\theta}}{f(q_t, \tilde{\theta})} \right] \right)$.

Then, we can rewrite (57) as $2 \frac{1}{T} \sum_{t=1}^T L_{Tt}$. Noting again L_{Tt} is a triangular array of i.i.d random variables with mean zero, we bound (57) as follows. Consider

$$\begin{aligned}
\text{Var}[L_{Tt}] &= E[L_{Tt} L_{Tt}'] \leq E \left[\tau(q_t, r_t) \left(\frac{W^K(q_t) \tilde{\theta}}{f(q_t, \tilde{\theta})} \right)^2 W^K(q_t) W^K(q_t)' \right] \\
&= E \left[\tau(q_t, r_t) \frac{1}{f(q_t, \tilde{\theta})} W^K(q_t) W^K(q_t)' \right] \leq \sup_{z \in \mathcal{V}_\varepsilon} \frac{g^{(n-k_1-1:n)}(z)}{f(z, \tilde{\theta})} \int_{\mathcal{V}} W^K(z) W^K(z)' dz \leq C_1 I_K
\end{aligned}$$

since $g^{(n-k_1-1:n)}(z)$ is bounded from above and $f(z, \tilde{\theta})$ is bounded from below uniformly over $z \in \mathcal{V}_\varepsilon$. It follows that

$$E \left(\left\| \sum_{t=1}^T L_{Tt} / \sqrt{T} \right\| \right) \leq \sqrt{\text{tr}(\text{Var}[L_{Tt}])} \leq \sqrt{C_1 \text{tr}(I_K)} = O(\sqrt{K}).$$

Thus, we bound (57) as $O_p(\sqrt{K/T})$ from the Markov inequality. We conclude

$$\left\| \partial \hat{Q}_T(\tilde{\theta}) / \partial \theta - \partial Q(\tilde{\theta}) / \partial \theta \right\| = O_p(\sqrt{K/T})$$

under $s > 2$ and $-4\beta_\eta + 4\alpha < 0$. Thus, noting $O_p\left(\sqrt{K/T}\right) = o_p\left(T^{-1/2+\alpha/2+\delta}\right)$ for $K = T^\alpha$ and sufficiently small δ , we have

$$\begin{aligned} & \sup_{\|\theta - \theta_K^*\| \leq o(\eta_T), \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| \\ & \leq \sup_{\|\theta - \theta_K^*\| \leq o(\eta_T), \theta \in \Theta_T} \left\| \partial(\widehat{Q}_T(\theta) - Q(\theta)) / \partial\theta \right\| \sup_{\|\theta - \theta_K^*\| \leq o(\eta_T), \theta \in \Theta_T} \|\theta - \theta_K^*\| \\ & = o_p\left(\eta_T T^{-1/2+\alpha/2+\delta}\right) \end{aligned}$$

from (55) applying the Cauchy-Schwarz inequality. ■

G.3 Proof of Lemma E.1

Proof. In the following proof, we treat $\epsilon_0 = 0$ to simplify discussions since we can pick ϵ_0 arbitrary small but still we need to impose that $f(\cdot, \theta) \in \underline{\mathcal{F}}_T$ is bounded away from zero.

Now define $Q_0 = E[\tau(q_t, r_t) \ln L(f_0; q_t, r_t)]$ and recall that $\theta_K^* = \underset{\theta \in \Theta_T}{\operatorname{argmax}} Q(\theta)$. Consider

$$\underset{\theta \in \Theta_T}{\operatorname{argmin}} Q_0 - Q(\theta) = \underset{\theta \in \Theta_T}{\operatorname{argmax}} Q(\theta)$$

which implies that among the parametric family $\{f(z, \theta) : \theta = (\vartheta_1, \dots, \vartheta_K)\}$, $Q(\theta_K^*)$ will have the minimum distance to Q_0 noting for all $\theta \in \Theta_T$, $Q(\theta) \leq Q_0$ from the information inequality (see Gallant and Nychka (1987, p.484)). First, we show that

$$Q_0 - Q(\theta_K) = O(K^{-s}). \quad (60)$$

Define $F(z, \theta_K) = \int_v^z f(v, \theta_K) dv$ and consider

$$\begin{aligned} |Q_0 - Q(\theta_K)| & \leq E \left[\tau(q_t, r_t) \left| \ln \frac{L(f_0; q_t, r_t)}{L(f(\cdot, \theta); q_t, r_t)} \right| \right] \\ & \leq E \left[\tau(q_t, r_t) \left| \ln \frac{f_0(q_t)}{f(q_t, \theta_K)} \right| \right] + E \left[(k_1 - 1) \tau(q_t, r_t) \left| \ln \frac{1 - F_0(q_t)}{1 - F(q_t, \theta_K)} \right| \right] \\ & \quad + E \left[(k_2 - 1) \tau(q_t, r_t) \left| \ln \frac{1 - F_0(r_t)}{1 - F(r_t, \theta_K)} \right| \right] \end{aligned} \quad (61)$$

by the triangular inequality. We bound the three terms in (61) in turns.

(i) *Bound of the first term in (61)*

Now denoting a random variable Z to follow the distribution with the density f_0 with the support \mathcal{V} and using $|\ln(1 + a)| \leq 2|a|$ in a neighborhood of $a = 0$, consider

$$\begin{aligned} & E \left[\tau(q_t, r_t) \left| \ln \frac{f_0(q_t)}{f(q_t, \theta_K)} \right| \right] \leq E \left[2 \left| \frac{f_0(q_t)}{f(q_t, \theta_K)} - 1 \right| \right] \\ & = 2 \int_{\mathcal{V}} \frac{1}{f(z, \theta_K)} |f_0(z) - f(z, \theta_K)| g^{(n-k_1+1:n)}(z) dz \\ & = 2 \int_{\mathcal{V}} \frac{g^{(n-k_1+1:n)}(z)}{f(z, \theta_K)} \frac{\sqrt{f_0(z)}}{\sqrt{f_0(z)}} \left| h_{f_0}(z) e^{-z^2/4} + W^K(z)' \theta_K \right| \left| h_{f_0}(z) e^{-z^2/4} - W^K(z)' \theta_K \right| dz \\ & \leq 2 \sup_{z \in \mathcal{V}} \frac{g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)}}{f(z, \theta_K)} \sup_{z \in \mathcal{V}} \left| h_{f_0}(z) e^{-z^2/4} - W^K(z)' \theta_K \right| E \left[\frac{|h_{f_0}(Z) e^{-Z^2/4}| + |W^K(Z)' \theta_K|}{\sqrt{f_0(Z)}} \right]. \end{aligned}$$

Note

$$E \left[\left| h_{f_0}(Z) e^{-Z^2/4} \right| / \sqrt{f_0(Z)} \right] \leq \sqrt{E \left[h_{f_0}^2(Z) e^{-Z^2/2} / f_0(Z) \right]} < 1 \quad (62)$$

since $0 < h_{f_0}^2(z)/f_0(z) < 1$ for all $z \in \mathcal{V}$ by construction and note

$$E \left[\left| W^K(Z)' \theta_K \right| / \sqrt{f_0(Z)} \right] \leq \sqrt{E \left[\theta_K' W^K(Z) W^K(Z)' \theta_K / f_0(Z) \right]} = \sqrt{\theta_K' \theta_K} = \|\theta_K\| < 1 \quad (63)$$

since $\|\theta_K\|^2 < 1$. Also note $\sup_{z \in \mathcal{V}} \frac{g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)}}{f(z, \theta_K)} < \infty$ since $g^{(n-k_1+1:n)}(\cdot)$ and $f_0(\cdot)$ are bounded from above and since $f(z, \theta_K)$ is bounded away from zero. From (41), it follows that

$$E \left[\left| \ln(f_0(q_t)/f(q_t, \theta_K)) \right| \right] = O(K^{-s}) \quad \text{and similarly} \quad E \left[\left| \ln(f_0(r_t)/f(r_t, \theta_K)) \right| \right] = O(K^{-s}).$$

(ii) *Bound of the second term in (61)*

For some $0 < \alpha < 1$, note

$$\begin{aligned} E \left[\left| F_0(q_t) - F(q_t, \theta_K) \right| \right] &= E \left[\left| \int_{\underline{v}}^{q_t} f_0(z) - f(z, \theta_K) dz \right| \right] \\ &\leq E \left[\left| (q_t - \underline{v}) |f_0(\alpha q_t + (1-\alpha)\underline{v}) - f(\alpha q_t + (1-\alpha)\underline{v}, \theta_K)| \right| \right] \\ &\leq (\bar{v} - \underline{v}) E \left[\left| f_0(\alpha q_t + (1-\alpha)\underline{v}) - f(\alpha q_t + (1-\alpha)\underline{v}, \theta_K) \right| \right], \end{aligned}$$

where the last inequality is from $\bar{v} > q_t > \underline{v}$. Using the change of variable $z = \alpha q_t + (1-\alpha)\underline{v}$, note

$$\begin{aligned} &E \left[\left| f_0(\alpha q_t + (1-\alpha)\underline{v}) - f(\alpha q_t + (1-\alpha)\underline{v}, \theta_K) \right| \right] \\ &\leq \sup_{z \in \mathcal{V}} \left| h_{f_0}(z) e^{-z^2/4} - W^K(z)' \theta_K \right| \int_{\underline{v}}^{\bar{v} - (1-\alpha)(\bar{v} - \underline{v})} \frac{1}{\alpha} g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)} \frac{|h_{f_0}(z) e^{-z^2/4} + W^K(z)' \theta_K|}{\sqrt{f_0(z)}} dz \\ &\leq \sup_{z \in \mathcal{V}} \left| h_{f_0}(z) e^{-z^2/4} - W^K(z)' \theta_K \right| \int_{\underline{v}}^{\bar{v}} \frac{1}{\alpha} g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)} \frac{|h_{f_0}(z) e^{-z^2/4} + W^K(z)' \theta_K|}{\sqrt{f_0(z)}} dz \end{aligned}$$

where the second inequality is from $(1-\alpha)(\bar{v} - \underline{v}) > 0$. Note

$$\begin{aligned} &\int_{\underline{v}}^{\bar{v}} \frac{1}{\alpha} g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)} \frac{|h_{f_0}(z) e^{-z^2/4} + W^K(z)' \theta_K|}{\sqrt{f_0(z)}} dz \\ &< \sup_{z \in \mathcal{V}} \frac{1}{\alpha} g^{(n-k_1+1:n)}(z) \sqrt{f_0(z)} E_{f_0} \left[\frac{|h_{f_0}(Z) e^{-Z^2/4}| + |W^K(Z)' \theta_K|}{\sqrt{f_0(Z)}} \right] < \infty \end{aligned}$$

by (62) and (63) and hence

$$E \left[\left| F_0(q_t) - F(q_t, \theta_K) \right| \right] = O(K^{-s}). \quad (64)$$

Now using $|\ln(1+a)| \leq 2|a|$ in a neighborhood of $a = 0$, note

$$\begin{aligned} E \left[\left| \tau(q_t, r_t) \left| \ln \frac{1-F_0(q_t)}{1-F(q_t, \theta_K)} \right| \right| \right] &\leq 2E \left[\left| \tau(q_t, r_t) \left| \frac{F(q_t, \theta_K) - F_0(q_t)}{1-F(q_t, \theta_K)} \right| \right| \right] \\ &\leq C \cdot E \left[\left| \tau(q_t, r_t) |F_0(q_t) - F(q_t, \theta_K)| \right| \right] \end{aligned}$$

since $F(q_t, \theta_K) < 1$ and hence from (64), we bound the second term of (61) to be $O(K^{-s})$.

(iii) *Bound of the third term in (61)*

Similarly with the second term of (61), we can show that the third term to be $O(K^{-s})$.

From these results (i)-(iii), we conclude $|Q_0 - Q(\theta_K)| = O(K^{-s})$. It follows that

$$0 \leq Q(\theta_K^*) - Q(\theta_K) \leq Q(\theta_K^*) - Q_0 + C_1 K^{-s} \leq C_1 K^{-s} \quad (65)$$

where the first inequality is by definition of $\theta_K^* = \operatorname{argmax}_{\theta \in \Theta_T} Q(\theta)$, the second inequality is by (60), and the last inequality is since $Q(\theta_K^*) \leq Q_0$ from the information inequality (see Gallant and Nychka (1987, p.484)). Now using the second order Taylor expansion where $\tilde{\theta}$ lies between a $\theta \in \Theta_T$ and θ_K^* , we have

$$\begin{aligned} & Q(\theta_K^*) - Q(\theta) \\ &= -\frac{\partial Q}{\partial \theta'}(\theta_K^*)(\theta - \theta_K^*) - \frac{1}{2}(\theta - \theta_K^*)' \frac{\partial^2 Q(\tilde{\theta})}{\partial \theta \partial \theta'} (\theta - \theta_K^*) = -\frac{1}{2}(\theta - \theta_K^*)' \frac{\partial^2 Q(\tilde{\theta})}{\partial \theta \partial \theta'} (\theta - \theta_K^*) \end{aligned} \quad (66)$$

since $\frac{\partial Q}{\partial \theta}(\theta_K^*) = 0$ by F.O.C of (44). Picking θ to be θ_K , from (66) and Condition 1, we conclude

$$\begin{aligned} Q(\theta_K^*) - Q(\theta_K) &\geq C_4 \|\theta_K - \theta_K^*\|^2 \text{ if } \|\theta_K^* - \theta_K\| = o(\zeta(K)^{-2}) \\ Q(\theta_K^*) - Q(\theta_K) &\geq O(\zeta(K)^{-2}) \|\theta_K - \theta_K^*\| \geq O(\zeta(K)^{-4}) \text{ otherwise} \end{aligned} \quad (67)$$

since $Q(\theta_K^*) - Q(\theta_K) \geq \lambda_{\min}\left(-\frac{1}{2} \frac{\partial^2 Q(\tilde{\theta})}{\partial \theta \partial \theta'}\right) \|\theta_K - \theta_K^*\|^2$. However, the case of (67) contradicts to (65) if $s > 2$, which means (65) implies $\|\theta_K^* - \theta_K\| = o(\zeta(K)^{-2})$ under $s > 2$ and hence

$$Q(\theta_K^*) - Q(\theta_K) \geq C_4 \|\theta_K - \theta_K^*\|^2.$$

Together with (65), it implies $C_1 K^{-s} \geq Q(\theta_K^*) - Q(\theta_K) \geq C_4 \|\theta_K - \theta_K^*\|^2$ and hence under $s > 2$

$$\|\theta_K - \theta_K^*\| = O(K^{-s/2}) \quad (68)$$

as claimed in Lemma E.1. Finally note $\|\theta_K - \theta_K^*\| = o(\zeta(K)^{-2})$ as long as (68) holds under $s > 2$.

Now consider

$$\begin{aligned} & \sup_{z \in \mathcal{V}_\varepsilon} |f(z, \theta_K) - f(z, \theta_K^*)| \leq \sup_{z \in \mathcal{V}_\varepsilon} \left\| (W^K(z)' \theta_K)^2 - (W^K(z)' \theta_K^*)^2 \right\| \\ & \leq \sup_{z \in \mathcal{V}_\varepsilon} \left\| W^K(z)' \theta_K - W^K(z)' \theta_K^* \right\| \sup_{z \in \mathcal{V}_\varepsilon} \left\| W^K(z)' \theta_K + W^K(z)' \theta_K^* \right\| \\ & \leq \sup_{z \in \mathcal{V}} \|W^K(z)\| \|\theta_K - \theta_K^*\| \sup_{z \in \mathcal{V}} \|W^K(z)\| (\|\theta_K\| + \|\theta_K^*\|) = O(\zeta(K)^2 K^{-s/2}) \end{aligned}$$

from the Cauchy-Schwarz inequality, (68), $\sup_{z \in \mathcal{V}} \|W^K(z)\| \leq \zeta(K)$, and $\|\theta\|^2 < 1$ for any $\theta \in \Theta_T$. It follows that

$$\begin{aligned} & \sup_{z \in \mathcal{V}_\varepsilon} |f_0(z) - f_K^*(z)| \leq \sup_{z \in \mathcal{V}_\varepsilon} |f_0(z) - f(z, \theta_K)| + \sup_{z \in \mathcal{V}_\varepsilon} |f(z, \theta_K) - f(z, \theta_K^*)| \\ & \leq O(\zeta(K) K^{-s}) + O(\zeta(K)^2 K^{-s/2}) = O(\zeta(K)^2 K^{-s/2}). \end{aligned}$$

■

G.4 Proof of Lemma E.2

Proof. Similarly with Kim (2007), we can show $\left\| \hat{\theta}_K - \theta_K^* \right\| = o_p(T^{-\kappa}) = o_p(T^{-1/2+\alpha/2+\delta})$. The idea - a similar proof strategy used in Ai and Chen (2003) for a sieve minimum distance estimation - is that an improvement in the convergence rate contributed iteratively by the local curvature of $\hat{Q}_T(\theta)$ around θ_K^* can be achieved up to the uniform convergence rate of $\hat{Q}_T(\theta)$ to $Q(\theta)$ and hence

we can obtain the convergence rate of the SNP density estimator up to $o_p(T^{-1/2+\alpha/2+\delta})$, which is the uniform convergence rate we derived in Section G.2 (Lemma G.6). A formal proof follows below.

From Condition 1 and (66), similarly with (67), we note that

$$\begin{aligned} Q(\theta_K^*) - Q(\theta) &\geq C_1 \cdot \|\theta - \theta_K^*\|^2 \text{ if } \|\theta - \theta_K^*\| = o(\zeta(K)^{-2}) \\ Q(\theta_K^*) - Q(\theta) &\geq C_2 \cdot \zeta(K)^{-2} \|\theta - \theta_K^*\| \text{ otherwise.} \end{aligned} \quad (69)$$

Denote $\kappa = 1/2 - \alpha/2 - \delta$ and $\eta_{0T} = o(T^{-\kappa})$. We derive the convergence rate in two cases: one is when η_{0T} has the equal or a smaller order than $o(\zeta(K)^{-4})$ and the other case is when η_{0T} has a larger order than $o(\zeta(K)^{-4})$.

1) When η_{0T} has equal or smaller order than $o(\zeta(K)^{-4})$, which holds under $\alpha < 1/5$:

Now let $\delta_{0T} = \sqrt{2\eta_{0T}}$. For any c such that $C_1 c^2 > 1$, it follows

$$\begin{aligned} &\Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\delta_{0T}\right) \\ &\leq \Pr\left(\sup_{\|\theta - \theta_K^*\| \geq c\delta_{0T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) \geq \widehat{Q}_T(\theta_K^*)\right) \\ &\leq \Pr\left(\sup_{\theta \in \Theta_T} \left|\widehat{Q}_T(\theta) - Q(\theta)\right| > \eta_{0T}\right) \\ &\quad + \Pr\left(\left\{\sup_{\theta \in \Theta_T} \left|\widehat{Q}_T(\theta) - Q(\theta)\right| \leq \eta_{0T}\right\} \cap \left\{\sup_{\|\theta - \theta_K^*\| \geq c\delta_{0T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) \geq \widehat{Q}_T(\theta_K^*)\right\}\right) \\ &\leq \Pr\left(\sup_{\theta \in \Theta_T} \left|\widehat{Q}_T(\theta) - Q(\theta)\right| > \eta_{0T}\right) + \Pr\left(\sup_{\|\theta - \theta_K^*\| \geq c\delta_{0T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - 2\eta_{0T}\right) \\ &= P_1 + P_2. \end{aligned} \quad (70)$$

Now note $P_1 \rightarrow 0$ by Lemma G.6. Now we show $P_2 \rightarrow 0$. This holds since $Q(\theta)$ has its maximum at θ_K^* . To be precise, note

$$Q(\theta_K^*) - Q(\theta) \geq C_1 \|\theta - \theta_K^*\|^2 \geq 2C_1 c^2 \eta_{0T}$$

and hence

$$\begin{aligned} \sup_{\|\theta - \theta_K^*\| \geq c\delta_{0T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) &\leq -2C_1 c^2 \eta_{0T} \text{ if } \|\theta - \theta_K^*\| = o(\zeta(K)^{-2}) \\ \sup_{\|\theta - \theta_K^*\| \geq c\delta_{0T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) &< -C_3 \zeta(K)^{-4} \text{ otherwise.} \end{aligned}$$

Therefore, as long as $C_1 c^2 > 1$ and $\zeta(K)^4 \eta_{0T} \rightarrow 0$, we have $P_2 \rightarrow 0$. $\zeta(K)^4 \eta_{0T} \rightarrow 0$ holds under $\alpha < 1/5$. Thus, we have proved $\|\widehat{\theta}_K - \theta_K^*\| = o_p(T^{-\kappa/2})$.

Now we refine the convergence rate by exploiting the local curvature of $\widehat{Q}_T(\theta)$ around θ_K^* . Let $\eta_{1T} = n^{-\kappa} \delta_{0T} = o(n^{-(\kappa+\kappa/2)})$ and $\delta_{1T} = \sqrt{\eta_{1T}} = o(T^{-(\kappa/2+\kappa/4)})$. For any c such that $C_1 c^2 > 1$,

we have²¹

$$\begin{aligned}
& \Pr \left(\left\| \widehat{\theta}_K - \theta_K^* \right\| \geq c\delta_{1T} \right) \\
& \leq \Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) \geq \widehat{Q}_T(\theta_K^*) \right) \\
& \leq \Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| > \eta_{1T} \right) \\
& \quad + \Pr \left(\left\{ \sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| \leq \eta_{1T} \right\} \right. \\
& \quad \left. \cap \left\{ \sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) \geq \widehat{Q}_T(\theta_K^*) \right\} \right) \quad (71) \\
& \leq \Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| > \eta_{1T} \right) \\
& \quad + \Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - \eta_{1T} \right) \\
& = P_3 + P_4
\end{aligned}$$

where $P_3 \rightarrow 0$ from Lemma G.7 and $P_4 \rightarrow 0$ similarly with P_2 noting

$$\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) \leq -C_1 c^2 \eta_{1T}$$

by (69) and since $\|\theta - \theta_K^*\| = o(\zeta(K)^{-2})$ for any θ such that $\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}$ under $\alpha < \frac{1}{5}$. This shows that $\left\| \widehat{\theta}_K - \theta_K^* \right\| = o_p(T^{-(\kappa/2 + \kappa/4)})$. Repeating this refinement for infinite number of times, we obtain

$$\left\| \widehat{\theta}_K - \theta_K^* \right\| = o_p(T^{-(\kappa/2 + \kappa/4 + \kappa/8 + \dots)}) = o_p(T^{-\kappa}) = o_p(T^{-1/2 + \alpha/2 + \delta})$$

under $\alpha < 1/5$.

2) Now we consider when η_{0T} has larger order than $o(\zeta(K)^{-4})$ (which holds under $\alpha \geq \frac{1}{5}$): Let $\widetilde{\delta}_{0T} = o(\zeta(K)^{-2}) T^\beta$ for $\beta > 0$. Then, from (70), we have

$$\begin{aligned}
& \Pr \left(\left\| \widehat{\theta}_K - \theta_K^* \right\| \geq c\widetilde{\delta}_{0T} \right) \\
& \leq \Pr \left(\sup_{\theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - Q(\theta) \right| > \eta_{0T} \right) + \Pr \left(\sup_{\|\theta - \theta_K^*\| \geq c\widetilde{\delta}_{0T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - 2\eta_{0T} \right) \\
& = P_1 + P_2.
\end{aligned}$$

Again note $P_1 \rightarrow 0$ from Lemma G.6. Now we show $P_2 \rightarrow 0$. Note from (69),

$$\sup_{\|\theta - \theta_K^*\| \geq c\widetilde{\delta}_{0T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) \leq -C_2 \zeta(K)^{-2} \|\theta - \theta_K^*\| \leq -o(\zeta(K)^{-4}) T^\beta$$

since $\|\theta - \theta_K^*\| > o(\zeta(K)^{-2})$ for any θ such that $\|\theta - \theta_K^*\| \geq c\widetilde{\delta}_{0T}$. It follows that $P_2 \rightarrow 0$ as long

²¹Note $\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| \leq \eta_{1T}$ implies, for any θ such that $\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}$,

$$\begin{aligned}
& -\eta_{1T} - \sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} (Q(\theta) - Q(\theta_K^*)) \\
& \leq \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) \leq \eta_{1T} + \sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} (Q(\theta) - Q(\theta_K^*))
\end{aligned}$$

and hence we obtain

$\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) \leq \eta_{1T} + \sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} (Q(\theta) - Q(\theta_K^*))$. Therefore, $\Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) > 0 \right) \leq \Pr \left(\sup_{\delta_{0T} \geq \|\theta - \theta_K^*\| \geq c\delta_{1T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - \eta_{1T} \right)$, from which we obtain the third inequality.

as $\zeta(K)^4 T^{-\beta} \eta_{0T} \rightarrow 0$, which holds under

$$\beta > 5\alpha/2 - 1/2 + \delta \quad (72)$$

and hence the convergence rate will be $o_p(\delta_{0T}) = o_p(T^{-\alpha+\beta})$. Now we refine the convergence rate by exploiting the local curvature of $\widehat{Q}_T(\theta)$ around θ_K^* again. Let $\widetilde{\eta}_{1T} = T^{-\kappa} \widetilde{\delta}_{0T} = o(T^{-((\alpha-\beta)+\kappa)})$ and $\widetilde{\delta}_{1T} = \sqrt{\widetilde{\eta}_{1T}} = o(T^{-((\alpha-\beta)/2+\kappa/2)})$. Then, from (71), we have

$$\begin{aligned} & \Pr \left(\left\| \widehat{\theta}_K - \theta_K^* \right\| \geq c \widetilde{\delta}_{1T} \right) \\ & \leq \Pr \left(\sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| > \widetilde{\eta}_{1T} \right) \\ & \quad + \Pr \left(\sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - \widetilde{\eta}_{1T} \right) \\ & = P_3 + P_{41} \end{aligned} \quad (73)$$

where $P_3 \rightarrow 0$ from Lemma G.7. Now we show $P_{41} \rightarrow 0$ similarly with P_2 . Here again we need to consider two cases:

2-1) When $\widetilde{\delta}_{1T}$ has equal or smaller order than $o(\zeta(K)^{-2})$, which holds under $\beta \leq \frac{1}{2} - \frac{3}{2}\alpha - \delta$ and hence from $\alpha > \beta$ and (72) it requires $1/5 \leq \alpha < 1/4$. Under this case, note

$$\begin{aligned} & \sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) \leq -C_1 \|\theta - \theta_K^*\|^2 \leq -C_1 c^2 \widetilde{\delta}_{1T}^2 = -C_1 c^2 \widetilde{\eta}_{1T} \\ & \text{if } \|\theta - \theta_K^*\| = o(\zeta(K)^{-2}) \\ & \sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} Q(\theta) - Q(\theta_K^*) \leq -C_2 \zeta(K)^{-2} \|\theta - \theta_K^*\| \leq -C_2 \zeta(K)^{-4} \text{ otherwise} \end{aligned}$$

by (69) and hence $P_{41} \rightarrow 0$ under $C_1 c^2 > 1$ and $\zeta(K)^4 \widetilde{\eta}_{1T} = \zeta(K)^4 \widetilde{\eta}_{1T} T^{-\kappa} \widetilde{\delta}_{0T} = \zeta(K)^2 o(T^{-\kappa} T^\beta) \rightarrow 0$, which holds under $\beta \leq 1/2 - 3\alpha/2 - \delta$. Repeating this refinement for infinite times (noting that for any θ such that $\widetilde{\delta}_{1T} \geq \|\theta - \theta_K^*\|$, we have $\|\theta - \theta_K^*\| = o(\zeta(K)^{-2})$), we obtain

$$\left\| \widehat{\theta}_K - \theta_K^* \right\| = o_p \left(T^{-\left(\lim_{L \rightarrow \infty} \left(\sum_{l=1}^L \kappa/2^l + (\alpha-\beta)/L \right) \right)} \right) = o_p(T^{-\kappa})$$

and hence the effect of η_{0T} 's having larger order than $o(\zeta(K)^{-4})$ disappears ($\frac{(\alpha-\beta)}{L}$ goes to zero as L goes to infinity). This makes sense because the iterated convergence rate improvement using the local curvature will dominate the convergence rate from the uniform convergence.

2-2) When $\widetilde{\delta}_{1T}$ has bigger order than $o(\zeta(K)^{-2})$, which holds under $\beta > 1/2 - 3\alpha/2 - \delta$ and $1/4 \leq \alpha < 1/3$:

In this case, we let $\widetilde{\delta}_{1T} = \widetilde{\delta}_{0T} n T^{-\gamma}$ for some $\gamma > 0$ and hence we require $\beta > \gamma$. From (73), we note

$$\begin{aligned} & \Pr \left(\left\| \widehat{\theta}_K - \theta_K^* \right\| \geq c \widetilde{\delta}_{1T} \right) \\ & \leq \Pr \left(\sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} \left| \widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_K^*) - (Q(\theta) - Q(\theta_K^*)) \right| > \widetilde{\eta}_{1T} \right) \\ & \quad + \Pr \left(\sup_{\widetilde{\delta}_{0T} \geq \|\theta - \theta_K^*\| \geq c \widetilde{\delta}_{1T}, \theta \in \Theta_T} Q(\theta) \geq Q(\theta_K^*) - \widetilde{\eta}_{1T} \right) \\ & = P_3 + P_{42}. \end{aligned}$$

We have seen that P_3 goes to zero since $\widetilde{\eta}_{1T} = T^{-\kappa} \widetilde{\delta}_{0T}$ and by Lemma G.7. Now we verify P_{42} goes to zero. From (69), to have $P_{42} \rightarrow 0$, we require that $\zeta(K)^{-2} \widetilde{\delta}_{1T}$ have a bigger order than $\widetilde{\eta}_{1T}$ and hence we need $\gamma < 1/2 - 3\alpha/2 - \delta$. Now we improve the convergence rate again using the local curvature

by defining $\tilde{\eta}_{2T} = T^{-\kappa}\tilde{\delta}_{1T} = o(T^{-((\alpha-\beta+\gamma)+\kappa)})$ and $\tilde{\delta}_{2T} = \sqrt{\tilde{\eta}_{2T}} = o(T^{-((\alpha-\beta+\gamma)/2+\kappa/2)})$. Then, similarly as before, at the end, we will obtain $\|\hat{\theta}_K - \theta_K^*\| = o_p(T^{-\kappa})$ as long as $\tilde{\delta}_{2T}$ has equal or smaller order than $o(\zeta(K)^{-2})$. The complicating case is again when $\tilde{\delta}_{2T}$ has a bigger order than $o(\zeta(K)^{-2})$, which happens when $\beta - \gamma > 1/2 - 3\alpha/2 - \delta$ but applying the same argument as before, we will obtain the same convergence rate of $\|\hat{\theta}_K - \theta_K^*\| = o_p(T^{-\kappa})$ as long as $\alpha < 1/3$. Combining these results, we conclude that under $\alpha < 1/3$, we have

$$\|\hat{\theta}_K - \theta_K^*\| = o_p(T^{-\kappa}) = o_p(T^{-1/2+\alpha/2+\delta}).$$

This result is intuitive in the sense that ignoring δ , we obtain $o(\zeta(K)^{-2}) = o(T^{-\kappa}) = T^{-1/3}$ at $\alpha = 1/3$ and hence when $\alpha \geq 1/3$, there is no room to improve the convergence rate using the local curvature. ■