University of California
Los Angeles

Essays in Microeconometrics and Industrial Organization

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Federico Zincenko

2013

Abstract of the Dissertation

Essays in Microeconometrics and Industrial Organization

by

Federico Zincenko

Doctor of Philosophy in Economics

University of California, Los Angeles, 2013

Professor Rosa L. Matzkin, Chair

This dissertation is divided into three chapters. In Chapter 1, I propose a nonparametric estimator for the bidders' utility function and the density of private values in a first-price sealed-bid auction model. Specifically, I study a setting with risk-averse bidders within the independent private value paradigm. I adopt a fully nonparametric approach by not placing any restrictions on the shape of the bidders' utility function beyond strict monotonicity, concavity, and differentiability. In contrast to previous literature, I characterize such utility function and the density of private values by a minimizer of a certain functional. I estimate this minimizer, which is a smooth real-valued function, in two steps by the method of sieves. Then, the estimators for the bidders' utility function and the density of private values are smooth functionals of the estimator for the minimizer. The estimator for the utility function is uniformly consistent and shape-preserving, while the estimator for the density is uniformly consistent and asymptotically normal.

Chapter 2, which is a joint paper with I. Obara, studies a model of repeated Bertrand competition among asymmetric firms that produce a homogeneous product. The discounting rates and marginal costs may vary across firms. We identify the critical level of discount factor such that a collusive outcome can be sustained if and only if the average discount factor within the lowest cost firms is above the

critical level. We also characterize the set of all efficient collusive equilibria when firms differ only in their discounting rates. Due to differential discounting, impatient firms gain a larger share of the market at an earlier stage of the game and patient firms gain a larger share at a later stage in efficient equilibrium. Although there are many efficient collusive equilibria, our model provides a unique prediction in the long run in the sense that every efficient collusive equilibrium converges to the unique efficient stationary collusive equilibrium within finite time.

Chapter 3 develops a weighted average derivative estimator for β in the context $E(y|x^c, x^d) = G(x^{c\prime}\beta, x^d)$, where x^c and x^d are continuous and discrete random vectors, respectively, and G is an unknown function. A distinguishing feature of the proposed estimator is the use of kernel smoothing for the discrete covariates. Under standard regularity conditions, such an estimator is root-N-consistent, asymptotically normal, and non-iterative.

The dissertation of Federico Zincenko is approved.

Sushil Bikhchandani

Jinyong Hahn

Ichiro Obara

Rosa L. Matzkin, Committee Chair

University of California, Los Angeles

2013

To my parents and my wife,

who —among other things—

gave me unconditional support and encouragement

while I was writing this manuscript

Table of Contents

List of Figures

List of Tables

Acknowledgments

Chapter 1 is the last version of my job market paper. I am especially grateful to my advisors for their guidance and encouragement. For useful comments and discussions, I thank Arie Beresteanu, Zhipeng Liao, Enrique Seira, Tridib Sharma, and Artyom Shneyerov, as well as, the seminar participants at UC Los Angeles, ITAM, University of Pittsburgh, 2012 California Econometrics Conference (UC Davis), and 2012 Graduate Students Conference (WU Saint Louis). Chapter 1 benefited from the financial support of the Dissertation Year Fellowship and the Welton Prize Award.

Chapter 2 is a joint paper with Ichiro Obara. We thank the audience at Cal Poly, Cornell, UCLA, Yale, 2011 Asian Meeting of the Econometric Society, and 2012 Southwest Economic Theory Conference. A previous version of Chapter 3 was submitted as the field paper for the Econometrics sequence, during the 2nd year of the Ph.D. program.

VITA

Education

2010 M.A. in Economics, UCLA

2007 M.A. in Economics, Universidad de San Andrés (Argentina)

2004 B.A. in Economics, Universidad de San Andrés (Argentina)

Awards

2012 Dissertation Year Fellowship and Welton Prize Award, UCLA

2009 Department Progress Award, UCLA

Employment

2009-2012 Teaching Assistant, Department of Economics, UCLA

2005-2008 Teaching Assistant, Department of Economics, Universidad de San Andrés (Argentina)

Presentations

2013 Instituto Tecnológico Autónomo de Mexico; University of Pittsburgh; New Economic School (Moscow, Russia); Royal Holloway, University of London

2012 Southwest Economic Theory Conference (UC San Diego); California Econometrics Conference (UC Davis); 7th Annual Graduate Student Conference (WU St. Louis)

CHAPTER 1

Nonparametric Sieve Estimation in First-Price Auctions with Risk-Averse Bidders

1.1 Introduction

Risk aversion is essential to understanding economic decisions under uncertainty. In first-price sealed-bid auctions, risk aversion plays a fundamental role in explaining bidders' behavior. Although several families of utility functions have been employed to describe different attitudes toward risk, in practice, we do not know which one accurately explains bidders' behavior.

In this paper, I develop an estimator for the utility function of risk-averse bidders, which in contrast to previous work, is nonparametric. I consider a first-price sealed-bid auction with risk-averse bidders within the paradigm of independent private values. In this setting, each potential buyer has his own private value for the item being sold and makes a sealed bid. The buyer who makes the highest bid wins the item and pays the seller the amount of that bid. This model is completely characterized by two objects. The first is the bidders' utility function, which describes bidders' risk preferences. The second is the density of private values, which describes the distribution of valuations for the auctioned item. Estimation of these two objects has proceeded by assuming that the econometrician observes all the bids and that the common utility function of the bidders is known up to a finite-dimensional parameter.

This paper develops an estimator that imposes no parametric specification on the common utility function of the bidders. Only strict monotonicity, concavity, and differentiability of this utility function is assumed. These assumptions are satisfied by linear, constant relative risk aversion (CRRA), and constant absolute risk aversion (CARA) utility functions, as well as, many others. In this sense, my paper generalizes the empirical analysis of first-price auctions by nesting many existing estimators within a fully nonparametric framework.

This paper has two objectives. The first is to nonparametrically estimate the bidders' utility function. Despite its relevance, only a few papers have proposed an estimator for such a function. [CGP11], for instance, adopts a semi-parametric approach and propose an estimator for the bidders' risk aversion parameter. Their approach requires that the researcher imposes a parametric specification –such as CRRA or CARA– on the bidders' utility function before estimating the risk aversion parameter and the density of private values. In real-world applications, the choice of the parametric specification may be arbitrary and not always realistic. In addition, there is no general agreement on which specification is the right one; when the choice is incorrect, the resulting estimator is invalid.

The second objective of my paper is to estimate the latent density of private values following a fully nonparametric perspective. To that end, I propose an estimator for the density of private values that does not rely on any parametric specification of the bidders' utility function. The main advantage of this approach is that the resulting estimator is robust to misspecification of such utility function. A common practice when estimating first-price auctions is to first assume a specific family of risk aversion for the bidders' utility, and then, estimate the density of private values. This procedure has been justified so far because of its low implementation costs and the possibility of attaining the optimal global rate of convergence ([GPV00] and [CGP11]). However, it can be criticized because an incorrect choice of the family of risk aversion invalidates the estimator for the density of private values.

Several papers have developed nonparametric estimators for the density of private values under the assumption that bidders are risk-neutral. The pioneering work [GPV00] constructed the first estimator to attain the optimal global rate of convergence. Recently, [MS12] has proposed an alternative estimator that is asymptotically normal and also attains the optimal rate. [BS12] has used integrated simulated moments to propose an estimator and construct a test for the validity of the first-price auction model. Here, I build on previous work by allowing bidders to be risk-averse.

My estimator for the density of private values is asymptotically normal and uniformly consistent. I derive these asymptotic properties by extending the approach of [MS12] to accommodate risk aversion from a nonparametric perspective. This has two advantages over existing work. First, empirical and experimental evidence indicates that risk aversion is a fundamental component of bidders' behavior (see [GPV09], Section 1, as well as the references cited therein); therefore, invoking risk neutrality is likely to generate erroneous conclusions. Second, there is no evidence telling us which concept of risk aversion is the most appropriate to describe bidders' risk preferences; therefore, it is essential to adopt a nonparametric approach.[1]

To my knowledge, only two papers have analyzed the identification of the bidders' utility function from a nonparametric perspective. [LP08] identified and estimated such a function by exploiting two auction designs, namely, ascending and first-price sealed-bid auctions. [GPV09] improved on [LP08] and identified the bidders' utility function by using the latter design only. They showed that the bidders' utility function is nonparametrically identified under some exclusion restrictions. Their primary exclusion restriction was exogenous bidders' participation. This exclusion restriction means that the distribution of valuations, or

[1] Regarding the experimental evidence, I highlight [Del08], whose "findings are not inconsistent with a role for risk aversion in the tendency to bid too high."

equivalently, the density of private values, is independent of the number of bidders. Exploiting this restriction, [GPV09] developed their constructive identification strategy. However, such a strategy is recursive and based on an infinite series of differences in quantiles, so it does not lead to a natural estimator for the bidders' utility function. Here, my contribution is to develop a valid estimator.

Assuming that bidders' participation is exogenous, I develop a convenient identification procedure that allows us to estimate the objects of interest: the bidders' utility function and the density of private values. Specifically, I characterize these objects by an argument that minimizes a certain functional over a space of smooth functions; in other words, the bidders' utility function and the density of private values are characterized by a minimizer of a certain functional. Such a minimizer is a smooth real-valued function and becomes the (infinite-dimensional) parameter of interest. I nonparametrically estimate this infinite-dimensional parameter in two steps by the method of sieve extremum estimation. This method optimizes an empirical criterion function over a sequence of finite-dimensional approximation spaces (sieve spaces); see [Che07]. The validity of the resulting estimator for the parameter of interest relies on the assumption of exogenous bidders' participation.

The estimators for the bidders' utility function and the density of private values are smooth nonlinear functionals of the sieve estimator for the parameter of interest. In particular, the estimator for the utility function is uniformly consistent and preserves the basic properties of the utility function (strict monotonicity, concavity, and differentiability). This shape-preserving feature arises as I use the Bernstein polynomials to estimate the parameter of interest. As noted by [Mat94], shape-preserving estimators have many advantages, among others, they decrease the variance and improve the quality of an extrapolation beyond the support of the data.

This paper is related to a vast literature on empirical industrial organization. First, it relates to the literature on structural econometrics of auction data. This literature is extensive and has expanded at an extraordinary rate; for example, see the surveys [HP95], [Laf97], [PV99], [AH07], and [HP07], as well as the textbook [PHH06]. I remark that nonparametric approaches have become very popular as auction data has become more available. Second, this paper is also related to the literature on recovering risk preferences from observed behavior. Within this line of research, I highlight [Lu04] and [AHS11]. The former proposes a semiparametric method to estimate the risk aversion parameter, as well as the risk premium, in the context of a first-price sealed-bid auction with stochastic private values. The latter considers a buy price auction framework and nonparametrically identifies both time and risk preferences of the bidders. Furthermore, it is worthwhile to mention [CE07] that estimate risk preferences from data on deductible choices in auto insurance contracts.

The results obtained in my paper are relevant for public policy recommendations. First-price sealed-bid auctions are used in many socio-economic contexts such as timber sales ([ALS11]), outer continental shelf wildcat auctions ([LPV03]), as well as competitive sales of municipal bonds ([Tan11]). In particular, using data from U.S. timber auctions, [ALS11] showed that first-price sealed-bid auctions generate higher revenue than open ascending auctions. In order to establish adequate auction rules that maximize the auctioneer's revenue, we need robust information about bidders' risk preferences. For instance, when bidders are risk-neutral, the auctioneer's expected revenue is the same under a first-price and second-price sealed-bid auction. However, when bidders are risk-averse, the first-price auction design generates more expected revenue than the latter ([HMZ10]). Moreover, the optimal reserve price depends on both the risk preferences and the distribution of valuations.

The rest of this paper is organized as follows. The remaining part of this section presents an sketch of my methodology at a technical level. Section 1.2 describes the auction model and establishes the key identification assumption, that is, exogenous bidders' participation. Section 1.3 introduces the (infinite-dimensional) parameter of interest within a general framework and presents the main mathematical results. Section 1.4 defines the two-step nonparametric sieve estimator for the parameter of interest and establishes its uniform consistency. Section 1.5 provides estimators for the auction model's objects, the bidders' utility function and the density of private values, and establishes their asymptotic properties. Section 1.6 reports the results of a limited Monte Carlo study, and also, presents an empirical illustration. Section 1.7 concludes with a discussion of possible extensions. Proofs of all results are given in the Appendix.

Sketch of Methodology: Let $\lambda_0^{-1}(\cdot)$ be a smooth real-valued function, which characterizes both the bidders' utility function and the density of private values (Section 1.2). The function $\lambda_0^{-1}(\cdot)$ is the parameter of interest (subsection 1.3.1), so the main objective of this paper is to nonparametrically estimate $\lambda_0^{-1}(\cdot)$. The idea behind my approach is summarized as follows.

In subsection 1.3.2, I construct two population criterion functions $Q_1 : \mathscr{A} \to \mathbb{R}_{\geq 0}$ and $Q_2 : \mathscr{A} \times \mathscr{H}_R \to \mathbb{R}_{\geq 0}$, where $\mathscr{A}$ is a set of sequences of functions and $\mathscr{H}_R$ is a space of smooth functions that contains $\lambda_0^{-1}(\cdot)$, that is, $\lambda_0^{-1}(\cdot) \in \mathscr{H}_R$. In particular, $Q_1(\cdot)$ and $Q_2(\cdot,\cdot)$ satisfy the following identification property: $\|\phi(\cdot) - \lambda_0^{-1}(\cdot)\|_\infty^c \leq Q_1[(\alpha_t)_t] + Q_2[(\alpha_t)_t, \phi]$ for all $[(\alpha_t)_t, \phi] \in \mathscr{A} \times \mathscr{H}_R$, where $\|\cdot\|_\infty$ denotes the sup-norm and c is a finite constant greater than one that depends on the smoothness of $\lambda_0^{-1}(\cdot)$. From this identification property, $\lambda_0^{-1}(\cdot)$ can be characterized as the unique minimizer of the functional $Q_2[(\tilde{\alpha}_t)_t, \cdot] : \mathscr{H}_R \to \mathbb{R}_{\geq 0}$, where $(\tilde{\alpha}_t)_t$ is a sequence of functions that satisfies $Q_1[(\tilde{\alpha}_t)_t] = 0$.

In subsection 1.3.3, I construct the sieve: an increasing sequence $\{\mathscr{A}^{(n)} \times \mathscr{H}^{(n)} \subseteq \mathscr{A} \times \mathscr{H}_R : n \in \mathbb{N}\}$ of finite-dimensional approximation spaces. The sieve spaces $\mathscr{A}^{(n)}$ and $\mathscr{H}^{(n)}$ are built on wavelet and Bernstein polynomial spaces, respectively, and satisfy the following approximation property: there exists a sequence $[(A_t^{(n)})_t, P^{(n)}] \in \mathscr{A}^{(n)} \times \mathscr{H}^{(n)}$ such that $Q_1[(A_t^{(n)})_t] \to 0$ and $Q_2[(A_t^{(n)})_t, P^{(n)}] \to 0$ as n grows to infinity. Basically, $(A_t^{(n)})_t$ approximates the sequence of functions $(\tilde{\alpha}_t)_t$, which satisfies $Q_1[(\tilde{\alpha}_t)_t] = 0$, while $P^{(n)}(\cdot)$ approximates to $\lambda_0^{-1}(\cdot)$ because $\|P^{(n)}(\cdot) - \lambda_0^{-1}(\cdot)\|_\infty^c \leq Q_1[(A_t^{(n)})_t] + Q_2[(A_t^{(n)})_t, P^{(n)}] \to 0$.

In Section 4, I define the estimator for $\lambda_0^{-1}(\cdot)$ using observed bids, which are obtained from a sample of N independent auctions. The shapes of the population criterion functions $Q_1(\cdot)$ and $Q_2(\cdot,\cdot)$ naturally lead to their empirical counterparts $\hat{Q}_1(\cdot)$ and $\hat{Q}_2(\cdot,\cdot)$, respectively. Both $\hat{Q}_1(\cdot)$ and $\hat{Q}_2(\cdot,\cdot)$ are computed from observed bids and converge uniformly in probability to $Q_1(\cdot)$ and $Q_2(\cdot,\cdot)$, respectively, as the sample size N grows to infinity. The estimator $\hat{\lambda}^{-1}(\cdot)$ of $\lambda_0^{-1}(\cdot)$ is computed in two steps. In the first one, we define $(\hat{A}_t^{(N)})_t$ as the argument that minimizes $\hat{Q}_1(\cdot)$ over $\mathscr{A}^{(N)}$; basically, $(\hat{A}_t^{(N)})_t$ is the empirical counterpart of $(A_t^{(N)})_t$. In the second step, $\hat{\lambda}^{-1}(\cdot)$ is defined as the argument that minimizes $\hat{Q}_2[(\hat{A}_t^{(N)})_t, \cdot]$ over $\mathscr{H}^{(N)}$; basically, $\hat{\lambda}^{-1}(\cdot)$ is the empirical counterpart of $P^{(N)}(\cdot)$. The uniform consistency of $\hat{\lambda}^{-1}(\cdot)$ relies on the idea that $P^{(N)}(\cdot)$ converges uniformly to $\lambda_0^{-1}(\cdot)$ as $N \to +\infty$. Then, in Section 1.5, the estimators for the bidders' utility function and the density of private values are constructed as nonlinear smooth functionals of $\hat{\lambda}^{-1}(\cdot)$.

1.2 First-Price Auction Model

In this section, I present the model and establish the key identification assumption. Subsection 1.2.1 describes the model, which is standard in the auction literature: a first-price sealed-bid auction with risk-averse bidders, independent private values, and a non-binding reserve price. Within this framework, I set my objective: estimating the bidders' utility function and the density of private values. Subsection 1.2.2 discusses existing identification results and establishes the key identification assumption, that is, exogenous bidders' participation.

1.2.1 Model

A single indivisible object is sold through a first-price sealed-bid auction with non-binding reserve price. In other words, the object is sold to the highest bidder who pays his bid to the seller and each bidder does not know others' bids when forming his bid. Within the independent private value (IPV) paradigm, each bidder knows his own private value v, but not other bidders' private values. There are $I \geq 2$ bidders and private values are drawn independently from a common cumulative distribution function (c.d.f.) $F(\cdot|I)$. Such a distribution is twice continuously differentiable with density $f(\cdot|I)$ and a compact support $[\underline{v}(I), \bar{v}(I)] \subseteq \mathbb{R}_{\geq 0}$. Both I and $F(\cdot|I)$ are common knowledge.

All bidders are identical ex ante and the game is symmetric. Each bidder has the same univariate utility function $U(\cdot)$ that is independent of I. If a bidder with value v wins and pays $b \geq 0$, his utility is $U(v-b)$, and if he loses, his utility is $U(0)$; see [MR84], Section 1, Case 1. Since any bidder's payment must be smaller or equal than his own valuation, the domain of $U(\cdot)$ is restricted to $\mathbb{R}_{\geq 0}$. Bidder i with valuation v_i maximizes his expected utility

$$E(\Pi_i) = U(v_i - b_i)\Pr(b_i \geq b_j, j \neq i),$$

with respect to his bid b_i, where b_j is the jth-player's bid. It is also assumed that $U(\cdot)$ is twice continuously differentiable, $U(0) = 0$, $U'(\cdot) > 0$, and $U''(\cdot) \leq 0$.

Only symmetric Bayesian Nash equilibria are considered. As a consequence, there exists a unique symmetric equilibrium bidding function $s(\cdot; I)$; see [HMZ10], Section 2, and the references cited therein. Such a function is strictly increasing, continuous on $[\underline{v}(I), \bar{v}(I)]$, and continuously differentiable on $(\underline{v}(I), \bar{v}(I)]$. Moreover, it satisfies the differential equation

$$s'(v; I) = (I-1)\frac{f(v|I)}{F(v|I)}\lambda_0(v-b) \tag{1.1}$$

with boundary condition $s[\underline{v}(I); I] = \underline{v}(I)$, where $b = s(v; I)$ is the optimal bid, $s'(v; I)$ denotes the first derivative of $s(v; I)$ with respect to v, and

$\lambda_0(\cdot) \equiv U(\cdot)/U'(\cdot)$. From equation (1.1), the equilibrium bidding function can also be written as

$$s(v;I) = v - \lambda_0^{-1}\left\{\frac{s'(v;I)F(v|I)}{(I-1)f(v|I)}\right\},$$

where $\lambda_0^{-1}(\cdot)$ denotes the inverse of $\lambda_0(\cdot)$. Note that the negative difference between a bid and its corresponding valuation depends crucially on both the bidders' risk preferences, which are represented by $\lambda_0^{-1}(\cdot)$, and the distribution of valuations.

Given the above framework, the model can be characterized by the objects $U(\cdot)$ and $F(\cdot|I)$. It is assumed that $U(\cdot)$ and $F(\cdot|I)$ satisfy the regularity conditions of [GPV09]. Specifically, $[U(\cdot), F(\cdot|I)] \in \mathcal{U}_R \times \mathcal{F}_R$ where $R \in \mathbb{N}$ and the corresponding sets are defined as follows.

Definition 1. *Let $\mathcal{U}_R$ be the set of utility functions $U : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ that satisfy the next conditions: $U(0) = 0$ and $U(\bar{y}) = 1$ for some $\bar{y} > 0$; $U(\cdot)$ is continuous on $\mathbb{R}_{\geq 0}$ and admits $R+1$ continuous derivatives on $\mathbb{R}_{>0}$ with $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$; $\lim_{x\downarrow 0} \nabla^r \lambda_0(x)$ is finite for $r = 1, 2, \ldots, R+1$, where $\nabla^r \lambda_0$ stands for the r-th derivative of λ_0.*

Definition 2. *Let $\mathcal{F}_R$ be the set of distribution functions $F(\cdot|I)$ that satisfy the next conditions: $F(\cdot|I)$ is a c.d.f. with support $[\underline{v}(I), \bar{v}(I)]$ and $0 \leq \underline{v}(I) < \bar{v}(I) < +\infty$; $F(\cdot|I)$ admits $R+1$ continuous derivatives on $[\underline{v}(I), \bar{v}(I)]$; and $f(\cdot|I)$ is bounded away from zero on $[\underline{v}(I), \bar{v}(I)]$.*

The main objective of this paper is to estimate the utility function $U(\cdot)$ and the density of private values $f(\cdot|I)$. Before doing so, the next subsection discusses existing identification results and establishes the key identification assumption.

1.2.2 Identification Assumption: Exogenous Participation

Suppose that the number of bidders I is observed and the distribution $G(\cdot|I)$ of an equilibrium bid is known. Let $v(\alpha|I)$ and $b(\alpha|I)$ denote the α-quantile of

$F(\cdot|I)$ and $G(\cdot|I)$, respectively.[2] Since $s(\cdot;I)$ is strictly increasing, we have that $b(\alpha|I) = s[\upsilon(\alpha|I);I]$ for any $\alpha \in [0,1]$. From [GPV09], we can reformulate the differential equation (1.1) as

$$1 = (I-1)\frac{g(b|I)}{G(b|I)}\lambda_0(\upsilon - b),$$

where $b \in (\underline{\upsilon}(I), \bar{b}(I)]$, $\bar{b}(I) = s[\bar{\upsilon}(I);I]$, $\upsilon = s^{-1}(b;I)$, and $g(\cdot|I) = G'(\cdot|I)$. After elementary algebra and since $\lambda_0(\cdot)$ is strictly increasing, it follows that

$$\upsilon(\alpha|I) = b(\alpha|I) + \lambda_0^{-1}\left\{\frac{\alpha}{(I-1)g[b(\alpha|I)|I]}\right\} \tag{1.2}$$

for any $\alpha \in [0,1]$.

Expression (1.2) is useful to derive the smoothness conditions of the equilibrium bid distribution $G(\cdot|I)$. To characterize such conditions, the following set of univariate distributions is defined.

Definition 3. *Let $\mathcal{G}_R$ be the set of distributions $G(\cdot|I)$ that satisfy the next conditions: $G(\cdot|I)$ is a c.d.f. with support $[\underline{b}(I), \bar{b}(I)]$ and $0 \leq \underline{b}(I) < \bar{b}(I) < +\infty$; $G(\cdot|I)$ admits $R+1$ derivatives on $[\underline{b}(I), \bar{b}(I)]$; $g(\cdot|I)$ admits $R+1$ continuous derivatives on $(\underline{b}(I), \bar{b}(I)]$ and is bounded away from zero on its support; $\lim_{b \downarrow \underline{b}(I)} d^r[G(b|I)/g(b|I)]/db^r$ exists and is finite for $r = 1, 2, \ldots, R+1$.*

A distribution $G(\cdot|I)$ is said to be *rationalized* by an auction model if there exists a structure $[U(\cdot), F(\cdot|I)]$ whose equilibrium bid distribution is $G(\cdot|I)$. From [GPV09], Proposition 1, any bid distribution $G(\cdot|I) \in \mathcal{G}_R$ can be rationalized by $[U(\cdot), F(\cdot|I)]$ if and only if $[U(\cdot), F(\cdot|I)] \in \mathcal{U}_R \times \mathcal{F}_R$. Furthermore, [CGP11] improved on this result and showed that any $G(\cdot|I) \in \mathcal{G}_R$ can be rationalized by some $[U(\cdot), F(\cdot|I)] \in \mathcal{U}_R \times \mathcal{F}_R$ even when $U(\cdot)$ is restricted to belong to parametric families of risk aversion such as CRRA and CARA.

[2]In this paper, the α-quantile of any c.d.f. $F(\cdot)$ is defined by $q(\alpha) \equiv \inf\{b \in \mathbb{R} : F(b) \geq \alpha\}$ for $\alpha \in (0,1]$, whereas $q(0) \equiv \sup\{b \in \mathbb{R} : F(b) \leq 0\}$. Note $q(0)$ and $q(1)$ become the infimum and supremum, respectively, of the support of the density $f(\cdot)$.

A structure $[U(\cdot), F(\cdot|I)] \in \mathcal{U}_R \times \mathcal{F}_R$ is said to be *nonidentified* if there is a another different structure $[\tilde{U}(\cdot), \tilde{F}(\cdot|I)] \in \mathcal{U}_R \times \mathcal{F}_R$ that leads to the same equilibrium bid distribution. If no such a structure exists, the model is said to be *identified*. [GPV09] has shown that any element of $\mathcal{U}_R \times \mathcal{F}_R$ is nonidentified from the knowledge of $G(\cdot|I)$. However, they have also shown that model is identified under some exclusion restrictions. Their main exclusion restriction is exogenous bidders' participation: $F(\cdot|I)$ does not depend on the number of bidders I, or more specifically, $F() \equiv F(\cdot|I_1) = F(\cdot|I_2)$ for at least two number of bidders I_1 and I_2 with $2 \leq I_1 < I_2$. Under this restriction, [GPV09] proved that $[U(\cdot), F(\cdot)]$ is identified from the knowledge of the conditional bid distributions $G(\cdot|I_1)$ and $G(\cdot|I_2)$.

In the rest of the paper, it is assumed that that bidders' participation is exogenous, or equivalently, $f(\cdot) \equiv f(\cdot|I_1) = f(\cdot|I_2)$ for at least two numbers of bidders $2 \leq I_1 < I_2$. As noted earlier, the main objective of this paper is to estimate the utility function $U(\cdot)$ and the density of private values $f(\cdot)$. By definition of $\lambda_0(\cdot)$ and expression (1.2), both $U(\cdot)$ and $f(\cdot)$ can be characterized by $\lambda_0^{-1}(\cdot)$ from the knowledge of $G(\cdot|I_1)$ and $G(\cdot|I_2)$, where $2 \leq I_1 < I_2$. As a consequence, $\lambda_0^{-1}(\cdot)$ becomes the (infinite-dimensional) parameter of interest for the next two sections.

1.3 Approximation via Sieve Spaces

This section presents the main mathematical results. In preparation, subsection 1.3.1 formally introduces the parameter of interest, $\lambda_0^{-1}(\cdot)$, within a general framework. Subsection 1.3.2 establishes the nonparametric identification result, which allows us to uniquely characterize $\lambda_0^{-1}(\cdot)$ within a space of smooth functions. The identification result is achieved using two population criterion functions. Subsection 1.3.3 constructs sieve spaces to approximate the zeros of those functions.

Before proceeding, I lay out the notation for the remaining discussion. The set of nonnegative integers $\mathbb{N} \cup \{0\}$ is denoted by $\mathbb{N}_0$. The usual conventions $0! \equiv 1$ and $0^0 \equiv 1$ are adopted. For $x \in \mathbb{R}$, the ceiling function is denoted by $\lceil x \rceil \equiv \min\{n \in \mathbb{Z} : x \leq n\}$. For a real-valued function f and a set Z, the range of f over Z is denoted by $f(Z)$, and I use standard notation for L_q-norms: $\|f\|_{q,Z} = [\int_Z |f(z)|^q dz]^{1/q}$ if $1 \leq q < +\infty$, whereas $\|f\|_{\infty,Z} = \sup_{z \in Z} |f(z)|$. The indicator function is denoted by $\mathbb{1}\{\cdot\}$, $\nabla^r f$ stands for the r-th derivative of f, and $\nabla^0 f \equiv f$.

1.3.1 Parameter of Interest

Given i_1 and i_2 integers that satisfy $2 \leq i_1 < i_2$, consider two univariate distributions $G_1(\cdot) \equiv G(\cdot|i_1)$ and $G_2(\cdot) \equiv G(\cdot|i_2)$ that belong to $\mathcal{G}_R$ (see Definition 3). The pair $[G_1, G_2] \in \mathcal{G}_R^2$ is the underlying data distribution, and given the previous auction model, it is associated to a pair of bids (b_1, b_2) where the integers (i_1, i_2) represent the numbers of participants in each auction. In other words, b_1 and b_2 can be regarded as random variables generated by two different auctions with i_1 and i_2 bidders, respectively.

The distribution of an equilibrium bid $G_j(\cdot)$, $j = 1, 2$, varies with the number of bidders i_j because the equilibrium bidding strategy varies with i_j. In addition, the exclusion restriction $F(\cdot) \equiv F(\cdot|i_1) = F(\cdot|i_2)$, together with equation (1.1), imposes additional constraints on the pair $[G_1, G_2]$. To formalize these constraints, I introduce the set of pairs of distributions $\mathcal{G}_R^*$.

Let $R \in \mathbb{N}$ and $\bar{H} \in \mathbb{R}_{>0}$ be fixed constants, and also, let $b_j(\alpha)$ denote the α-quantile of $G_j(\cdot)$.

Definition 4. *Let $\mathcal{G}_R^*$ be the collection of pairs of univariate distributions $[G_1, G_2] \in \mathcal{G}_R^2$ that satisfy the following conditions:*

1. $b_1(0) = b_2(0)$ *and* $b_1(\alpha) < b_2(\alpha)$ *for* $\alpha \in (0, 1]$.

2. *There exists a function* $\lambda_0 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *with* $R + 1$ *continuous derivatives,* $\lambda_0(0) = 0$*,* $\lambda_0'(\cdot) \geq 1$*, and* $\|\nabla^{R+1} \lambda_0^{-1}\|_{\infty, \mathbb{R}_{\geq 0}} \leq \bar{H}$*. Such function satisfies:*

(a) *For* $\alpha \in [0,1]$, *the compatibility condition*

$$\Delta b(\alpha) = \lambda_0^{-1}[R_1(\alpha)] - \lambda_0^{-1}[R_2(\alpha)],$$

where $\Delta b(\alpha) \equiv b_2(\alpha) - b_1(\alpha)$ *and* $R_j(\alpha) \equiv \alpha b_j'(\alpha)/(i_j - 1)$.

(b) *For* $j \in \{1,2\}$ *and* $b \in [\underline{b}, \bar{b}_j]$, $\xi_j'(b) > 0$, *where* $\xi_j(b) \equiv b + \lambda_0^{-1}\{G_j(b)/[(i_j - 1)g_j(b)]\}$ *and* $g_j(\cdot)$ *stands for the density of* $G_j(\cdot)$.

For the remaining discussion, I assume that $[G_1, G_2] \in \mathcal{G}_R^*$, and the (infinite-dimensional) parameter of interest is the function $\lambda_0^{-1}(\cdot)$. To save notation, I write $\underline{b} \equiv b_1(0) = b_2(0)$ and $\bar{b}_j \equiv b_j(1)$. Basically, Definition 4 captures the restrictions on the distributions of bids derived from the auction model of subsection 1.2.1. The first condition means that participants bid more aggressively as the number of bidders increases. The second statement establishes a compatibility condition between two auctions with different numbers of bidders. More specifically, $\lambda_0(0) = 0$ is simply a normalizing restriction, $\lambda_0'(\cdot) \geq 1$ indicates that bidders are risk-averse, and $\|\nabla^{R+1}\lambda_0^{-1}\|_{\infty,\mathbb{R}_{\geq 0}}\bar{H} < +\infty$ is a regularity condition about the degree of smoothness of $\lambda_0^{-1}(\cdot)$. Condition 2.(a) formalizes the assumption that the distribution of valuations, as well as the bidders' utility function, does not depend on the number of bidders i_1 and i_2. Condition 2.(b) establishes the existence of an inverse bidding function that is consistent with expression (1.2).

From [GPV09], any element of $\mathcal{G}_R^*$ can be rationalized by the auction model of subsection 1.2.1 with the exclusion restriction $F(\cdot) \equiv F(\cdot|i_1) = F(\cdot|i_2)$, that is, exogenous participation. In other words, we already know that, for any $[G_1, G_2] \in \mathcal{G}_R^*$, there exists a (unique) structure $[U(\cdot), F(\cdot)] \in \mathcal{U}_R \times \mathcal{F}_R$ that is independent of the number of bidders and whose equilibrium bid distribution is $G_j(\cdot)$, when the number of bidders is i_j. Specifically, $[U(\cdot), F(\cdot)]$ can be obtained as follows. The utility function $U(\cdot)$ is the solution of the differential equation $\lambda_0(\cdot)U'(\cdot) - U(\cdot) = 0$ with an additional normalizing restriction such as $U(\bar{y}) = 1$ for some $\bar{y} > 0$. The α-quantile of the private value distribution $F(\cdot)$, which is independent of the number of bidders, becomes $v(\alpha) = \xi_1[b_1(\alpha)] = \xi_2[b_2(\alpha)]$. Note that $v(\cdot)$ is well-defined due to the compatibility condition 2.(a), so the distribution of private values is given by $F(\cdot) = v^{-1}(\cdot)$.

As an illustration, consider the pair of uniform distributions $[G_1, G_2]$ given by

$$G_j(b) = \int_{-\infty}^{b} \left(\frac{i_j}{i_j - 1} \right) \mathbb{1} \left\{ 0 \le v \le \frac{i_j - 1}{i_j} \right\} dv, \tag{1.3}$$

where $j = 1, 2$ and $b \in \mathbb{R}$. In this particular case, each distribution $G_j(\cdot)$ is generated by an auction model with i_j participants, risk-neutral bidders ($U(y) = y$), and independent private values distributed as uniform $[0,1]$ ($f(v) = \mathbb{1}\{0 \le v \le 1\}$). This illustration is useful to show that $\mathcal{G}_R^*$ is nonempty for any $2 \le i_1 < i_2$, $R \in \mathbb{N}$, and $\bar{H} > 0$; specifically, the pair $[G_1, G_2]$ defined by (1.3) belongs to $\mathcal{G}_R^*$ regardless of the values of $2 \le i_1 < i_2$, $R \in \mathbb{N}$, and $\bar{H} > 0$. Conditions 1-2 of Definition 4 can be easily checked. The first is satisfied as the α-quantiles of $G_1(\cdot)$ and $G_2(\cdot)$ are $b_1(\alpha) = (i_1 - 1)\alpha/i_1$ and $b_2(\alpha) = (i_2 - 1)\alpha/i_2$, respectively. The second condition is also satisfied because $\Delta b(\alpha) = (\alpha/i_1) - (\alpha/i_2)$ and $R_j(\alpha) = \alpha/i_j$, so the functions $\lambda_0(\cdot)$ and $\xi_j(\cdot)$ become $\lambda_0(u) = u$ and $\xi_j(b) = i_j b/(i_j - 1)$. Finally, note that both functions satisfy all the requirements of condition 2; in particular, $\|\nabla^{R+1}\lambda_0^{-1}\|_{\infty,\mathbb{R}_{\ge 0}} = 0 \le \bar{H}$.

1.3.2 Population Criterion Functions

In this subsection, I introduce the first methodological innovation of the paper: constructing appropriate population criterion functions that allow us to identify the parameter of interest, $\lambda_0^{-1}(\cdot)$, within a space of smooth functions. Recall that our auction model can be completely characterized by $\lambda_0^{-1}(\cdot)$, so after identifying this function, we can recover both the bidders' utility function and the density of private values.

Before proceeding to the nonparametric case, consider the polynomial case $\lambda_0^{-1}(u) = \beta_1 u + \beta_2 u^2 + \cdots + \beta_L u^L$ as an illustration, where $L \in \mathbb{N}$ and $(\beta_1, \beta_2, \ldots, \beta_L) \in$

$\mathbb{R}^L_{>0}$ are unknown coefficients. Since the compatibility condition $\Delta b(\alpha) = \lambda_0^{-1}[R_1(\alpha)] - \lambda_0^{-1}[R_2(\alpha)]$ holds for every $\alpha \in [0,1]$, it follows that

$$\begin{bmatrix} \Delta b(\alpha_0) \\ \Delta b(\alpha_1) \\ \vdots \\ \Delta b(\alpha_K) \end{bmatrix} = \begin{bmatrix} R_1(\alpha_0) - R_2(\alpha_0) & \dots & R_1(\alpha_0)^L - R_2(\alpha_0)^L \\ R_1(\alpha_1) - R_2(\alpha_1) & \dots & R_1(\alpha_1)^L - R_2(\alpha_1)^L \\ \vdots & \dots & \vdots \\ R_1(\alpha_K) - R_2(\alpha_K) & \dots & R_1(\alpha_K)^L - R_2(\alpha_K)^L \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{bmatrix}$$

for any $K \in \mathbb{N}$ and $(\alpha_0, \alpha_1, \dots, \alpha_K) \in [0,1]^{K+1}$. Observe that there are $K+1$ linear equations and L unknowns, $(\beta_1, \beta_2, \dots, \beta_L)$. To recover these coefficients, we need (at least) L independent equations, so the choice of K and the arguments $(\alpha_0, \alpha_1, \dots, \alpha_K)$ becomes crucial. The problem is that an arbitrary choice of $(\alpha_0, \alpha_1, \dots, \alpha_K)$ does not necessarily lead to a system of K independent equations because Definition 4 imposes mild restrictions on $R_1(\cdot)$ and $R_2(\cdot)$.

In what follows, I solve this problem by constructing two criterion functions. Basically, the first one will provide the proper arguments $(\alpha_0, \alpha_1, \dots, \alpha_K)$ to plug-in the compatibility condition and obtain at least L independent equations. The second criterion function will recover $(\beta_1, \beta_2, \dots, \beta_L)$ using the compatibility condition evaluated at the arguments provided by the first criterion function. Basically, a nonparametric specification of $\lambda_0^{-1}(\cdot)$ can be approximated by letting L grow to infinity, so for identification purposes, K must grow to infinity as well. The problem arises as some arguments $\alpha_t \in [0,1]$, $t = 0, 1, \dots, K$, will be close to each other, and so the rows of the above system of linear equations.

Turning to the nonparametric case where $\lambda_0^{-1}(\cdot)$ is a smooth function, let $[0, \bar{r}]$ denote the range of $R_1(\cdot)$ over $[0,1]$ with $\bar{r} \equiv \max\{R_1(\alpha) : \alpha \in [0,1]\}$. From [GPV09], we already know that $\lambda_0^{-1}(\cdot)$ is nonparametrically identified on $[0, \bar{r}]$, or in other words, $\lambda_0^{-1}(\cdot)$ is uniquely defined on $[0, \bar{r}]$. It can also be shown that $\lambda_0^{-1}(u)$ cannot be identified when $u > \bar{r}$, so the identification region $[0, \bar{r}]$ cannot be improved. To be specific, [GPV09] has established that

$$\lambda_0^{-1}(u_0) = \sum_{t \in \mathbb{N}_0} \Delta b(\tilde{\alpha}_t), \tag{1.4}$$

where $u_0 \in (0, \bar{r}]$ and $(\tilde{\alpha}_t)_{t \in \mathbb{N}_0} \subseteq (0,1)$ is a strictly decreasing sequence that satisfies the nonlinear recursive relation $R_1(\tilde{\alpha}_t) = R_2(\tilde{\alpha}_{t-1})$ with initial condition $R_1(\tilde{\alpha}_0) = u_0$. Since $R_1(\cdot)$ is not necessarily increasing, the sequence $(\tilde{\alpha}_t)_{t \in \mathbb{N}_0}$ is not necessarily unique.

At this point, it is not known whether expression (1.4) can lead to a valid estimator of $\lambda_0^{-1}(\cdot)$. As noted by [GPV09], Section 5, an estimation strategy based on (1.4) would have several problems. First, expression (1.4) does not provide the rate of at which $\lambda_0^{-1}(u_0)$ can be estimated because the characterization of $\lambda_0^{-1}(u_0)$ is recursive and based on an infinity series of differences in quantiles. Second, since Definition 4 does not guarantee the existence of a "Polynomial Minorant" for $R_1(\cdot)$, it is impossible to establish the rate at which $\lambda_0^{-1}(u_0)$ can be estimated. The reason is that $\lambda_0^{-1}(u_0)$ depends crucially on $\tilde{\alpha}_0$, which is implicitly defined by the equation $R_1(\tilde{\alpha}_0) = u_0$, and the rate of convergence of any estimator for $\tilde{\alpha}_0$ would depend on the "Polynomial Minorant" of $R_1(\cdot)$; see [CHT07], Condition C.2 and Theorem 3.1.

In the rest of this subsection, I develop a convenient identification approach based on two population criterion functions, which will allow us to build a valid estimator for $\lambda_0^{-1}(\cdot)$. As a starting point, I define the domains of these functions: a set of sequence of functions $\mathscr{A}$ and a space of smooth functions $\mathscr{H}_R$. Let $\bar{u} \in [\bar{r}, +\infty)$ be a fixed real number.

Definition 5. *Let $\mathscr{A}$ be the set of sequences of functions $(\alpha_t)_{t \in \mathbb{N}_0} : [0, \bar{u}] \to [0,1]^\infty$ that satisfy the next conditions: for each $t \in \mathbb{N}_0$, $\alpha_t : [0, \bar{u}] \to [0,1]$ is a Lebesgue measurable function, and there is $T \in \mathbb{N}$ such that $\alpha_t(\cdot) = 0$ for all $t \geq T$.*

Let $\mathscr{H}_R$ be the space of functions $\phi : [0, \bar{u}] \to \mathbb{R}_{\geq 0}$ that satisfy the next conditions: $\phi(0) = 0$, $\phi(\cdot)$ admits $R+1$ continuous derivatives on $[0, \bar{u}]$, $0 \leq \phi'(\cdot) \leq 1$, *and* $\|\nabla^{R+1}\phi\|_{\infty,[0,\bar{u}]} \leq \bar{H}$.

Several remarks are noteworthy. First, T is not uniformly bounded across $\mathscr{A}$, or in other words, each sequence of functions $(\alpha_t)_t \in \mathscr{A}$ possesses its own finite T. Second, $R+1$ indicates the degree of smoothness of $\mathscr{H}_R$ due to the restriction $\|\nabla^{R+1}\phi\|_{\infty,[0,\bar{u}]} \leq \bar{H}$. Third, by the second condition of Definition 4, $\lambda_0^{-1} : [0,\bar{u}] \to \mathbb{R}_{\geq 0}$ belongs to $\mathscr{H}_R$ because $\lambda_0'(\cdot) \geq 1$ and $\|\nabla^{R+1}\lambda_0^{-1}\|_{\infty,\mathbb{R}_{\geq 0}} \leq \bar{H}$. As a result, $\mathscr{H}_R$ becomes the (infinite-dimensional) parameter space.

Given Definition 5, I construct two population criterion functions $Q_1 : \mathscr{A} \to \mathbb{R}_{\geq 0}$ and $Q_2 : \Theta_R \to \mathbb{R}_{\geq 0}$, where the domain of the latter is given by $\Theta_R \equiv \mathscr{A} \times \mathscr{H}_R$. To be specific, such functions are defined as follows:

$$
\begin{aligned}
Q_1[(\alpha_t)_t] &= \int_0^{\bar{r}} |R_1[\alpha_0(u)] - u| du + \sum_{t=1}^{+\infty} \int_0^{\bar{r}} |R_1[\alpha_t(u)] - R_2[\alpha_{t-1}(u)]| du, \\
Q_2[(\alpha_t)_t, \phi] &= \sum_{t=0}^{+\infty} \int_0^{\bar{r}} |\Delta b[\alpha_t(u)] + \phi\{R_2[\alpha_t(u)]\} - \phi\{R_1[\alpha_t(u)]\}| du, \qquad (1.5)
\end{aligned}
$$

and also, we set $Q[(\alpha_t)_t, \phi] \equiv Q_1[(\alpha_t)_t] + Q_2[(\alpha_t)_t, \phi]$. Observe that $Q_1(\cdot)$ and $Q_2(\cdot,\cdot)$ are well-defined. In particular, we have that $Q_1(\cdot), Q_2(\cdot,\cdot) < +\infty$ because, by construction of $\mathscr{A}$, the series of expression (1.5) always involve a finite number of terms. Recall that for each $(\alpha_t)_t \in \mathscr{A}$, there is $T \in \mathbb{N}$ such that $\alpha_t(\cdot) = 0$ for all $t \geq T$, as well as, $R_1(0) = R_2(0)$ and $\Delta b(0) = 0$.

The idea behind the construction of these criterion functions and the choice of their functional forms can be described as follows. Observe first that $Q_2[(\alpha_t)_t, \lambda_0^{-1}] = 0$ for any $(\alpha_t)_t \in \mathscr{A}$ because of condition 2.(a) in Definition 4. However, $Q_2[(\alpha_t)_t, \phi] = 0$ does not necessarily imply $\phi(\cdot) = \lambda_0^{-1}(\cdot)$; for instance, if $0_{\mathscr{A}}$ denotes the sequence of zero functions, then $Q_2[0_{\mathscr{A}}, \phi] = 0$ for any $\phi \in \mathscr{H}_R$. In view of identification and estimation purposes, it is useful to establish a uniqueness result in the sense that $Q_2[(\alpha_t)_t, \phi] = 0$ implies $\phi(\cdot) = \lambda_0^{-1}(\cdot)$. A criterion function based exclusively on the compatibility condition $\Delta b(\cdot) - \lambda_0^{-1}[R_1(\cdot)] - \lambda_0^{-1}[R_2(\cdot)] = 0$ will fail to achieve this uniqueness result, so we need to introduce $Q_1(\cdot)$, which can be interpreted as a "first-step" criterion function. The role of $Q_1(\cdot)$ is to identify a sequence of functions $(\tilde{\alpha}_t)_t$ so that $Q_2[(\tilde{\alpha}_t)_t, \phi] = 0$ implies $\phi(\cdot) = \lambda_0^{-1}(\cdot)$, and therefore, $\lambda_0^{-1}(\cdot)$ becomes the unique minimizer of the functional $Q_2[(\tilde{\alpha}_t)_t, \cdot]$. Roughly speaking, such a sequence of functions will be characterized as the zero of $Q_1(\cdot)$, that is $Q_1[(\tilde{\alpha}_t)_t] = 0$.

The next proposition formalizes the above arguments by establishing a convenient identification result. Let define $c(r) = (R+2)/(R+1-r)$ for $r = 0, 1, \ldots, R$.

Proposition 1.3.1. *If $[G_1, G_2] \in \mathcal{G}_R^*$, there exists a constant $\underline{K} > 0$ such that $\underline{K}\|\nabla^r \phi - \nabla^r \lambda_0^{-1}\|_{\infty,[0,\bar{r}]}^{c(r)} \leq Q[(\alpha_t)_t, \phi]$ for all $[(\alpha_t)_t, \phi] \in \Theta_R$ and $r \in \{0, 1, \ldots, R\}$.*

The proof of this proposition is detailed in Appendix 1.A.1.1 and can be divided into two parts. The first determines an upper bound for $\|\nabla^r \phi - \nabla^r \lambda_0^{-1}\|_{\infty,[0,\bar{r}]}$ in terms of R, $\bar{H}$, r, and $\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]}$. Specifically, using Theorem 1 of [Gab67], it can be shown that there is $\underline{K} > 0$ such that $\underline{K}\|\nabla^r f - \nabla^r g\|_{\infty,[0,\bar{r}]}^{c(r)} \leq \|f - g\|_{1,[0,\bar{r}]}$ for all $f, g \in \mathscr{H}_R$; this approach is similar to that of [CS98], Appendix B. I remark that the inequality between the *sup*-norm and the L_1-norm depends crucially on the existence of a finite constant $\bar{H}$ such that $\|\nabla^{R+1}\phi\|_{\infty,[0,\bar{u}]} \leq \bar{H}$. Moreover, Theorem 2 of [Gab67] proves that the exponent $c(r)$ cannot be essentially improved. The second part of the proof exploits the shape of the criterion functions, and then, employs repeated triangular inequalities to show that $\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]} \leq 2Q[(\alpha_t)_t, \phi]$ for any $[(\alpha_t)_t, \phi] \in \Theta_R$. In order to obtain this result, the function $Q_1(\cdot)$ plays a fundamental role as the inequality $\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]} \leq \bar{K} Q_2[(\alpha_t)_t, \phi]$ does not necessarily holds for an arbitrary $(\tilde{\alpha}_t)_t \in \mathscr{A}$ and a fixed constant $\bar{K} < +\infty$. In addition, we also require that $\lambda_0^{-1}(0) = \phi(0) = 0$.

The distinguishing feature of Proposition 1.3.1, in comparison with the existing identification result, is that the (Sobolev) distance between $\lambda_0^{-1}(\cdot)$ and any function of $\mathscr{H}_R$ is bounded above by a known criterion function. This feature is essential to develop an estimator for $\lambda_0^{-1}(\cdot)$ and obtain its rate of convergence. In this sense, Proposition 1.3.1 plays the role of Identification Condition 3.1 in [Che07].

An implication of Proposition 1.3.1 is the following: if there exists a sequence of functions $(\tilde{\alpha}_t)_t$ that satisfies $Q_1[(\tilde{\alpha}_t)_t] = 0$, then $\lambda_0^{-1}(\cdot)$ can be characterized as the minimizer of the functional $Q_2[(\tilde{\alpha}_t)_t, \cdot] : \mathscr{H}_R \to \mathbb{R}_{\geq 0}$. The main problem with this approach is that $Q_1[(\alpha_t)_t] > 0$ for any $(\alpha_t)_t \in \mathscr{A}$ by construction of $\mathscr{A}$ and $Q_1(\cdot)$. Nevertheless, the next implication is still valid: if there is a sequence $\{[(\alpha_t^{(n)})_t, \phi^{(n)}] \in \Theta_R : n \in \mathbb{N}\}$ such that $Q[(\alpha_t^{(n)})_t, \phi^{(n)}] \to 0$ as n grows to infinity, then $\nabla^r \phi^{(n)} \to \nabla^r \lambda_0^{-1}$ uniformly on $[0, \bar{r}]$ for any $r = 0, 1, \ldots, R$. In addition, the rate of approximation of $\nabla^r \phi^{(n)}$ toward $\nabla^r \lambda_0^{-1}$ can be bounded by the rate at which $Q[(\alpha_t^{(n)})_t, \phi^{(n)}]$ converges to zero, The next subsection will prove not only that the sequence $[(\alpha_t^{(n)})_t, \phi^{(n)}]$ exists, but also that it belongs to finite-dimensional approximation spaces (sieve spaces).

1.3.3 Sieve Spaces: Definition and Approximation Result

The innovation here is to build sieve spaces that approximate the zeros of the population criterion functions $Q_1(\cdot)$ and $Q_2(\cdot, \cdot)$. These sieves will be employed later, in subsection 1.4.3, to define the estimator for $\lambda_0^{-1}(\cdot)$.

Roughly speaking, the elements of $\mathscr{A}$ and $\mathscr{H}_R$ have different degree of smoothness. The components of a sequence $(\alpha_t)_t \in \mathscr{A}$ are bounded Lebesgue measurable functions, whereas any element of $\mathscr{H}_R$ is just a nondecreasing function with degree of smoothness $R + 1$. Due to this difference, $\mathscr{A}$ and $\mathscr{H}_R$ must be approximated using different sieve spaces. On the one hand, $\mathscr{A}$ will be approximated by wavelets with basis function $\kappa(x) \equiv (1 - |x|)\mathbb{1}\{|x| < 1\}$ where $x \in \mathbb{R}$. On the other hand, $\mathscr{H}_R$ will be approximated by Bernstein polynomials. The notation for the Bernstein polynomials basis is

$$p_{J,j}(u) \equiv \binom{J}{j} \left(\frac{u}{\bar{u}}\right)^j \left(1 - \frac{u}{\bar{u}}\right)^{J-j},$$

where $u \in [0, \bar{u}]$, $j, J \in \mathbb{N}_0$, $j \leq J$, and $p_{j,-1}(u) = p_{j-1,j}(u) = 0$.

To explain the dimensions of the sieve spaces, let consider two increasing divergent sequences of positive integers: $(K_n)_{n\in\mathbb{N}}$ and $(L_n)_{n\in\mathbb{N}}$. Then, the sieve spaces $\mathscr{A}^{(n)} \subseteq \mathscr{A}$ and $\mathscr{H}^{(n)} \subseteq \mathscr{H}_R$ can be defined as follows, while the sequences $(K_n)_n$ and $(L_n)_n$ are employed to explain their dimensions.

Definition 6. *Let $\mathscr{A}^{(n)}$ be the space of sequence of functions $(A_t)_{t\in\mathbb{N}_0} : [0,\bar{u}] \to [0,1]^\infty$ of the form*

$$A_t(u) \quad = \quad \begin{cases} \sum_{j=0}^{\lceil 2^{J_t} \rceil} a_{t,j}\kappa[2^{J_t}(u/\bar{u}) - j] & \textit{if } 0 \le t \le K_n - 1, \\ 0 & \textit{otherwise}, \end{cases} \tag{1.6}$$

where $1 \le J_t < +\infty$ and $0 \le a_{t,j} \le 1$ for any $t = 0, 1, \dots, K_n - 1$ and $j = 0, 1, \dots, \lceil 2^{J_t} \rceil$.

Let $\mathscr{H}^{(n)}$ be the space of Bernstein polynomials $P : [0,\bar{u}] \to \mathbb{R}_{\ge 0}$ with degree L_n of the form

$$P(u) = \sum_{j=0}^{L_n} b_j p_{L_n,j}(u),$$

where the coefficients $\{b_j : j = 0, 1, \dots, L_n\}$ satisfy the following conditions:

1. *$b_0 = 0$,*

2. *$\bar{u}L_n^{-2} \le b_{j+1} - b_j \le \bar{u}L_n^{-1}$ for $0 \le j \le L_n - 1$,*

3. *and*

$$\left| \sum_{i=0}^{R+1} (-1)^{R+1-i} \binom{R+1}{i} b_{j+i} \right| \le \left(\frac{\bar{u}}{L_n} \right)^{R+1} \bar{H} \quad \textit{for } 0 \le j \le L_n - (R+1).$$

Several remarks are noteworthy. First, $\mathscr{A}^{(n)}$ and $\mathscr{H}^{(n)}$ are finite-dimensional spaces spanned by $\kappa(\cdot)$ and $p_{J,j}(\cdot)$, respectively. Note that $\kappa(\cdot)$ and $p_{J,j}(\cdot)$ are weighting functions that satisfy $\sum_{j\in\mathbb{Z}} \kappa(x - j) = 1$ and $\sum_{j=0}^{J} p_{J,j}(u) = 1$ for any $x \in \mathbb{R}$ and $u \in [0,\bar{u}]$. Second, $\mathscr{A}^{(n)}$ is well-defined and $\mathscr{A}^{(n)} \subseteq \mathscr{A}^{(n+1)} \subseteq \mathscr{A}$ for all $n \in \mathbb{N}$. In particular, the range of any function $A_t(\cdot)$ given by (1.6) is a subset of $[0,1]$ due to the restriction $0 \le a_{t,j} \le 1$. Third, the dimension of $\mathscr{A}^{(n)}$ is given by $\dim[\mathscr{A}^{(n)}] = \sum_{t=0}^{K_n-1} \lceil 2^{J_t} + 1 \rceil$, and as can be noted, is extremely large because $\mathscr{A}$

is a set of sequence of functions whose components are not necessarily continuous functions. Fourth, due to the conditions 1-3, any polynomial $P \in \mathscr{H}^{(n)}$ satisfies the restrictions $P(0) = 0$, $0 \le P'(\cdot) \le 1$, and $\|\nabla^{R+1} P\|_{\infty,[0,\bar{u}]} \le \bar{H}$. Basically, the equality $P(0) = 0$ follows immediately from $b_0 = 0$. The inequality $0 \le P'(\cdot) \le 1$ is obtained from the formulas

$$P'(u) = \frac{L_n}{\bar{u}} \sum_{j=0}^{L_n} b_j [p_{L_n-1,j-1}(u) - p_{L_n-1,j}(u)] = \frac{L_n}{\bar{u}} \sum_{j=0}^{L_n-1} (b_{j+1} - b_j) p_{L_n-1,j}(u)$$

and $\sum_{j=0}^{L_n-1} p_{L_n-1,j}(u) = 1$; see [DL93], Chapter 10. Proceeding in a similar manner and exploiting the third condition, it can also be shown that $\|\nabla^{R+1} P\|_{\infty,[0,\bar{u}]} \le \bar{H}$. Finally, the dimension of $\mathscr{H}^{(n)}$ is $\dim[\mathscr{H}^{(n)}] = L_n$ because $b_0 = 0$, and trivially, $\dim[\Theta^{(n)}] = \dim[\mathscr{A}^{(n)}] + \dim[\mathscr{H}^{(n)}]$.

The main result of this section is that the proposed sieve spaces satisfy the following approximation property.

Proposition 1.3.2. *Let $\gamma_J > 0$ be a finite constant. If $[G_1, G_2] \in \mathcal{G}_R^*$, the following results hold:*

1. *If J_t is of the form $J_t = \gamma_J (t+5) \log_2(\lceil K_n^{1/2} \rceil)$, $t = 0, 1, \ldots, K_n - 1$, then there exists a sequence $\{(A_t^{(n)})_{t \in \mathbb{N}_0} \in \mathscr{A}^{(n)} : n \in \mathbb{N}\}$ such that $Q_1[(A_t^{(n)})_t] = O(K_n^{-1})$ as $n \to +\infty$.*

2. *There exists a sequence of polynomials $\{P^{(n)} \in \mathscr{H}^{(n)} : n \in \mathbb{N}\}$ such that $\|Q_2[\cdot, P^{(n)}]\|_{\infty,\mathscr{A}^{(n)}} = O(K_n/L_n)$ as $n \to +\infty$.*

The proof of this proposition is constructive, in the sense that it provides explicit expressions for $(A_t^{(n)})_t$ and $P^{(n)}$ in terms of $R_1(\cdot)$, $R_2(\cdot)$, and $\lambda_0^{-1}(\cdot)$. All details are given in Appendix 1.A.1.2. The proof of the first item is very involved, but the main idea can be summarized as follows. For a fixed $u \in [0, \bar{r}]$, recall from expression (1.4) that $\lambda_0^{-1}(u) = \sum_{t \in \mathbb{N}_0} \Delta b(\tilde{\alpha}_t)$, where $R_1(\tilde{\alpha}_0) = u$ and $R_1(\tilde{\alpha}_t) = R_2(\tilde{\alpha}_{t-1})$ when $t \ge 1$. Instead of considering a fixed $u \in [0, \bar{r}]$, I regard $(\tilde{\alpha}_t)_t$ as a sequence of functions $\{\tilde{\alpha}_t(\cdot) : t \in \mathbb{N}_0\}$ defined on $[0, \bar{r}]$. More specifically, I consider $(\tilde{\alpha}_t)_t$ as a sequence of functions $\{\tilde{\alpha}_t(\cdot) : t \in \mathbb{N}_0\}$ with domain $[0, \bar{r}]$ that satisfies the following recursive relation: $R_1[\tilde{\alpha}_0(u)] = u$ and $R_1[\tilde{\alpha}_t(u)] = R_2[\tilde{\alpha}_{t-1}(u)]$ for all $t \geq 1$ and $u \in [0, \bar{r}]$. By construction, it follows that $Q_1[(\tilde{\alpha}_t)_t] = 0$, so now the objective is to approximate $(\tilde{\alpha}_t)_t$ using elements of $\mathscr{A}^{(n)}$.

Such an approximation has two major difficulties. The first is that $(\tilde{\alpha}_t)_t$ is not uniquely defined because $R_1(\cdot)$ is not necessarily strictly increasing. The second is that $(\tilde{\alpha}_t)_t$ does not belong to $\mathscr{A}$ because $\tilde{\alpha}_t(u) > 0$ for all $(t, u) \in \mathbb{N}_0 \times (0, \bar{r}]$; note that $R_1(\alpha) = R_2(\alpha) = 0$ if and only if $\alpha = 0$. To address both difficulties, for each $t = 0, 1, \ldots, K_n - 1$, I propose a function $A_t^{(n)}(\cdot)$ that converges uniformly to a well-defined $\tilde{\alpha}_t(\cdot)$, while for each $t \geq K_n$, I simply set $A_t^{(n)}(\cdot) = 0$. To complete the proof, I show that the resulting sequence, $\{A_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$, belongs to $\mathscr{A}^{(n)}$ and approaches $\tilde{\alpha}_t(\cdot)$ so that $Q_1[(A_t^{(n)})_t] = O(K_n^{-1})$.

Turning to the second item of Proposition 1.3.2, note that $\|Q_2[\cdot, \lambda_0^{-1}]\|_{\infty, \mathscr{A}^{(n)}} = 0$, so the idea is to approximate $\lambda_0^{-1}(\cdot)$ by Bernstein polynomials. The natural candidate is the Bernstein operator of degree L_n, namely,

$$P^{(n)}(u) = \sum_{j=0}^{L_n} b_j^{(n)} p_{L_n, j}(u), \tag{1.7}$$

where $b_j^{(n)} = \lambda_0^{-1}(j\bar{u}/L_n)$ for $j = 0, 1, \ldots, L_n$. Although the function $\lambda_0^{-1}(\cdot)$ cannot be identified on its entire domain $\mathbb{R}_{\geq 0}$, the second condition of Definition 4 ensures that $\lambda_0^{-1}(\cdot)$ exists outside $[0, \bar{r}]$, so the coefficients $\{b_j^{(n)} : j = 0, 1, \ldots, L_n\}$ are well-defined. Regarding the approximation rate of (1.7), it is well-known that $\|P^{(n)} - \lambda_0^{-1}\|_{\infty, [0, \bar{r}]} = O(L_n^{-1})$; see [DL93], Chapter 10, Theorem 3.1. To complete the proof, I show that $P^{(n)} \in \mathscr{H}^{(n)}$ when n is sufficiently large, and also, that $Q_2(\cdot, \cdot)$ can be bounded above as follows:

$$Q_2[(A_t)_t, P^{(n)}] \leq \bar{C} K_n \|P^{(n)} - \lambda_0^{-1}\|_{\infty, [0, \bar{r}]}$$

for all $(A_t)_t \in \mathscr{A}^{(n)}$, where $\bar{C} > 0$ is a finite constant independent of $(A_t)_t$ and n. Finally, the desired result emerges from $\|P^{(n)} - \lambda_0^{-1}\|_{\infty, [0, \bar{r}]} = O(L_n^{-1})$.

1.4 Estimation Method: Uniform Consistency

The objective of this paper is to estimate the bidders' utility function and the density of private values. A valid estimator for $\lambda_0^{-1}(\cdot)$ is sufficient for estimating these functions. To that end, in this section, I construct a valid estimator for $\lambda_0^{-1}(\cdot)$ from observed bids. Roughly speaking, such an estimator is the empirical counterpart of (1.7).

This section is divided into four subsections. Subsection 1.4.1 presents the assumption regarding the data generating process. Subsections 1.4.2 and 1.4.3 build the estimator of $\lambda_0^{-1}(\cdot)$, which is denoted by $\hat{\lambda}^{-1}(\cdot)$. Exploiting the mathematical results of the previous section, subsection 1.4.4 derives the main result of the paper: the weak uniform consistency of $\hat{\lambda}^{-1}(\cdot)$ together with its rate of convergence.

1.4.1 Data Generating Process

In practice, the auctioned object can be heterogeneous, so I introduce an additional random vector X to account for heterogeneity in the auctioned object. The set of numbers of potential bidders $\mathcal{I}$ may contain more than two elements, so I also consider this case. The econometrician observes a collection of random vectors $\{(B_{pl}, I_l, X_l) : p = 1, \ldots, I_l; l = 1, \ldots, N\}$, where B_{pl} is the bid placed by the pth individual in the lth auction, I_l is the number of bidders in the lth auction, and X_l is a vector of continuous auction-specific covariates. The following assumption is satisfied.

Assumption 1. *There exists a collection of independent random vectors* $\{(B_{1,l}, \ldots, B_{I_l,l}, I_l, X_l) : l = 1, 2, \ldots, N\}$ *defined on a probability space* $(\Sigma, \mathscr{F}, \mathbb{P})$ *and the following conditions hold:*

1. $\{(I_l, X_l) : l = 1, 2, \ldots, N\}$ *are identically distributed.*

2. *The marginal p.d.f.* $\varphi(\cdot)$ *of* X_l *has compact support* $\mathcal{X} \subseteq \mathbb{R}^d$*, is bounded away from zero on* $\mathcal{X}$*, admits* $R+1$ *continuous derivatives on* $\text{int}(\mathcal{X})$*, and* $2R+1 \geq d$.

3. *For each* $x \in \mathcal{X}$*, the conditional p.d.f.* $\pi(\cdot|x)$ *of* I_l *given* $X_l = x$ *has finite support* $\mathcal{I} = \{i_1, i_2, \ldots, i_M\} \subseteq \mathbb{N}_{\geq 2}$ *with* $M \geq 2$*. For all* $i \in \mathcal{I}$*,* $\pi(i|\cdot)$ *is strictly positive and admits* $R+1$ *continuous derivatives on* $\text{int}(\mathcal{X})$.

4. *For each* $(i,x) \in \mathcal{I} \times \mathcal{X}$*,* $\{(B_{1,l}, \ldots, B_{I_l,l}) : l = 1,2,\ldots,N\}$ *are identically distributed conditional on* $\{(I_l, X_l) = (i,x) : l = 1,2,\ldots,N\}$ *with joint c.d.f.* $\mathbf{G}(\cdot|i,x)$*. For all* $(i_1,i_2,x) \in \mathcal{I}^2 \times \mathcal{X}$ *with* $i_1 < i_2$*,* $\mathbf{G}(b|i_j,x) = \prod_{h=1}^{i_j} G(b_h|i_j,x)$ *where* $j = 1,2$*,* $b \in \mathbb{R}^{i_j}$*, and* $[G(\cdot|i_1,x), G(\cdot|i_2,x)] \in \mathcal{G}_R^*$.

This assumption formalizes the idea that the observations have been generated by the auction model of subsection 1.2.1. It also imposes standard regularity conditions that will be used later to establish the asymptotic results; among others, I highlight that the auctions must be independent.

In the rest of the paper, the asymptotic properties of the estimators are detailed as the sample size N grows to infinity, whereas M (the cardinality of $\mathcal{I}$) is fixed. In other words, all limits are taken as $N \to +\infty$ keeping M constant. In addition, I consider the case in which $X_l = x$, where x is fixed and belongs to $\text{int}(\mathcal{X})$. If there is no ambiguity, the dependence of x is omitted from the notation; for example, the α-quantile of $G(\cdot|i_j,x)$ is simply denoted by $b_j(\alpha)$, where $j = 1,\ldots,M$. The abbreviation"w.p.a.1" stands for "with $\mathbb{P}$-probability approaching one."

1.4.2 Preliminary Kernel Estimators

This subsection constructs nonparametric estimators for $\Delta b(\cdot)$ and $R_j(\cdot)$. These estimators will be used later, in subsection 1.4.3, to compute the empirical criterion functions associated with (1.5). Following recent literature on the subject, we employ a kernel approach to estimate the components of $\Delta b(\cdot)$ and $R_j(\cdot)$, namely, $b_j(\cdot)$ and $b'_j(\cdot)$. For this purpose, let $k(\cdot)$ and $k_b(\cdot)$ be univariate kernels, and also, let h_G, h_g, h_∂, and h_b be bandwidths. The following assumptions are satisfied.

Assumption 2. *The kernels $k(\cdot)$ and $k_b(\cdot)$ are symmetric, have support $[-1,1]$, have $R+1$ continuous derivatives on $\mathbb{R}$, and satisfy $\int k(v)dv = \int k_b(v)dv = 1$, as well as, $k_b(\cdot) \geq 0$. The order of $k(\cdot)$ is $R+1$, that is, moments of order strictly smaller than the given order vanish. For $x \in \mathbb{R}^d$, denote the product kernel by $K(x) = \prod_{j=1}^d k(x_j)$.*

Assumption 3. *Let γ_G, γ_g, and γ_∂ be positive constants. The bandwidths are of the form:*

$$h_G = \gamma_G[\log(N)/N]^{1/(2R+d+2)}, h_g = \gamma_g[\log(N)/N]^{1/(2R+d+3)},$$

and $h_\partial = \gamma_\partial(1/N)^{1/(d+1)}$ if $d > 0$. Moreover, h_b satisfies $h_b \to 0$ and $h_b/h_g \to +\infty$ as $N \to +\infty$.

The estimators of $\varphi(x)$, $\pi(i|x)$, $G(b|i,x)$, and the density $g(b|i,x)$ are obtained directly from eqs. (8)-(9) in [MS12]. For any $(b,i,x) \in \mathbb{R}_+ \times \mathcal{I} \times \text{int}(\mathcal{X})$, I define

$$\begin{aligned}
\hat{\varphi}(x) &= \frac{1}{Nh_G^d}\sum_{l=1}^N K\left(\frac{x-X_l}{h_G}\right), \\
\hat{\pi}(i|x) &= \frac{1}{\hat{\varphi}(x)Nh_G^d}\sum_{l=1}^N \mathbb{1}\{I_l = i\}K\left(\frac{x-X_l}{h_G}\right), \\
\hat{G}(b|i,x) &= \frac{1}{\hat{\pi}(i|x)\hat{\varphi}(x)Nh_G^d i}\sum_{l=1}^N\sum_{p=1}^{I_l} \mathbb{1}\{B_{lp} \leq b, I_l = i\}K\left(\frac{x-X_l}{h_G}\right), \text{ and} \\
\hat{g}(b|i,x) &= \frac{1}{\hat{\pi}(i|x)\hat{\varphi}(x)Nh_g^{d+1} i}\sum_{l=1}^N\sum_{p=1}^{I_l} \mathbb{1}\{I_l = i\}k\left(\frac{b-B_{pl}}{h_g}\right)K\left(\frac{x-X_l}{h_g}\right).
\end{aligned}$$

To estimate the boundaries of the support of $g(\cdot|i_j,x)$, where $j = 1,\ldots,M$, I define the following hypercube containing $x = (x_1,\ldots,x_d)$:

$$\pi(x) = [x_1 - h_\partial, x_1 + h_\partial] \times \cdots \times [x_d - h_\partial, x_d + h_\partial].$$

Then, the boundaries of the support of $g(\cdot|i_j,x)$, $\underline{b}$ and $\bar{b}_j$, are estimated by

$$\begin{aligned}\hat{\underline{b}} &= \min\{B_{pl} : p = 1,\dots,i_l; l = 1,\dots,N; X_l \in \pi(x)\} \text{ and} \\ \hat{\bar{b}}_j &= \max\{B_{pl} : p = 1,\dots,i_l; l = 1,\dots,N; I_l = i_j; X_l \in \pi(x)\},\end{aligned}$$

respectively. Recall that $\underline{b}$ is independent of i_j by Assumption 1.4, thus there is no need to restrict bids such that $I_l = i_j$.

The conditional quantile function $b_j(\cdot)$, as well as its derivative $b'_j(\cdot)$, is estimated as in [MS12]. However, I give special attention to boundary issues, so the estimator of $b_j(\cdot)$ is defined by

$$\hat{b}_j(\alpha) = \begin{cases} \hat{\underline{b}} & \text{if } 0 \le \alpha < h_b, \\ \inf\{b \in \mathbb{R}_{\ge 0} : \hat{G}(b|i_j,x) \ge \alpha\} & \text{if } h_b \le \alpha \le 1 - h_b, \\ \hat{\bar{b}}_j & \text{otherwise.} \end{cases}$$

Due to the boundary correction on $[0, h_b) \cup (1 - h_b, 1]$ and since $h_b/h_g \to +\infty$ (Assumption 3), Lemma 1.A.9 (Appendix 1.A.2) shows that $\hat{g}[\hat{b}_j(\cdot)|i_j,x]$ is bounded away from zero on $[h_b, 1-h_b]$ w.p.a.1. Hence, the estimator of $b'_j(\cdot)$ can be defined by

$$\hat{b}'_j(\alpha) = \begin{cases} 1/\hat{\underline{g}}_j & \text{if } 0 \le \alpha < h_b, \\ 1/\hat{g}[\hat{b}_j(\alpha)|i_j,x] & \text{if } h_b \le \alpha \le 1 - h_b, \\ 1/\hat{\bar{g}}_j & \text{otherwise,} \end{cases}$$

where $\hat{\underline{g}}_j$ and $\hat{\bar{g}}_j$ are consistent estimators of $g(\underline{b}|i_j,x)$ and $g(\bar{b}_j|i_j,x)$, respectively:

$$\begin{aligned}\hat{\underline{g}}_j &= \frac{1}{\hat{\pi}(i_j|x)\hat{\varphi}(x)Nh_g^{d+1}i_j}\sum_{l=1}^{N}\sum_{p=1}^{I_l}\mathbb{1}\{I_l = i_j\}\tilde{k}_b\left(\frac{\hat{\underline{b}} - B_{pl}}{h_g}\right)K\left(\frac{x - X_l}{h_g}\right) \text{ and} \\ \hat{\bar{g}}_j &= \frac{1}{\hat{\pi}(i_j|x)\hat{\varphi}(x)Nh_g^{d+1}i_j}\sum_{l=1}^{N}\sum_{p=1}^{I_l}\mathbb{1}\{I_l = i_j\}\tilde{k}_b\left(\frac{B_{pl} - \hat{\bar{b}}_j}{h_g}\right)K\left(\frac{x - X_l}{h_g}\right),\end{aligned}$$

where $\tilde{k}_b(\cdot)$ is a one-sided version of $k_b(\cdot)$, namely, $\tilde{k}_b(v) = 2k_b(v)\mathbb{1}(v \le 0)$. I remark that below results do not change if $\hat{\underline{g}}_j$ and $\hat{\bar{g}}_j$ are replaced by other estimators whose rates of convergence are equal or faster than h_b.[3]

[3] For a recent discussion about estimation of conditional quantile functions, see [GS12] and the references cited therein.

For the rest of this paper, I consider a fixed pair $(i_1, i_2) \in \mathcal{I}^2$ with $i_1 < i_2$. Definition 4 imposes sign and smoothness conditions on the difference $\Delta b(\alpha) = b_2(\alpha) - b_1(\alpha)$. In order to preserve those properties, $\Delta b(\alpha)$ is estimated by the following convolution operation:

$$\Delta\hat{b}(\alpha) = \frac{1}{h_b \bar{k}_b(\alpha)} \int_0^1 \max\{0, \hat{b}_2(v) - \hat{b}_1(v)\} k_b\left(\frac{\alpha - v}{h_b}\right) dv, \tag{1.8}$$

where $\alpha \in [0,1]$ and $\bar{k}_b(\alpha) = \int_{-1}^{(1-\alpha)/h_b} k_b(v)dv$. Similarly, $R_1(\alpha)$ is estimated by

$$\hat{R}_1(\alpha) = \frac{1}{h_b \bar{k}_b(\alpha)} \int_0^1 \frac{v\hat{b}_1'(v)}{(i_1 - 1)} k_b\left(\frac{\alpha - v}{h_b}\right) dv, \tag{1.9}$$

and then, the estimator of $\bar{r}$ becomes $\hat{r} = \max\{\hat{R}_1(\alpha) : \alpha \in [0,1]\}$. In order to preserve the inequality $R_2(\cdot) \leq R_1(\cdot)$, the estimator of $R_2(\alpha)$ is given by

$$\hat{R}_2(\alpha) = \frac{1}{h_b \bar{k}_b(\alpha)} \int_0^1 \min\left\{\frac{v\hat{b}_1'(v)}{(i_1 - 1)}, \frac{v\hat{b}_2'(v)}{(i_2 - 1)}\right\} k_b\left(\frac{\alpha - v}{h_b}\right) dv. \tag{1.10}$$

Observe that I have not imposed a boundary correction on $[0, h_b)$ because $\Delta b(\alpha)$ and $R_j(\alpha)$ converge zero as $\alpha \to 0^+$. In addition, such implementation would not improve the uniform rate of convergence of $\Delta\hat{b}(\cdot)$ and $\hat{R}_j(\cdot)$. Note also that these convolution operators maintain the degree of smoothness of $\Delta b(\cdot)$ and $R_j(\cdot)$, in particular, the continuity of $\Delta\hat{b}(\cdot)$ and $\hat{R}_j(\cdot)$ guarantees the existence of solutions for the minimization problems discussed in the next subsection.

Finally, the uniform rates of convergence of $\Delta\hat{b}(\cdot)$ and $\hat{R}_j(\cdot)$ are established by the next lemma. The asymptotic properties of the components of $\Delta\hat{b}(\cdot)$ and $\hat{R}_j(\cdot)$ are detailed in Appendix 1.A.2, Lemma 1.A.9.

Lemma 1.4.1. *Under Assumptions 1-3,* $\|\Delta\hat{b}(\cdot) - \Delta b(\cdot)\|_{\infty,[0,1]} = O_P(h_b)$ *and* $\|\hat{R}_j(\cdot) - R_j(\cdot)\|_{\infty,[0,1]} = O_P(h_b)$ *for* $j = 1, 2$. *As a consequence,* $|\hat{r} - \bar{r}| = O_P(h_b)$.

1.4.3 Two-Step Nonparametric Sieve Estimator

In this subsection, I formally define the estimator of $\lambda_0^{-1}(\cdot)$ and describe how to implement it. As a starting point, I construct the empirical criterion functions

$\hat{Q}_1(\cdot)$ and $\hat{Q}_2(\cdot,\cdot)$ associated with $Q_1(\cdot)$ and $Q_2(\cdot,\cdot)$, respectively. Recall that $(K_n)_n$ and $(L_n)_n$ are increasing divergent sequences of positive integers, which are related to the dimensions of the sieve spaces, and that N denotes the number of auctions in the data set.

A natural way to proceed is to combine expression (1.5) with the estimators (1.8)-(1.10). Then, the empirical criterion function $\hat{Q}_1 : \mathscr{A}^{(N)} \to \mathbb{R}_{\geq 0}$ is defined by

$$\begin{aligned}\hat{Q}_1[(A_t)_t] &= \int_0^{\hat{r}} |\hat{R}_1[A_0(u)] - u| du + \sum_{t=1}^{K_N-1} \int_0^{\hat{r}} |\hat{R}_1[A_t(u)] - \hat{R}_2[A_{t-1}(u)]| du \\ &\quad + \int_0^{\hat{r}} |\hat{R}_2[A_{K_N-1}(u)]| du,\end{aligned}$$

while $\hat{Q}_2 : \Theta^{(N)} \to \mathbb{R}_{\geq 0}$ becomes

$$\hat{Q}_2[(A_t)_t, P] = \sum_{t=0}^{K_N-1} \int_0^{\hat{r}} |\Delta\hat{b}[A_t(u)] + P\{\hat{R}_2[A_t(u)]\} - P\{\hat{R}_1[A_t(u)]\}| du.$$

Note that the dimension of the sieve spaces depends on the sample size N. To facilitate technical details, $\bar{u}$ is taken to be large enough so that $\bar{u} > \bar{r}$, which implies $\bar{u} > \hat{r}$ w.p.a.1.

The estimator of $\lambda_0^{-1}(\cdot)$ is computed in two steps. In the first, we set $(\hat{A}_t)_{t\in\mathbb{N}_0}$ as the argument that minimizes $\hat{Q}_1(\cdot)$ over $\mathscr{A}^{(N)}$, formally,

$$(\hat{A}_t)_t = \arg\min \left\{ \hat{Q}_1[(A_t)_t] : (A_t)_t \in \mathscr{A}^{(N)} \right\}.$$

By definition of $\mathscr{A}^{(N)}$ and since the integrals of $\hat{Q}_1(\cdot)$ are supported on $[0, \hat{r}]$, the solution of this minimization problem is characterized by a sequence of functions $(\hat{A}_t)_{t\in\mathbb{N}_0} : [0, \hat{r}] \to [0,1]^\infty$ of the form

$$\hat{A}_t(u) = \begin{cases} \sum_{j=0}^{S_t} \hat{a}_{t,j} \kappa[2^{J_t}(u/\bar{u}) - j] & \text{if } 0 \leq t \leq K_N - 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1.11}$$

where $S_t = \lceil \hat{r} 2^{J_t} / \bar{u} \rceil$. Note that $\kappa[2^{J_t}(u/\bar{u}) - j] = 0$ when $u \in [0, \hat{r}]$ and $j > \lceil \hat{r} 2^{J_t}/\bar{u} \rceil$; for this reason, $\{0, 1, \ldots, S_t\}$ has become the support of j in the sum of expression (1.11) when $0 \leq t \leq K_N - 1$. The coefficients associated with $(\hat{A}_t)_t$,

$(\hat{a}_{0,0}, \hat{a}_{0,1}, \ldots, \hat{a}_{0,S_0}, \hat{a}_{1,0}, \ldots, \hat{a}_{K_N-1,S_{K_N-1}})$, are the argument that minimizes the expression

$$\int_0^{\hat{r}} \left| \hat{R}_1 \left(\sum_{j=0}^{S_0} a_{0,j} \kappa[2^{J_0}(u/\bar{u}) - j] \right) - u \right| du$$
$$+ \sum_{t=1}^{K_N-1} \int_0^{\hat{r}} \left| \hat{R}_1 \left(\sum_{j=0}^{S_t} a_{t,j} \kappa[2^{J_t}(u/\bar{u}) - j] \right) - \hat{R}_2 \left(\sum_{j=0}^{S_{t-1}} a_{t-1,j} \kappa[2^{J_{t-1}}(u/\bar{u}) - j] \right) \right| du$$
$$+ \int_0^{\hat{r}} \left| \hat{R}_2 \left(\sum_{j=0}^{S_{K_N-1}} a_{t,j} \kappa[2^{J_{K_N-1}}(u/\bar{u}) - j] \right) \right| du$$

with respect to $(a_{0,0}, \ldots, a_{0,S_0}, a_{1,0}, \ldots, a_{K_N-1,S_{K_N-1}}) \in [0,1]^{\dim[\mathscr{A}^{(N)}]}$. To conclude the first step, I remark that the above minimization problem has a well-defined solution because both $\Delta\hat{b}(\cdot)$ and $\hat{R}_j(\cdot)$ are uniformly continuous.

In the second step, we define the estimator of $\lambda_0^{-1}(\cdot)$ as the argument that minimizes $\hat{Q}_2[(\hat{A}_t)_t, \cdot]$ over $\mathscr{H}^{(N)}$, formally,

$$\hat{\lambda}^{-1}(\cdot) = \arg\min \left\{ \hat{Q}_2[(\hat{A}_t)_t, P] : P \in \mathscr{H}^{(N)} \right\}.$$

The solution of this minimization problem is a Bernstein polynomial with degree L_N of the form

$$\hat{\lambda}^{-1}(u) = \sum_{j=1}^{L_N} \hat{b}_j p_{L_N,j}(u),$$

where $(\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{L_N})$ is the argument that minimizes the expression

$$\sum_{t=0}^{K_N-1} \int_0^{\hat{r}} \left| \Delta\hat{b}[\hat{A}_t(u)] + \sum_{j=1}^{L_N} b_j \left[p_{L_N,j}\{\hat{R}_2[\hat{A}_t(u)]\} - p_{L_N,j}\{\hat{R}_1[\hat{A}_t(u)]\} \right] \right| du$$

with respect to $(b_1, b_2, \ldots, b_{L_N})$ and subject to conditions 1-3 of Definition 6. Clearly, this minimization problem has a well-defined solution as the objective function is continuous in $(b_1, b_2, \ldots, b_{L_N})$.

1.4.4 Uniform Consistency

In this subsection, I establish the main result of the paper: the weak uniform consistency of $\hat{\lambda}^{-1}(\cdot)$ with its rate of convergence. In order to achieve the fastest rate, I make the assumption that K_N diverges at the rate of $h_b^{-1/2}$. In addition, I set the dimensions of the sieve spaces as an increasing function of the sample size N. More specifically, the assumption is as follows.

Assumption 4. *Let γ_K, γ_J, and γ_L be positive constants. The sequences K_N and L_N, as well as J_t ($t = 0, 1, \ldots, K_N - 1$), are of the form: $K_N = \lceil \gamma_K h_b^{-1/2} \rceil$, $L_N = \lceil \gamma_L K_N^2 \rceil$, and $J_t = \gamma_J (t+5) \log_2(\lceil K_N^{1/2} \rceil)$.*

Recall that h_b is a bandwidth that satisfies Assumption 3, K_N indicates to the support of the sums of the empirical criterion functions $\hat{Q}_1(\cdot)$ and $\hat{Q}_2(\cdot,\cdot)$, while L_N indicates the degree of the Bernstein polynomial associated with $\hat{\lambda}^{-1}(\cdot)$. Since K_N and L_N must be positive integers, Assumption 4 uses the ceiling function $\lceil \cdot \rceil$ in the construction of these sequences.

The next lemma states the uniform rate of convergence in probability of $\hat{Q}_1(\cdot)$ and $\hat{Q}_2(\cdot,\cdot)$ over the corresponding sieve spaces.

Lemma 1.4.2. *Under Assumptions 1-4 and when $N \to +\infty$, $\|\hat{Q}_1(\cdot) - Q_1(\cdot)\|_{\infty,\mathscr{A}^{(N)}} = O_P(h_b^{1/2})$ and $\|\hat{Q}_2(\cdot,\cdot) - Q_2(\cdot,\cdot)\|_{\infty,\Theta^{(N)}} = O_P(h_b^{1/2})$.*

The proof of this lemma is simple, basically, it is an immediate consequence of Lemma 1.4.1 and Assumption 4; see Appendix 1.A.2.2. The next theorem presents the main finding of this paper: the uniform rate of convergence in probability of $\hat{\lambda}^{-1}(\cdot)$ (and its derivatives) over the identification region $[0,\bar{r}]$.

Theorem 1.4.1. *Under Assumptions 1-4 and for $r = 0, 1, \ldots, R$,*

$$\left\| \nabla^r \hat{\lambda}^{-1}(\cdot) - \nabla^r \lambda_0^{-1}(\cdot) \right\|_{\infty,[0,\bar{r}]} = O_P\left(h_b^{1/[2c(r)]} \right),$$

where $c(r)$ and h_b have been defined in subsections 1.3.2 and 1.4.2, respectively.

Given Propositions 1.3.1-1.3.2 and Lemma 1.4.2, the proof of this Theorem relies on simple arguments. Here, we present a sketch for the case $r = 0$ and all the details are relegated to Appendix 1.A.2.3. From Proposition 1.3.1 and Lemma 1.4.2, it follows that $\underline{K}\|\hat{\lambda}^{-1} - \lambda_0^{-1}\|_{\infty,[0,\bar{r}]}^{c(0)} \leq Q_1[(\hat{A}_t)_t] + Q_2[(\hat{A}_t)_t, \hat{\lambda}^{-1}] \approx \hat{Q}_1[(\hat{A}_t)_t] + \hat{Q}_2[(\hat{A}_t)_t, \hat{\lambda}^{-1}]$. By construction of $[(\hat{A}_t)_t, \hat{\lambda}^{-1}]$, $\hat{Q}_1[(\hat{A}_t)_t] + \hat{Q}_2[(\hat{A}_t)_t, \hat{\lambda}^{-1}]$ is bounded above by $\hat{Q}_1[(A_t^{(N)})_t] + \hat{Q}_2[(\hat{A}_t)_t, P^{(N)}] \approx Q_1[(A_t^{(N)})_t] + Q_2[(\hat{A}_t)_t, P^{(N)}]$. By definition of sup-norm $\|\cdot\|_{\infty,\mathscr{A}^{(N)}}$ and since $(\hat{A}_t)_t \in \mathscr{A}^{(N)}$, the right-hand side can be bounded above by $Q_1[(A_t^{(N)})_t] + \|Q_2[\cdot, P^{(N)}]\|_{\infty,\mathscr{A}^{(N)}}$, so we obtain $\underline{K}\|\hat{\lambda}^{-1} - \lambda_0^{-1}\|_{\infty,[0,\bar{r}]}^{c(0)} \leq Q_1[(A_t^{(N)})_t] + \|Q_2[\cdot, P^{(N)}]\|_{\infty,\mathscr{A}^{(N)}}$. Then, the desired result emerges from Proposition 1.3.2 and Assumption 4, which implies $K_N^{-1} \leq \gamma_K^{-1} h_b^{1/2}$. I remark that the validity of the two-step procedure, described in subsection 1.4.3, relies on the uniform convergence of $Q_2[\cdot, P^{(N)}]$ over $\mathscr{A}^{(N)}$.

An immediate corollary of Theorem 1.4.1 is that $\hat{\lambda}^{-1}(\cdot)$ converges uniformly to $\lambda_0^{-1}(\cdot)$ at the rate of $[\log(N)/N]^{\frac{\nu(R+2)}{2(R+1)(2R+d+3)}}$ for any fixed value of $\nu \in (0,1)$. Not surprisingly, this rate is slower than the optimal semiparametric rate, $N^{-(R+1)/(2R+3)}$, obtained by [CGP11].

1.5 Estimating the First-Price Auction Model

The previous section developed an estimator for $\lambda_0^{-1}(\cdot)$; this section applies it to the auction model of subsection 1.2.1. Exploiting the asymptotic property of $\hat{\lambda}^{-1}(\cdot)$, subsection 1.5.1 proposes an estimator for the bidders' utility function $U(\cdot)$, while subsection 1.5.2 suggests a simple procedure to recover the density of private values $f(\cdot)$. The distinguishing feature of the proposed estimators is that they do not place a parametric restriction on the shape of $U(\cdot)$.

1.5.1 Bidders' Utility Function: Uniform Consistency

In this subsection, I propose an estimator for the bidders' utility function $U(\cdot)$; then, I show that it is uniformly consistent and provide its rate of convergence. As far as I know, this is the first nonparametric estimator for the bidders' utility function in the context of a first-price sealed-bid auction model.

As $\lambda_0^{-1}(\cdot)$ is identified on $[0,\bar{r}]$, it follows immediately that $\lambda_0(\cdot)$ is identified on $[0,\lambda_0^{-1}(\bar{r})]$. To facilitate the subsequent analysis, and since the scale of an utility function is irrelevant, it is assumed that $U[\lambda_0^{-1}(\bar{r})]=1$. From this normalization, $U(\cdot)$ can be identified on $[0,\lambda_0^{-1}(\bar{r})]$ as the solution of the differential equation $\lambda_0(\cdot)U'(\cdot)-U(\cdot)=0$; more specifically, $U(y)=\exp\{-\int_y^{\lambda_0^{-1}(\bar{r})}[1/\lambda(t)]dt\}$ where $y\in[0,\lambda_0^{-1}(\bar{r})]$. It can also be shown that $\lambda_0^{-1}(\bar{r})=\max\{v-s(v;I_1):v\in[\underline{v},\bar{v}]\}$, where $v-s(v;I_1)$ represents the monetary gain. Then, the identification region $[0,\lambda_0^{-1}(\bar{r})]$ cannot be improved because bidders cannot obtain a monetary gain greater than $\lambda_0^{-1}(\bar{r})$.

In order to recover $U(\cdot)$ from the data, I propose to use the natural estimator:

$$\hat{U}(y) = \begin{cases} 0 & \text{if } y=0, \\ \exp\left\{-\int_y^{\hat{\lambda}^{-1}(\hat{r})}[1/\hat{\lambda}(t)]dt\right\} & \text{if } 0<y<\hat{\lambda}^{-1}(\hat{r}), \\ 1 & \text{otherwise;} \end{cases} \tag{1.12}$$

where $\hat{\lambda}(\cdot)$ is an estimator of $\lambda_0(\cdot)$ defined as the inverse of $\hat{\lambda}^{-1}:[0,\hat{r}]\to[0,\hat{\lambda}^{-1}(\hat{r})]$, namely,

$$\hat{\lambda}(y) = \begin{cases} (\hat{\lambda}^{-1})^{-1}(y) & \text{if } 0\le y<\hat{\lambda}^{-1}(\hat{r}), \\ \hat{r} & \text{otherwise.} \end{cases}$$

I remark that $\hat{\lambda}(\cdot)$ is well-defined and $\hat{\lambda}'(\cdot)\ge 1$ because $L_N^{-1}\le\nabla^1\hat{\lambda}^{-1}(\cdot)\le 1$. Since $\hat{\lambda}^{-1}(\cdot)$ is infinitely differentiable on $[0,\hat{r}]$ and preserves the shape of $\lambda_0^{-1}(\cdot)$, $\hat{U}(\cdot)$ also preserves the shape and smoothness properties of $U(\cdot)$. In particular, $\hat{U}(\cdot)$ is continuous on $\mathbb{R}_{\ge 0}$ because $\hat{\lambda}(\cdot)$ is continuous, $\hat{\lambda}(0)=0$, and $\int_y^{\hat{\lambda}^{-1}(\hat{r})}[1/\hat{\lambda}(t)]dt\to+\infty$ as $y\to 0^+$. Besides, $\hat{U}(\cdot)$ is strictly increasing and concave on $[0,\hat{r}]$ regardless of the sample size. The computation of $\hat{\lambda}(\cdot)$ is not involved: since $\hat{\lambda}^{-1}(\cdot)$ has a polynomial representation, $\hat{\lambda}(y)$ is equal to the root of $\hat{\lambda}^{-1}(\cdot)-y$ that belongs to $[0,\hat{r}]$. Such a root is unique because $\hat{\lambda}^{-1}(\cdot)$ is strictly increasing on $[0,\hat{r}]$.

To conclude this subsection, the next proposition states that both $\hat{\lambda}(\cdot)$ and $\hat{U}(\cdot)$ inherit the uniform rate of convergence in probability of $\hat{\lambda}^{-1}(\cdot)$. In addition, it establishes the rate of convergence of the estimators for $\nabla^r U(\cdot)$ with $r = 1, \dots, R+1$. This information is useful because, for instance, $\nabla^2 U(\cdot)$ and $\nabla^2 U(\cdot)/\nabla^1 U(\cdot)$ describe the bidders' risk preferences.

Proposition 1.5.1. *Under Assumptions 1-4,*

$$\|\hat{\lambda}(\cdot) - \lambda_0(\cdot)\|_{\infty,[0,\lambda_0^{-1}(\bar{r})]} = O_P\left(h_b^{1/[2c(0)]}\right) \quad \text{and}$$
$$\|\hat{U}(\cdot) - U(\cdot)\|_{\infty,[0,\lambda_0^{-1}(\bar{r})]} = O_P\left(h_b^{1/[2c(0)]}\right).$$

Moreover, $\|\nabla^r \hat{U}(\cdot) - \nabla^r U(\cdot)\|_{\infty,[0,\lambda_0^{-1}(\bar{r})]} = O_P(h_b^{1/[2c(r-1)]})$ *for* $r = 1, \dots, R+1$.

1.5.2 Density of Private Values: Asymptotic Normality

Following [MS12], I propose an estimator for the density of private values $f(\cdot)$. Then, I show that it is uniformly consistent and asymptotically normal with an appropriate choice of the bandwidth. The asymptotic normality is useful to facilitate inference and testing procedures.

Combining expression (1.2) with the equality $1/f(v) = v'[F(v)]$, it follows that

$$\frac{1}{f(v)} = b'[F(v)|I] + \nabla^1 \lambda_0^{-1}\left\{\frac{F(v)b'[F(v)|I]}{I-1}\right\}\frac{1}{(I-1)}\left\{b'[F(v)|I] - \frac{F(v)g'\{b[F(v)|I]|I\}}{[g\{b[F(v)|I]|I\}]^3}\right\},$$

where $v \in [\underline{v}, \bar{v}]$ and $[\underline{v}, \bar{v}]$ denotes the support of $f(\cdot)$. Note that $f(\cdot)$ is over-identified on $[\underline{v}, \bar{v}]$ because $M \geq 2$; see Assumption 1. In particular, if we replace $\lambda_0^{-1}(\cdot)$ by the identity function, then $\nabla^1 \lambda_0^{-1}(\cdot) = 1$, and we easily obtain equation (3) of [MS12].

To estimate $f(\cdot)$, first, I define a preliminary estimator for the α-quantile of $F(\cdot)$, $v(\alpha)$, by $\tilde{v}_j(\alpha) = \hat{b}_j(\alpha) + \hat{\lambda}^{-1}[\hat{R}_j(\alpha)]$ where $j = 1, \dots, M$ and $\alpha \in [0,1]$. If $j = 1$, $\hat{R}_j(\cdot)$ is obtained from formula (1.9), and if $j \geq 1$, we can use either formula

(1.9) or (1.10). Second, since $\tilde{v}_j(\cdot)$ is not necessarily increasing on $[0,1]$, I define a monotone version of $\tilde{v}_j(\cdot)$ as follows:

$$\hat{v}_j(\alpha) = \begin{cases} \inf\{\tilde{v}_j(t) : t \in [\alpha, 1/2]\} & \text{if } 0 \le \alpha < 1/2, \\ \sup\{\tilde{v}_j(t) : t \in [1/2, \alpha]\} & \text{if } 1/2 \le \alpha \le 1. \end{cases}$$

Third, the estimator for the distribution of private values $F(\cdot)$ becomes $\hat{F}_j(v) = \sup\{\alpha \in [0,1] : \hat{v}_j(\alpha) \le v\}$. Lastly, the density of private values $f(\cdot)$ is estimated by

$$\hat{f}_j(v) = \left[\hat{b}'_j[\hat{F}_j(v)] + \frac{\nabla^1\hat{\lambda}^{-1}\{\hat{R}_j[\hat{F}_j(v)]\}}{i_j - 1}\left\{\hat{b}'_j[\hat{F}_j(v)] - \frac{\hat{F}_j(v)\tilde{g}'\{\hat{b}_j[\hat{F}_j(v)]|i_j, x\}}{[\hat{g}\{\hat{b}_j[\hat{F}_j(v)]|i_j, x\}]^3}\right\}\right]^{-1}, \tag{1.13}$$

where $\tilde{g}'(\cdot|i_j, x)$ is simply the derivative of the kernel estimator $\hat{g}(\cdot|i_j, x)$, namely,

$$\tilde{g}'(b|i_j, x) = \frac{1}{\hat{\pi}(i_j|x)\hat{\varphi}(x)Nh_f^{d+2}i_j}\sum_{l=1}^{N}\sum_{p=1}^{I_l}\mathbb{1}\{I_l = i_j\}k'\left(\frac{b - B_{pl}}{h_f}\right)K\left(\frac{x - X_l}{h_f}\right).$$

In the above expression, when estimating $\tilde{g}'(\cdot|i_j, x)$, the bandwidth h_g of $\hat{g}(\cdot|i_j, x)$ has been replaced with another bandwidth h_f. The reason is that $\tilde{g}'(\cdot|i_j, x)$ is the main term in the asymptotically expansion of $\hat{f}_j(v)$. As discussed in Proposition 1.5.2.3 below, to obtain the asymptotic normality result, we require $\sqrt{Nh_f^{d+3}}h_b^{1/[2c(1)]} \to 0$, whereas h_g satisfies $\sqrt{Nh_g^{d+3}}h_b^{1/[2c(1)]} \to +\infty$ by Assumption 3. To avoid any confusion, I remark that h_g is always employed to compute $\hat{g}(\cdot|i_j, x)$.

The main results of this subsection are summarized in the next proposition. In order to obtain this result, the uniform consistency of $\hat{\lambda}^{-1}(\cdot)$ over the entire set $[0, \bar{r}]$ is crucial.

Proposition 1.5.2. *Under Assumptions 1-4, the following statements hold for $j = 1, \ldots, M$:*

1. $\|\hat{v}_j(\cdot) - v(\cdot)\|_{\infty,[0,1]} = O_P(h_b^{1/[2c(0)]})$ *and* $\|\hat{F}_j(\cdot) - F(\cdot)\|_{\infty,\mathbb{R}} = O_P(h_b^{1/[2c(0)]})$.

2. *Let $\mathcal{C}$ be a fixed closed inner subset of $(\underline{v}, \bar{v})$. If the bandwidth h_f satisfies $h_f \to 0$ and $Nh_f^{d+3}/\log(N) \to +\infty$, then $\|\hat{f}_j(\cdot) - f(\cdot)\|_{\infty,\mathcal{C}} = O_P(h_b^{1/[2c(1)]})$.*

3. *If $Nh_f^{d+3}h_b^{1/c(1)} \to 0$ and $Nh_f^{d+1} \to +\infty$, for any fixed $v \in (\underline{v}, \bar{v})$ we have that*

$$\sqrt{Nh_f^{d+3}}[\hat{f}_j(v) - f(v)] \xrightarrow{d} \mathcal{N}[0, V_f(v, i_j)],$$

where

$$V_f(v, i_j) = \left[\nabla^1 \lambda_0^{-1}\{R_j[F(v)]\}\right]^2 \frac{\{\int [k(u)]^2 du\}^d \{\int [k'(u)]^2 du\} [F(v)]^2 [f(v)]^4}{i_j(i_j - 1)^2 \pi(i_j|x)\varphi(x)[g\{b_j[F(v)]|i_j, x\}]^5}.$$

This asymptotic variance can be estimated by the plug-in estimator, that is,

$$\hat{V}_f(v, i_j) = \left[\nabla^1 \hat{\lambda}^{-1}\{\hat{R}_j[\hat{F}_j(v)]\}\right]^2 \frac{\{\int [k(u)]^2 du\}^d \{\int [k'(u)]^2 du\} [\hat{F}_j(v)]^2 [\hat{f}_j(v)]^4}{i_j(i_j - 1)^2 \hat{\pi}(i_j|x)\hat{\varphi}(x)[\hat{g}\{\hat{b}_j[\hat{F}_j(v)]|i_j, x\}]^5}.$$

This estimator satisfies $|\hat{V}_f(v, i_j) - V_f(v, i_j)| = O_P(h_b^{1/[2c(1)]})$.

This proposition extends [MS12]'s results to accommodate risk aversion from a nonparametric perspective. However, we do not attain the optimal global rate due to the presence of $\nabla^1 \hat{\lambda}^{-1}(\cdot)$. In the first item, $\hat{v}_j(\cdot)$ and $\hat{F}_j(\cdot)$ inherit the rate of $\hat{\lambda}^{-1}(\cdot)$. In the second item, the conditions on h_f has been chosen so that the derivative of the bid density, $\tilde{g}'(\cdot|i_j, x)$, is uniformly consistent; then, the uniform consistency of $\hat{f}_j(\cdot)$ follows by standard arguments. Note that $\tilde{g}'(\cdot|i_j, x)$ attains the optimal global rate, for example, when h_f is of the form $h_f = \gamma_f[\log(N)/N]^{R/(2R+d+3)}$ for some fixed $\gamma_f > 0$.

The third item of Proposition 1.5.2 establishes the asymptotic normality of $\hat{f}_j(v)$ for $v \in (\underline{v}, \bar{v})$. Since the bandwidth h_f satisfies $Nh_f^{d+3}h_b^{1/c(1)} \to 0$, which implies over-smoothing, the normality result is driven by the fact that $\nabla^1 \hat{\lambda}^{-1}(\cdot)$ converges faster than $\tilde{g}'(\cdot|i_j, x)$. Therefore, the latter drives the asymptotic normality as in [MS12]. Obtaining such an asymptotic distribution is important because several policy recommendations –such as the optimal reserve price– depend on the density of private values. As a consequence, the distribution of an estimator related to certain policy recommendation will be driven by $\hat{f}_j(\cdot)$, which is the slowest convergent term.

1.6 Monte Carlo Experiments and Empirical Illustration

This section investigates the finite sample performance of the proposed estimators. To that end, subsection 1.6.1 presents a limited Monte Carlo study, while subsection 1.6.2 gives an empirical illustration based on timber auction data.

1.6.1 Monte Carlo Experiments

In this subsection, I run few numerical experiments to study the finite sample performance of the estimators (1.12) and (1.13). The evaluation criteria are the bias and the variance, and for comparison purposes, I consider the estimator for the density of private values developed by [MS12].

The design of the experiment is as follows. It is assumed that there are no covariates ($d = 0$). The distribution of valuations is uniform over the interval $[0,1]$, that is, $f(v) = \mathbb{1}\{0 \leq v \leq 1\}$. For the utility function $U(\cdot)$, three functional forms are considered:

A) *Risk-averse bidders with CARA utility*: $U(y) = 1 - \exp(-5y)$

B) *Risk-averse bidders with CRRA utility*: $U(y) = y^{1/2}$

C) *Risk-neutral bidders*: $U(y) = y$

Such choices of $f(\cdot)$ and $U(\cdot)$ are convenient because the corresponding bidding functions have closed-form expressions. For each design, we run 100 replications of $N = 3,500$ auctions: $N_1 = 2,500$ auctions with $i_1 = 2$ bidders and $N_2 = 1,000$

auctions with $i_2 = 5$ bidders, giving a total of $10,000$ bids in each replication. The low number of replications is due to the high computational costs of the proposed estimators, particularly, the computation of (1.11) involves a heavy nonlinear optimization problem. The large value for N is due to the fully nonparametric nature of the estimators. Such a large sample can be found, for example, in first-price auctions of municipal bonds.

In each Monte Carlo replication, first, we generate randomly $10,000$ independent valuations $\{v_m : m = 1, 2, \ldots, 10,000\}$ from a uniform $[0, 1]$ distribution. Second, using these valuations, we compute the corresponding bids according to the equilibrium bidding function (1.1). The closed-form expression for such a function is obtained from [HMZ10], eq. (2), and depends on the design of the utility function. Third, using the generated bids, we estimate $U(\cdot)$ and $f(\cdot)$ employing (1.12) and (1.13), respectively; when estimating $f(\cdot)$, we only consider the case $j = 1$, that is, $\hat{f}_1(\cdot)$.

Similarly to previous literature, I let $R = 1$ and use the tri-weight kernel function for $k(\cdot)$ and $k_b(\cdot)$. Moreover, I employ the following bandwidths: $h_b = 0.01$, $h_f = 1.06\hat{\sigma}_j(i_j N_j)^{-1/7}$, and $h_g = 1.06\hat{\sigma}_j(i_j N_j)^{-1/5}$, where $j = 1, 2$ and $\hat{\sigma}_j$ is the estimated standard deviation of the bids from auctions with i_j participants. Regarding the construction of the sieves, I set $\bar{u} = 1.25$, $K_N = 3$, $L_N = 4$, and J_t is chosen so that $2^{J_t}/\bar{u} = 3$ for all $t = 0, 1, 2$. Neither the preceding sections nor the previous literature indicates how to choose the bandwidth h_b and establish the dimensions of the sieve spaces. In the absence of theoretical guidance, such values have been chosen to facilitate the implementation of the simulations.

Figure 1.1 shows the behavior of the estimator for the bidders' utility function, $\hat{U}(\cdot)$, under design A). The dashed line represents the simulated mean of $\hat{U}(\cdot)$, over the replications, while the dotted lines represent the $5th$ and $95th$ percentiles. The solid line depicts the true utility function $U(\cdot)$, which has been normalized to satisfy $U[\lambda_0^{-1}(\bar{r})] = 1$. Analogously, Figures 1.2 and 1.3 do the same exercise

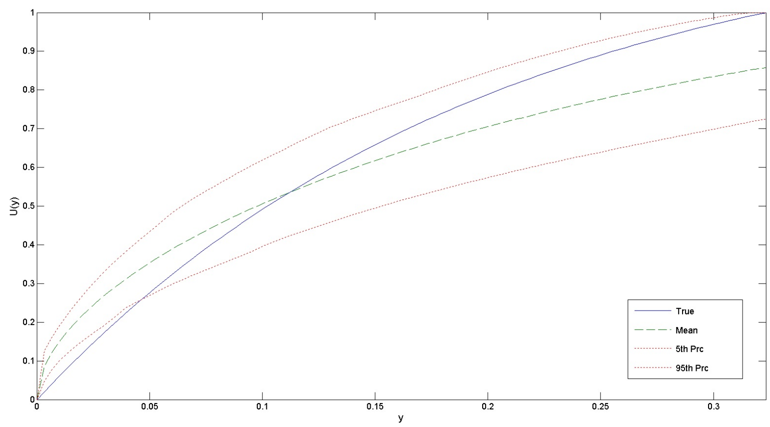

Figure 1.1: True and Estimated Utility Function - Design A)

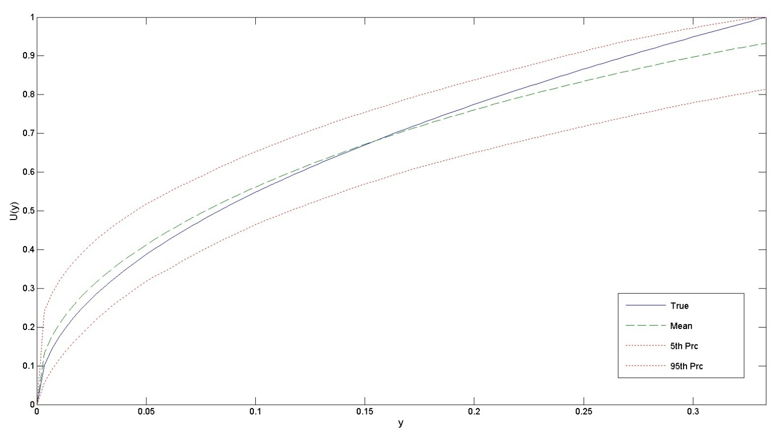

Figure 1.2: True and Estimated Utility Function - Design B)

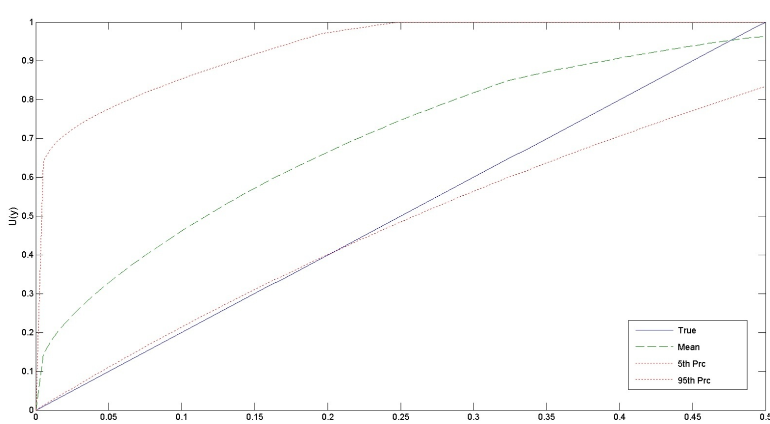

Figure 1.3: True and Estimated Utility Function - Design C)

Table 1.1: Density of Private Values - Simulated Bias and Variance

		Bias		Variance	
Design	υ	Eq. (1.13)	MS	Eq. (1.13)	MS
A)	0.1	-0.195	-0.081	0.006	0.003
	0.3	0.042	-0.240	0.011	0.006
	0.5	-0.008	-0.347	0.023	0.010
	0.7	-0.041	-0.380	0.046	0.018
B)	0.1	-0.226	-0.240	0.005	0.002
	0.3	0.030	-0.253	0.012	0.006
	0.5	0.094	-0.262	0.037	0.016
	0.7	0.088	-0.288	0.043	0.051
C)	0.1	-0.087	-0.004	0.068	0.005
	0.3	0.249	-0.028	0.057	0.019
	0.5	0.367	-0.021	0.165	0.034
	0.7	0.180	-0.142	0.971	0.058

for designs B) and C). The proposed estimator works well under designs A) and B), but it has a poor performance under design C). Note that, under the latter design, the corresponding $\lambda_0^{-1}(\cdot)$ lies at the boundary of the parameter space.

Each replication also gives us the estimated density function $\hat{f}_1(\cdot)$. Table 1.1 reports the simulated bias and variance of $\hat{f}_1(\cdot)$ evaluated at $\upsilon = 0.1, 0.3, 0.5, 0.7$, and for comparison purposes, [MS12]'s estimator (denoted by MS in the table) has been included. Under designs A) and B), $\hat{f}_1(\cdot)$ has smaller simulated bias than its competitor. However, the latter has always a smaller simulated variance. Not surprisingly, under design C), [MS12]'s estimator performs better than $\hat{f}_1(\cdot)$ in terms of bias and variance.[4]

[4]The programs were kindly provided by Vadim Marmer and Artyom Shneyrov. Simulations have been performed using MATLAB.

1.6.2 Empirical Illustration: US Forest Service Timber Auction

In this subsection, I present an empirical example that illustrates the usefulness of the proposed estimators. This empirical illustration is based on USFS timber auctions, which have been widely used in previous literature.

For comparison purposes, I use the same data set as in [LP08] and [CGP11]. The former nonparametrically estimates the bidders' utility function, as well as the density of private values, by exploiting two auction designs: ascending and first-price sealed-bid auctions. The latter adopts a semi-parametric approach by specifying risk aversion as CRRA and CARA. Then, it estimates the risk aversion parameter, as well as the density of private values, using data from first-price sealed-bid auctions only. Here, I employ the estimators (1.12) and (1.13) to non-parametrically estimate the bidders' utility function and the density of private values, respectively, using data from first-price sealed-bid auctions only.

I focus on the first-price sealed-bid auctions in 1979 for the Western half of the U.S.A. (Regions 1-6). Given the small number of auctions with more than three bidders, I only consider auctions with two and three bidders. The data provide 215 auctions: 107 auctions with two bidders and 108 auctions with three bidders, giving a total of 508 bids. For each auction, the data contain two variables characterizing each timber tract: the estimated volume in thousand board feet (mbf) and the appraisal value in dollars per mbf. The data also provide the sealed bids in dollars per mbf. As in previous literature, we consider a general specification of the utility function, so we are interested in the total bid for every tract, that is, the bid in dollars per mbf multiplied by the estimated volume in mbf. Table 1.2 reports summary statistics.[5]

Following [CGP11], I use the tract appraisal value, denoted by X_l hereafter, to explain the variability in the auctioned tracts. Specifically, X_l is computed

[5] For a detailed description of the data, see [LP08], Section 3. The data is publicly available at http://qed.econ.queensu.ca/jae/datasets/lu001.

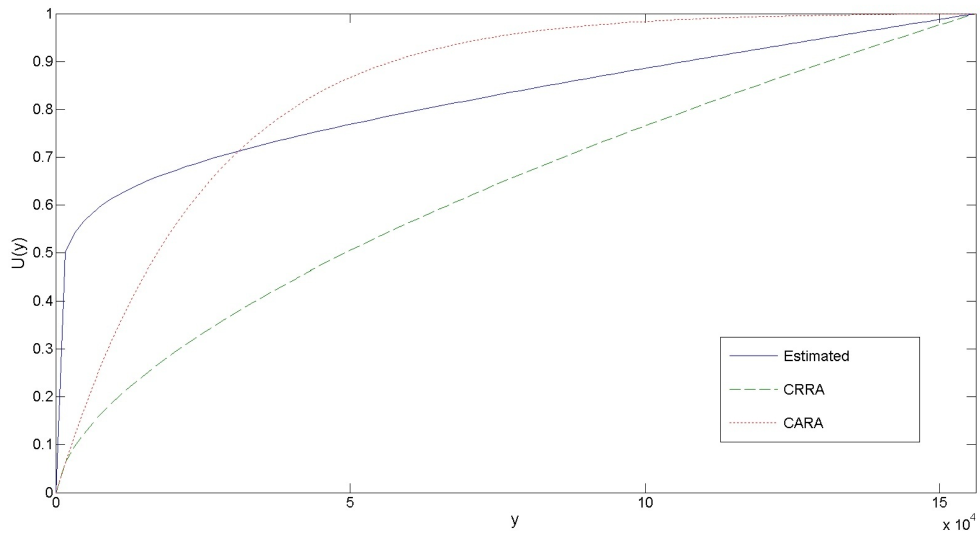

Figure 1.4: Estimated Utility Function

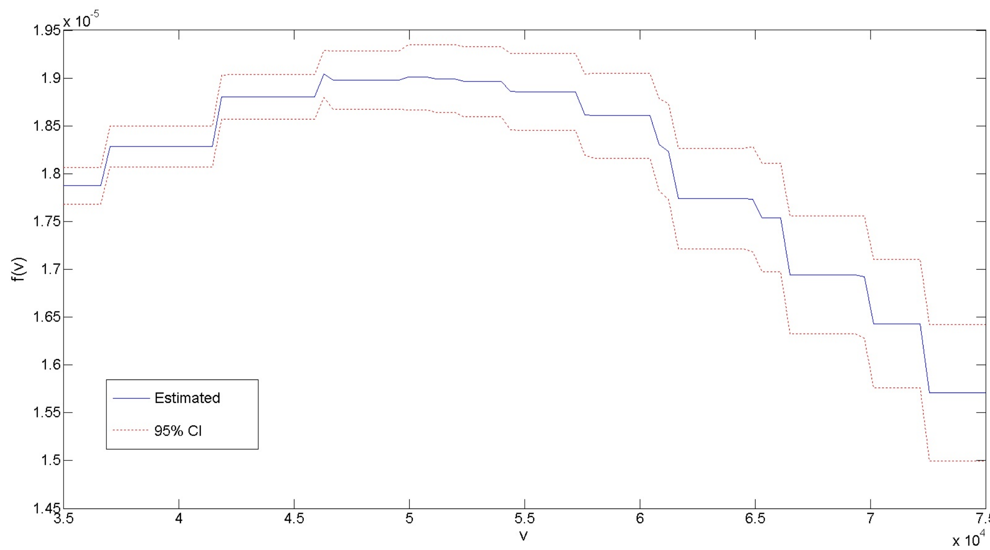

Figure 1.5: Estimated Density of Private Values

Table 1.2: Summary Statistics

	$I = 2$		$I = 3$	
	Mean	Std Dev	Mean	Std Dev
Bids ($)	77,417.70	191,996.45	96,747.53	159,255.81
Winning Bid ($)	83,625.14	201,094.98	119,356.93	188,753.48
Volume (mbf)	878.17	1,609.39	1,220.18	1,689.75
Appraisal Value ($ per mbf)	65.97	47.66	53.26	41.43

by multiplying the estimated volume in mbf and the appraisal value in dollars per mbf. To compute the preliminary kernel estimators (subsection 1.4.2), I use the same kernel and bandwidths as in the previous subsection. Regarding the construction of the sieves, I set $\bar{u} = 510,000$, $K_N = 2$, $L_N = 3$, and J_t is chosen so that $2^{J_t}/\bar{u} = 3$ for all $t = 0, 1, 2$.

Figure 1.4 displays the estimated bidders' utility function at the sample mean of X_l, which equals $\bar{X} = 49,574$. Specifically, the continuous line depicts such an estimated function using the estimator defined in (1.12). For comparison purposes, Figure 1.4 also includes an utility function corresponding to a CRRA (CARA) specification with parameter 0.6 (0.00004). These values have been obtained by [CGP11], Section 7, as well as [LP08]. As can be noted, the estimated utility function exhibits a significant concavity for values of y (monetary gain) close to zero. However, when y becomes large, the estimated function seems to be linear.[6]

In addition to estimating the bidders' utility function, I recover the density of private values using the estimator (1.13) with $j = 1$. Figure 1.5 displays the estimation results at the sample mean of X_l. The solid line depicts the estimated density of private values, while the dotted lines represent the 95% confidence

[6]To perform sensitivity analyses, the cases $L_N = 2$ and $L_N = 4$ have also been considered. Basically, we can reach similar conclusions. Additional results and figures are available upon request. I remark that [BCK07], which estimated Engel curves by the method of sieves, employed third order B-splines as sieve basis functions. Their sample size is 1,655 observations, while the dimension of the sieve space is 9.

intervals constructed with the estimated variance $\hat{V}(\cdot, 2)$ of Proposition 1.5.2.3. I highlight that, in contrast to previous literature, the estimated density is not single-peaked.

1.7 Concluding Remarks

This paper has studied a first-price sealed-bid auction with risk-averse bidders, independent private values, and a non-binding reserve price. In this context, I have proposed nonparametric estimators for the bidders' utility function and the density of private values. The key idea has been to characterize both functions by an argument that minimizes a functional over a space of smooth functions. Then, I have estimated such a minimizer in two steps by the method of sieves. The estimators for the bidders' utility function and the density of private values have been constructed as smooth nonlinear functionals of the estimator for the minimizer.

The estimator for the utility function is uniformly consistent and shape-preserving, while the estimator for the density is uniformly consistent and asymptotically normal. A limited Monte Carlo study suggests that the proposed estimators have good finite sample properties, particularly, in terms of bias reduction. To highlight the usefulness of such estimators, an application to the US Forest Service timber auctions has been provided.

The method proposed in this paper allows us to estimate bidders' risk preferences without placing any parametric restrictions –such as CARA or CRRA– on the utility function. In this way, this paper extends the literature on structural econometrics of first-price auctions by developing the first estimator for the bidders' utility function that can incorporate any type of risk preference. This extension is important as evidence strongly suggests that risk aversion is an essential component of bidders' behavior, but there is no consensus on which concept of risk aversion is the most appropriate to describe such behavior.

In comparison with existing literature, the proposed estimators have some limitations as they do not place parametric restrictions on the bidders' utility function. First, as expected, the rates of convergence of the proposed estimators are slower than the optimal global rates attained in [GPV00], [CGP11], as well as, [MS12]. Second, the validity of my results relies on the assumption that bidders' participation is exogenous, that is, the density of private values is independent of the number of bidders. Third, the proposed estimators involve a heavy nonlinear optimization problem, so the computational burden is high.

There are many interesting extensions for further research. First, it would be useful to establish the optimal rate for the estimator of the bidders' utility function, that is, the fastest rate at which the bidders' utility function can be estimated nonparametrically; see [Sto82]. Second, the asymptotic distribution of the utility function estimator needs to be determined, or alternatively, we could build conservative confidence intervals. Third, since the density of private values is over-identified, it is possible to construct a test to verify whether or not bidders' participation is exogenous. Recall that the validity of the proposed estimators depends on this exclusion restriction. Fourth, the results of this paper might be extended to more general cases such as a binding reserve price, affiliated private values, and asymmetric bidders ([GPV09], Section 4). In addition, we might consider relaxing the assumption of exogenous participation.

In view of future empirical applications, the proposed estimators can be employed to recover the set of optimal reserve prices, that is, the set of reserve prices that maximizes the expected auctioneer's revenue. This set depends on the bidders' risk-aversion and the distribution of valuations. So far, the optimal reserve price has been obtained only under the assumption that bidders are risk-neutral. For instance, [LPV03] has considered a first-price auction with affiliated private values, but assuming that bidders are risk-neutral. Their approach may be extended by allowing bidders to be risk-averse. In such a case, an estimate of the set of optimal reserve prices can be obtained by combining the obtained results with [CHT07]'s insight; the auctioneer's expected revenue may be used as the criterion function.

1.A Appendix

This appendix contains detailed proofs of the previous results. Subsection 1.A.1 provides the proofs of the mathematical results (Propositions 1.3.1-1.3.2), while subsection 1.A.2 presents the proofs of the statistical results (Lemmas 1.4.1-1.4.2 and Theorem 1.4.1). Proofs of all auxiliary lemmas are given in subsection 1.A.3.

1.A.1 Proof of Mathematical Results

In this subsection, I provide detailed proofs of Propositions 1.3.1 and 1.3.2.

1.A.1.1 Proof of Proposition 1.3.1

Before starting with the proof, the next auxiliary lemma is stated.

Lemma 1.A.1. *There exists* $\underline{K} > 0$ *such that* $\underline{K} \|\nabla^r f - \nabla^r g\|_{\infty,[0,\bar{r}]}^{c(r)} \leq \|f - g\|_{1,[0,\bar{r}]}$ *for all* $f, g \in \mathscr{H}_R$ *and* $r \in \{0, 1, \ldots, R\}$.

Proof of proposition: From the previous lemma, it suffices to show that the inequality $\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]} \leq 2Q[(\alpha_t)_t, \phi]$ holds for any $[(\alpha_t)_t, \phi] \in \Theta_R$. To that end, pick an arbitrary $[(\alpha_t)_t, \phi] \in \Theta_R$. By definition of $\mathscr{A}$, there is a finite T such that $\alpha_t(\cdot) = 0$ for all $t > T$. By a standard triangular inequality, the term $\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]}$ can be bounded above by

$$\int_0^{\bar{r}} |[\lambda_0^{-1}(u) - \phi(u)] - [\lambda_0^{-1}\{R_2[\alpha_T(u)]\} - \phi\{R_2[\alpha_T(u)]\}]| du$$
$$+2 \int_0^{\bar{r}} |R_2[\alpha_T(u)]| du.$$

Note that $|\lambda_0^{-1}(u_2) - \lambda_0^{-1}(u_1)| \le |u_2 - u_1|$ and $|\phi(u_2) - \phi(u_1)| \le |u_2 - u_1|$ for any $(u_1, u_2) \in [0, \bar{r}]^2$, then the condition $\lambda_0^{-1}(0) = \phi(0) = 0$ implies $|\lambda_0^{-1}\{R_2[\alpha_T(u)]|, |\phi\{R_2[\alpha_T(u)]| \le |R_2[\alpha_T(u)]|$. After repeated triangular inequalities, the first term can be bounded above by

$$\begin{aligned}
&\int_0^{\bar{r}} |[\lambda_0^{-1}(u) - \phi(u)] - [\lambda_0^{-1}\{R_1[\alpha_1(u)]\} - \phi\{R_1[\alpha_1(u)]\}]| + \sum_{t=1}^{T-1} \\
&|[\lambda_0^{-1}\{R_1[\alpha_t(u)]\} - \phi\{R_1[\alpha_t(u)]\}] - [\lambda_0^{-1}\{R_1[\alpha_{t+1}(u)]\} - \phi\{R_1[\alpha_{t+1}(u)]\}]| + \\
&|[\lambda_0^{-1}\{R_1[\alpha_T(u)]\} - \phi\{R_1[\alpha_T(u)]\}] - [\lambda_0^{-1}\{R_2[\alpha_T(u)]\} - \phi\{R_2[\alpha_T(u)]\}]|du \\
&\equiv \bar{Q}_1 + \bar{Q}_2 + \bar{Q}_3,
\end{aligned}$$

which immediately implies

$$\|\phi - \lambda_0^{-1}\|_{1,[0,\bar{r}]} \le \bar{Q}_1 + \bar{Q}_2 + \bar{Q}_3 + 2\int_0^{\bar{r}} |R_2[\alpha_T(u)]|du. \tag{1.14}$$

Next, I find proper upper bounds for $\bar{Q}_1$, $\bar{Q}_2$, and $\bar{Q}_3$. First, by Definition 4, condition 2, and $\Delta b[\alpha_0(u)] = \lambda_0^{-1}\{R_1[\alpha_0(u)]\} - \lambda_0^{-1}\{R_2[\alpha_0(u)]\}$, we have

$$\begin{aligned}
\bar{Q}_1 \;\le\; & 2\int_0^{\bar{r}} |R_1[\alpha_0(u)] - u|du + \\
& \int_0^{\bar{r}} |\Delta b[\alpha_0(u)] + \phi\{R_2[\alpha_0(u)]\} - \phi\{R_1[\alpha_0(u)]\}|du \\
& +2\int_0^{\bar{r}} |R_1[\alpha_1(u)] - R_2[\alpha_0(u)]|du.
\end{aligned} \tag{1.15}$$

Turning to the second term $\bar{Q}_2$, after applying a triangular inequality on each summand; more specifically, after adding and subtracting $\lambda_0^{-1}\{R_2[\alpha_t(u)]\} - \phi\{R_2[\alpha_t(u)]\}$), it follows that $\bar{Q}_2$ is bounded above by

$$\begin{aligned}
\bar{Q}_2 \;\le\; & \sum_{t=1}^{T-1} \int_0^{\bar{r}} |\Delta b[\alpha_t(u)] + \phi\{R_2[\alpha_t(u)]\} - \phi\{R_1[\alpha_t(u)]\}|du \\
& + \sum_{t=1}^{T-1} \int_0^{\bar{r}} |\lambda_0^{-1}\{R_1[\alpha_{t+1}(u)]\} - \lambda_0^{-1}\{R_2[\alpha_t(u)]\}|du \\
& + \sum_{t=1}^{T-1} \int_0^{\bar{r}} |\phi\{R_1[\alpha_{t+1}(u)]\} - \phi\{R_2[\alpha_t(u)]\}|du,
\end{aligned}$$

and as a result,

$$\bar{Q}_2 \le \sum_{t=1}^{T-1} \int_0^{\bar{r}} |\Delta b[\alpha_t(u)] + \phi\{R_2[\alpha_t(u)]\} - \phi\{R_1[\alpha_t(u)]\}| du + 2\sum_{t=1}^{T-1} \int_0^{\bar{r}} |R_1[\alpha_{t+1}(u)] - R_2[\alpha_t(u)]| du. \tag{1.16}$$

because $|\lambda_0^{-1}(u_2) - \lambda_0^{-1}(u_1)| \le |u_2 - u_1|$ and $|\phi(u_2) - \phi(u_1)| \le |u_2 - u_1|$ for any $(u_1, u_2) \in [0, \bar{r}]^2$. Regarding the third term $\bar{Q}_3$, by previous arguments we have

$$\bar{Q}_3 = \int_0^{\bar{r}} |\Delta b[\alpha_T(u)] + \phi\{R_2[\alpha_T(u)]\} - \phi\{R_1[\alpha_T(u)]\}| du. \tag{1.17}$$

In order to complete the proof, observe that $Q[(\alpha_t)_t, \phi]$ can be written as

$$Q[(\alpha_t)_t, \phi] = \int_0^{\bar{r}} |R_1[\alpha_0(u)] - u| du + \sum_{t=1}^{T-1} \int_0^{\bar{r}} |R_1[\alpha_t(u)] - R_2[\alpha_{t-1}(u)]| du + \int_0^{\bar{r}} |R_2[\alpha_T(u)]| du + \sum_{t=0}^{T} \int_0^{\bar{r}} |\Delta b[\alpha_t(u)] + \phi\{R_2[\alpha_t(u)]\} - \phi\{R_1[\alpha_t(u)\}]| du.$$

After combining together (1.14)-(1.17), the last expression yields

$$\|\phi - \lambda_0^{-1}\|_{q,[0,\bar{r}]} \le (\bar{Q}_1 + \bar{Q}_2 + \bar{Q}_3) + 2\int_0^{\bar{r}} |R_2[\alpha_T(u)]| du \le 2Q[(\alpha_t)_t, \phi].$$

Finally, the desired result follows immediately from Lemma 1.A.1.

1.A.1.2 Proof of Proposition 1.3.2

This proof is based on series of seven lemmas. The first one, Lemma 1.A.2, describes useful properties of the quantile function $b_j(\cdot)$, $j = 1, 2$, which can be extended to $R_j(\cdot)$ as $R_j(\alpha) = \alpha b_j'(\alpha)/(i_j - 1)$. In addition, Lemma 1.A.2 shows that $b_1'(0)/(i_1 - 1) > b_2'(0)/(i_2 - 1)$, which is a key result for the rest of this proof (in particular, for Lemma 1.A.8 below).

Lemma 1.A.2. *For $j = 1, 2$, $b_j(\cdot)$ has $R + 1$ continuous derivatives on $[0, 1]$, $\nabla^{R+2} b_j(\cdot)$ is continuous on $(0, 1]$, and $b_j'(\cdot)$ is bounded away from zero. Moreover, $b_1'(0)/(i_1 - 1) > b_2'(0)/(i_2 - 1)$.*

For the remaining discussion, I set $\gamma_J = 1$ to simplify technical details. Consider the partition $\{a_k^{(n)} \equiv k/\lceil K_n^{1/2} \rceil : k \in \mathbb{N}_0, k \leq \lceil K_n^{1/2} \rceil\}$ of $[0,1]$ whose size is $\lceil K_n^{1/2} \rceil$. Specifically, $(0,1]$ can be written as a disjoint union of the intervals $(a_k^{(n)}, a_{k+1}^{(n)}]$, $1 \leq k \leq \lceil K_n^{1/2} \rceil$:

$$(0,1] = \dot{\bigcup_{k=0}^{\lceil K_n^{1/2} \rceil - 1}} (a_k^{(n)}, a_{k+1}^{(n)}],$$

and the length of each interval $(a_k^{(n)}, a_{k+1}^{(n)}]$ is $\lceil K_n^{1/2} \rceil^{-1}$. Now for $\alpha \in [0,1]$, define the (left-continuous) functions $R_j^{(n)}(\cdot)$, $j = 1, 2$, as follows:

$$R_1^{(n)}(\alpha) = \begin{cases} R_1(a_k^{(n)}) + 2\bar{L}K_n^{-1/2}(\alpha - a_k^{(n)}) \\ \text{if } \alpha \in (a_k^{(n)}, a_{k+1}^{(n)}] \text{ \& } \inf_{(a_k^{(n)}, a_{k+1}^{(n)}]} |R_1'(\cdot)| \leq \bar{L}K_n^{-1/2}, \\ R_1(\alpha) \quad \text{otherwise,} \end{cases}$$

and

$$R_2^{(n)}(\alpha) = \begin{cases} R_2(a_k^{(n)}) - 2\bar{L}K_n^{-1/2}(\alpha - a_k^{(n)}) \\ \text{if } \alpha \in (a_k^{(n)}, a_{k+1}^{(n)}] \text{ \& } \inf_{(a_k^{(n)}, a_{k+1}^{(n)}]} |R_2'(\cdot)| \leq \bar{L}K_n^{-1/2}, \\ R_2(\alpha) \quad \text{otherwise,} \end{cases}$$

where

$$\bar{L} \equiv \max\{1, \|R_1''(\cdot)\|_{\infty,[\bar{\alpha}/5,1]}, \|R_2''(\cdot)\|_{\infty,[\bar{\alpha}/5,1]}, \|R_1'(\cdot)\|_{\infty,[0,1]}, \|R_2'(\cdot)\|_{\infty,[0,1]}\} < +\infty,$$

and $\bar{\alpha} \in (0,1)$ is a small constant such that $\min\{R_1'(\alpha), R_2'(\alpha)\} > b_1'(0)/[2(i_1 - 1)]$ whenever $\alpha \in [0, \bar{\alpha}]$. Note that such $\bar{\alpha}$ exists because $R_j'(\alpha) = [b_j'(\alpha) + \alpha b_j''(\alpha)]/(i_j - 1)$, $b_j'(\cdot)$ is continuous, and $b_j''(\cdot)$ is bounded (Lemma 1.A.2).

Several remarks are noteworthy. First and foremost, both $R_1^{(n)}(\cdot)$ and $R_2^{(n)}(\cdot)$ are strictly monotone on $(a_{k-1}^{(n)}, a_k^{(n)}]$ for all $1 \leq k \leq \lceil K_n^{1/2} \rceil$. Second, they are also continuously differentiable on any open interval $(a_{k-1}^{(n)}, a_k^{(n)})$. Third, if $R_1^{(n)}(\cdot)$ is differentiable at $\alpha \in (0,1)$, then it must be $|R_1^{(n)\prime}(\alpha)| \geq \bar{L}K_n^{-1/2}$. Fourth, if $R_1(\alpha) \geq 0$, then $R_1^{(n)}(\cdot)$ is strictly increasing on $(a_{k-1}^{(n)}, a_k^{(n)}] \ni \alpha$. Fifth, if $R_1^{(n)}(\cdot)$ is strictly decreasing on $(a_{k-1}^{(n)}, a_k^{(n)}]$, by construction $R_1^{(n)}(\cdot) = R_1(\cdot)$, and consequently, $R_1(\cdot)$ must be strictly decreasing on $(a_{k-1}^{(n)}, a_k^{(n)}]$ as well.

The next lemma describes in detail the relationship between $R_j(\cdot)$ and $R_j^{(n)}(\cdot)$. In particular, the second part studies the uniform rate of convergence of $R_j^{(n)}(\cdot)$ toward $R_j(\cdot)$.

Lemma 1.A.3. *There is $\tilde{N} \in \mathbb{N}$ such that $R_j^{(n)}(\alpha) = R_j(\alpha)$ for all $j \in \{1,2\}$, $n \geq \tilde{N}$, and $\alpha \in [0, \bar{\alpha}/2]$. Moreover, $\|R_j^{(n)}(\cdot) - R_j(\cdot)\|_{\infty,[0,1]} = O(K_n^{-1})$.*

Hereafter, we only consider $n \geq \tilde{N}$. Furthermore, $\tilde{N}$ is taken large enough so that $K_n^{-1/2} < \bar{\alpha}/4$ and $R_2^{(n)}(\cdot) > 0$ on $(0,1]$ for any $n \geq \tilde{N}$; the second requirement is possible due to Lemma 1.A.3 and the fact that $R_2(\cdot) > 0$ on $(0,1]$. Given these considerations, the next lemma states useful properties about $R_1^{(n)}(\cdot)$ and $R_2^{(n)}(\cdot)$.

Lemma 1.A.4. *Let $n \geq \tilde{N}$, $\alpha \in (0,1]$ and $1 \leq k \leq \lceil K_n^{1/2} \rceil$. The following statements hold:*

1. *$R_1^{(n)}(\alpha) \geq R_1(\alpha) > R_2(\alpha) \geq R_2^{(n)}(\alpha) > 0$, and obviously, $R_1^{(n)}(0) = R_2^{(n)}(0) = 0$.*
2. *The set $\{1 \leq j \leq \lceil K_n^{1/2} \rceil : R_1^{(n)}(a_j^{(n)}) \geq \bar{r}\}$ is nonempty.*
3. *Let $u \in (0, \bar{r}]$ and define $l = \min\{1 \leq j \leq \lceil K_n^{1/2} \rceil : R_1^{(n)}(a_j^{(n)}) \geq u\}$. Then, $R_1^{(n)}(a_{l-1}^{(n)}) < R_1^{(n)}(a_l^{(n)})$ and $u \in (R_1^{(n)}(a_{l-1}^{(n)}), R_1^{(n)}(a_l^{(n)})]$.*
4. *If $R_1^{(n)}(a_{k-1}^{(n)}) < R_1^{(n)}(a_k^{(n)})$ and $u \in (R_1^{(n)}(a_{k-1}^{(n)}), R_1^{(n)}(a_k^{(n)})]$, there exists $\tilde{\alpha} \in (a_{k-1}^{(n)}, a_k^{(n)}]$ such that $R_1^{(n)}(\tilde{\alpha}) = u$. Moreover, $R_1^{(n)}(\cdot)$ is strictly increasing on $(a_{k-1}^{(n)}, a_k^{(n)}]$.*
5. *If $R_1^{(n)}(\alpha) \leq \bar{r}$, for any $u \in (0, R_1^{(n)}(\alpha)]$ there is $\tilde{\alpha} \in (0, \alpha]$ such that $R_1^{(n)}(\tilde{\alpha}) = u$.*

The third and fourth statements imply that the set $\{a \in [0,1] : R_1^{(n)}(a) = u\}$ is nonempty for any $u \in [0, \bar{r}]$. Furthermore, it can be easily seen that $\#\{a \in$

$[0,1] : R_1^{(n)}(a) = u\} \leq \lceil K_n^{1/2} \rceil$ because $R_1^{(n)}(\cdot)$ is strictly monotone on each interval $(a_{k-1}^{(n)}, a_k^{(n)}]$. For each integer $n \geq \tilde{N}$, I well-define the sequence of functions $(\alpha_t^{(n)})_{t \in \mathbb{N}_0} : [0, \bar{r}] \to [0,1]^\infty$ as follows: $\alpha_0^{(n)}(u) = \min\{a \in [0,1] : R_1^{(n)}(a) = u\}$, and when $t \geq 1$,

$$\alpha_t^{(n)}(u) = \min\{a \in [0,1] : R_1^{(n)}(a) = R_2^{(n)}[\alpha_{t-1}^{(n)}(u)]\}.$$

For future convenience, I highlight the subsequent properties of $\{\alpha_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$.

Lemma 1.A.5. *For each integer $n \geq \tilde{N}$, $\alpha_0^{(n)}(\cdot)$ is strictly increasing on $[0, \bar{r}]$ and $0 < \alpha_{t+1}^{(n)}(u) < \alpha_t^{(n)}(u) \leq 1$ for all $(t, u) \in \mathbb{N}_0 \times (0, \bar{r}]$.*

Next, I describe useful properties of $\{\alpha_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$ on the open interval $(0, \bar{r})$. As a starting point, for $1 \leq l_0 \leq \lceil K_n^{1/2} \rceil$ define the set

$$V_{(l_0)} = \left\{ u \in (0, \bar{r}) : \max_{j=0,\ldots,l_0-1} R_1^{(n)}(a_j^{(n)}) < u < R_1^{(n)}(a_{l_0}^{(n)}) \right\}, \tag{1.18}$$

which of course may be empty. For $t \in \mathbb{N}_0$, let $(l_0, l_1, \ldots, l_t) \in \mathbb{N}^{t+1}$ be a $(t+1)$-tuple whose components satisfy $1 \leq l_j \leq \lceil K_n^{1/2} \rceil$ being $0 \leq j \leq t$. Using a recursive argument on $t \in \mathbb{N}$, define further the set $V_{(l_0,\ldots,l_{t-1},l_t)}$ by

$$V_{(l_0,\ldots,l_{t-1},l_t)} = \left\{ u \in V_{(l_0,\ldots,l_{t-1})} : \max_{j=0,1,\ldots,l_t-1} R_1^{(n)}(a_j^{(n)}) < R_2^{(n)}[\alpha_{t-1}^{(n)}(u)] < R_1^{(n)}(a_{l_t}^{(n)}) \right\},$$

and observe that $V_{(l_0,\ldots,l_{t-1},l_t)} \cap V_{(l_0,\ldots,l_{t-1},l_t')} = \varnothing$ whenever $l_t' \neq l_t$.

Using an inductive argument, the next lemma proves that $\alpha_t^{(n)}(\cdot)$ is continuously differentiable and strictly monotone on each $V_{(l_0,l_1,\ldots,l_{t-1},l_t)}$. Let $|Z|$ denote the Lebesgue measure of a measurable set $Z \subseteq \mathbb{R}^m$ being $m \in \mathbb{N}$.

Lemma 1.A.6. *Let $n \geq \tilde{N}$. For $t \in \mathbb{N}_0$, let $(l_0, l_1, \ldots, l_t) \in \mathbb{N}^{t+1}$ be a $(t+1)$-tuple whose components satisfy $1 \leq l_j \leq \lceil K_n^{1/2} \rceil$ being $0 \leq j \leq t$. Then, the following statements hold for every $t \in \mathbb{N}_0$:*

1. *$V_{(l_0,\dots,l_{t-1},l_t)}$ is an open interval, and if $u \in V_{(l_0,\dots,l_{t-1},l_t)}$, then $\alpha_t^{(n)}(u) \in (a_{l_t-1}^{(n)}, a_{l_t}^{(n)})$.*

2. *Both functions $\alpha_t^{(n)}(\cdot)$ and $R_2^{(n)}[\alpha_t^{(n)}(\cdot)]$ are continuously differentiable and strictly monotone on $V_{(l_0,\dots,l_{t-1},l_t)}$. Besides, $|\alpha_t^{(n)\prime}(u)| \le K_n^{(t+1)/2}$ for any $u \in V_{(l_0,\dots,l_{t-1},l_t)}$.*

3. *$\sum |V_{(l_0,\dots,l_{t-1},l_t)}| = \bar{r}$ where the sum runs over all the $(t+1)$-tuples $(l_0, l_1, \dots, l_t) \in \mathbb{N}^{t+1}$ that satisfy $1 \le l_j \le \lceil K_n^{1/2} \rceil$ being $0 \le j \le t$. The cardinality of the support of the sum is $\lceil K_n^{1/2} \rceil^{t+1}$.*

Now I introduce a basic error estimation tool in approximation theory: the (first) 1-modulus of smoothness; see [DP97], pp. 190. Specifically, for a real function $f : \mathbb{R} \to \mathbb{R}$ and a positive real number $\epsilon > 0$, this concept is defined by

$$\omega[f(\cdot), \epsilon] = \sup_{0<|h|\le\epsilon} \|f(\cdot + h) - f(\cdot)\|_{1,\mathbb{R}} = \sup_{0<|h|\le\epsilon} \int_{\mathbb{R}} |f(u+h) - f(u)| du.$$

To apply this concept to the functions $\alpha_t^{(n)}(\cdot)$, $t \in \mathbb{N}_0$, I extend the domain of $\{\alpha_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$ from $[0, \bar{r}]$ to $\mathbb{R}$ in the straightforward way: $\alpha_t^{(n)}(\alpha) = 0$ whenever $\alpha \notin [0, \bar{r}]$.

The next lemma determines upper bounds on the 1-moduli of smoothness of $\{\alpha_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$.

Lemma 1.A.7. *Let $c > 0$ be a fixed constant. For any $t \in \mathbb{N}_0$, we have*

$$\omega\left[\alpha_t^{(n)}(\cdot), c\lceil K_n^{1/2} \rceil^{-(t+5)}\right] \le c(2^4 + \bar{r}) K_n^{-2}.$$

Now we establish a crucial result about the uniform rate of convergence of $\{\alpha_t^{(n)}(\cdot) : t \in \mathbb{N}_0\}$.

Lemma 1.A.8. *Let $(T_m)_{m\in\mathbb{N}}$ be an increasing divergent sequence of nonnegative integers. There exist constants $\bar{K}, \bar{N} < +\infty$ such that $T_m \|\alpha_{T_m}^{(n)}(\cdot)\|_{\infty,[0,\bar{r}]} \leq \bar{K}$ for all $m, n \geq \bar{N}$. As a consequence,*

1. *there is $N^* \in \mathbb{N}$ such that $K_m \|\alpha_{K_m - 1}^{(n)}(\cdot)\|_{\infty,[0,\bar{r}]} \leq 2\bar{K}$ for all $m, n \geq N^*$, and*
2. *there is $T \in \mathbb{N}$ such that $\|\alpha_t^{(n)}(\cdot)\|_{\infty,[0,\bar{r}]} \leq \bar{\alpha}/2$ for all $t, n \geq T$.*

Finally, we are ready to state the desired proof, which follows a constructive approach in the sense that it provides formulas for $(A_t^{(n)})_{t\in\mathbb{N}_0} \in \mathscr{A}^{(n)}$ and $P^{(n)} \in \mathscr{H}^{(n)}$.

Proof of proposition: To simplify the remaining discussion, just consider $n \in \mathbb{N}$ sufficiently large so that $\min\{n, K_n\}$ is strictly greater than $\max\{\tilde{N}, N^*, T\}$; see Lemmas 1.A.3 and 1.A.8 above. This proof is divided into three steps. In the first one, for each $n \in \mathbb{N}$, I propose a sequence of functions $(A_t^{(n)})_{t\in\mathbb{N}_0} \in \mathscr{A}^{(n)}$ to approximate $(\alpha_t^{(n)})_{t\in\mathbb{N}_0}$, and then, I describe its approximation error as n grows to infinity. The second step shows that $Q_1[(A_t^{(n)})_t] = O(K_n^{-1})$ as $n \to +\infty$. The third one proves that the Bernstein polynomial $P^{(n)}(\cdot)$, which has been defined in (1.7), belongs to $\mathscr{H}^{(n)}$ and satisfies the second part of the proposition, that is, $\|Q_2[\cdot, P^{(n)}]\|_{\infty,\mathscr{A}^{(n)}} = O(K_n^{-1})$.

Step 1: When $0 \leq t \leq K_n - 1$, each function $\alpha_t^{(n)}(\cdot)$ can be approximated by the wavelet operator at the J_t-th resolution level ([DP97], eq. 2.1.1):

$$A_t^{(n)}(u) = \sum_{j\in\mathbb{Z}} \tilde{a}_{t,j}^{(n)} \kappa[2^{J_t}(u/\bar{u}) - j],$$

where $u \in [0, \bar{u}]$ and the coefficients $\tilde{a}_{t,j}$ are defined by

$$\tilde{a}_{t,j}^{(n)} = \frac{2^{J_t}}{\bar{u}} \int_0^{\bar{r}} \alpha_t^{(n)}(v) \kappa[2^{J_t}(v/\bar{u}) - j] dv = \int_{-1}^{1} \alpha_t^{(n)} \left[\frac{\bar{u}(w+j)}{2^{J_t}}\right] \kappa(w) dw.$$

When $t \geq K_n$, $A_t^{(n)}(\cdot)$ simply becomes the zero function: $A_t^{(n)}(u) = 0$ for any $t \geq K_n$ and $u \in [0, \bar{u}]$.

It can be shown that $(A_t^{(n)})_{t\in\mathbb{N}_0} \in \mathscr{A}^{(n)}$. First, as the support of $\kappa(\cdot)$ is $[-1,1]$ and $\bar{u} \geq \bar{r}$, it follows that $\tilde{a}_{t,j}^{(n)} = 0$ when either $j < 0$ or $j > \lceil 2^{J_t} \rceil$. Second, since $\kappa(\cdot)$ integrates one and $0 \leq \alpha_t^{(n)}(\cdot) \leq 1$, then $0 \leq \tilde{a}_{t,j}^{(n)} \leq 1$. For future convenience, I remark that $\tilde{a}_{t,j}^{(n)} \leq \bar{\alpha}/2$ for any $t \geq T$ and $j \in \mathbb{Z}$ (Lemma 1.A.8.2), and consequently, $A_t^{(n)}(\cdot) \leq \bar{\alpha}/2$.

From [DP97], Theorem 2.2.1 and Corollary 2.1.1, we obtain that

$$\|A_t^{(n)}(\cdot)-\alpha_t^{(n)}(\cdot)\|_{1,[0,\bar{r}]} \leq 2c_2\omega\left[\alpha_t^{(n)}(\cdot), 2^{1-J_t}\right] = 2c_2\omega\left[\alpha_t^{(n)}(\cdot), 2\lceil K_n^{1/2}\rceil^{-(t+5)}\right] \leq C_2 K_n^{-2},$$

where c_2 and C_2 are finite constants (independent of t and n) given by [DP97] and Lemma 1.A.7 in Appendix 1.A.3.

Step 2: As a starting point, we can write $Q_1[(A_t^{(n)})_t] \equiv \bar{Q}_1 + \bar{Q}_2 + \bar{Q}_3$, where

$$\begin{aligned}
\bar{Q}_1 &= \int_0^{\bar{r}} |R_1[A_0^{(n)}(u)] - u| du, \\
\bar{Q}_2 &= \sum_{t=1}^{T} \int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]| du, \\
\bar{Q}_3 &= \sum_{t=T+1}^{K_n-1} \int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]| du + \int_0^{\bar{r}} |R_2[A_{K_n-1}^{(n)}(u)]| du,
\end{aligned}$$

and T is given by Lemma 1.A.8.2. To simplify the remaining discussion, the dependence of $\bar{Q}_1$, $\bar{Q}_2$, and $\bar{Q}_3$ on n has been omitted from the notation. The next step is to find proper upper bounds for $\bar{Q}_1$, $\bar{Q}_2$, and $\bar{Q}_3$. First, after applying the usual triangular inequalities and since $u = R_1^{(n)}[\alpha_0^{(n)}(u)]$, we have that

$$\begin{aligned}
\bar{Q}_1 &\leq \int_0^{\bar{r}} |R_1[A_0^{(n)}(u)] - R_1^{(n)}[A_0^{(n)}(u)]| du \\
&\quad + \int_0^{\bar{r}} |R_1^{(n)}[A_0^{(n)}(u)] - R_1^{(n)}[\alpha_0^{(n)}(u)]| du \\
&\leq \bar{r}\|R_1^{(n)}(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + \int_0^{\bar{r}} |R_1^{(n)}[A_0^{(n)}(u)] - R_1^{(n)}[\alpha_0^{(n)}(u)]| du \\
&\leq 3\bar{r}\|R_1^{(n)}(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + \bar{L}\|A_0^{(n)}(\cdot) - \alpha_0^{(n)}(\cdot)\|_{1,[0,\bar{r}]}.
\end{aligned}$$

Hence, $\bar{Q}_1 = O(K_n^{-1})$ by Lemma 1.A.3 and Step 1. Second, proceeding in a similar manner for $1 \le t \le T$, it follows that

$$\begin{aligned}
&\int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]|du \\
&\le \int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_1^{(n)}[\alpha_t^{(n)}(u)|du \\
&\quad + \int_0^{\bar{r}} |R_2^{(n)}[\alpha_{t-1}^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]|du \\
&\le 3\bar{r}\left[\|R_1^{(n)}(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + \|R_2^{(n)}(\cdot) - R_2(\cdot)\|_{\infty,[0,1]}\right] \\
&\quad + \bar{L}\left[\|A_t^{(n)}(\cdot) - \alpha_t^{(n)}(\cdot)\|_{1,[0,\bar{r}]} + \|A_{t-1}^{(n)}(\cdot) - \alpha_{t-1}^{(n)}(\cdot)\|_{1,[0,\bar{r}]}\right].
\end{aligned}$$

Since T is finite and independent of n (Lemma 1.A.8), the term $\bar{Q}_2$ also becomes $O(K_n^{-1})$. Third, when $t \ge T+1$, it follows that $0 \le \alpha_t^{(n)}(\cdot), A_t^{(n)}(\cdot) \le \bar{\alpha}/2$ (Lemma 1.A.8.2 and *Step 1*), and therefore

$$R_1[\alpha_t^{(n)}(\cdot)] = R_1^{(n)}[\alpha_t^{(n)}(\cdot)] = R_2^{(n)}[\alpha_{t-1}^{(n)}(\cdot)] = R_2[\alpha_{t-1}^{(n)}(\cdot)]$$

because of $n \ge \tilde{N}$ and Lemma 1.A.3. Then, each summand of $\bar{Q}_3$ can be bounded above as follows:

$$\begin{aligned}
&\int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]|du \\
&\le \int_0^{\bar{r}} |R_1[A_t^{(n)}(u)] - R_1[\alpha_t^{(n)}(u)]|du + \int_0^{\bar{r}} |R_2[\alpha_{t-1}^{(n)}(u)] - R_2[A_{t-1}^{(n)}(u)]|du \\
&\le \bar{L}\left[\|A_t^{(n)}(\cdot) - \alpha_t^{(n)}(\cdot)\|_{1,[0,\bar{r}]} + \|A_{t-1}^{(n)}(\cdot) - \alpha_{t-1}^{(n)}(\cdot)\|_{1,[0,\bar{r}]}\right],
\end{aligned}$$

and also,

$$\begin{aligned}
\int_0^{\bar{r}} |R_2[A_{K_n-1}^{(n)}(u)]|du &\le \int_0^{\bar{r}} |R_2[A_{K_n-1}^{(n)}(u)] - R_2[\alpha_{K_n-1}^{(n)}(u)]|du \\
&\quad + \int_0^{\bar{r}} |R_2[\alpha_{K_n-1}^{(n)}(u)]|du \\
&\le \bar{L}\left[\|A_{K_n-1}^{(n)}(\cdot) - \alpha_{K_n-1}^{(n)}(\cdot)\|_{1,[0,\bar{r}]} + \|\alpha_{K_n-1}^{(n)}(\cdot)\|_{\infty,[0,\bar{r}]}\right] \\
&= O(K_n^{-1}),
\end{aligned}$$

where the last equality follows from *Step 1* and Lemma 1.A.8.1. As a result, $\bar{Q}_3 = O(K_n^{-1})$.

Step 3: First, we prove that $P^{(n)} \in \mathscr{H}^{(n)}$ when n is sufficiently large, which implies that $P^{(n)}$ preserves the shape and smoothness of $\lambda_0^{-1}(\cdot)$ because $\mathscr{H}^{(n)} \subseteq$

$\mathscr{H}_R$. Observe that the coefficients $\{b_j^{(n)} : j = 0, 1, \ldots, L_n\}$ satisfy the conditions 1-3 of Definition 6. First, $b_0^{(n)} = \lambda_0^{-1}(0) = 0$. Second,

$$\frac{\bar{u}}{L_n^2} \leq \left(\frac{\bar{u}}{L_n}\right) \inf_{u \in [0,\bar{u}]} \nabla^1 \lambda_0^{-1}(u) \leq b_{j+1}^{(n)} - b_j^{(n)} \leq \left(\frac{\bar{u}}{L_n}\right) \sup_{u \in [0,\bar{u}]} \nabla^1 \lambda_0^{-1}(u) \leq \frac{\bar{u}}{L_n},$$

where $0 \leq j \leq L_n - 1$. For the first inequality, note that n is large enough and $\inf_{u \in [0,\bar{u}]} \nabla^1 \lambda_0^{-1}(u) > 0$ because $\lambda_0'(\cdot)$ is strictly positive and bounded above on $[0, \bar{u}]$. For the last inequality, note that $\sup_{u \in [0,\bar{u}]} \nabla^1 \lambda_0^{-1}(u) \leq 1$ because $\lambda_0'(\cdot) \geq 1$. Third, proceeding in a similar manner, the last condition follows from $\|\nabla^{R+1} \lambda_0^{-1}\|_{\infty, \mathbb{R}_{\geq 0}} \leq \bar{H}$.

Second, observe that

$$\begin{aligned} &Q_2[(A_t)_t, P^{(n)}] = \\ &\sum_{t=0}^{K_n - 1} \int_0^{\bar{r}} |\Delta b[A_t(u)] + P^{(n)}\{R_2[A_t(u)]\} - P^{(n)}\{R_1[A_t(u)]\}| du \\ &\leq 2K_n \bar{r} \|P^{(n)} - \lambda_0^{-1}\|_{\infty, [0,\bar{r}]}, \end{aligned}$$

any arbitrary $(A_t)_t \in \mathscr{A}^{(n)}$, because $\Delta b(\cdot) = \lambda_0^{-1}[R_1(\cdot)] - \lambda_0^{-1}[R_2(\cdot)]$ on $[0, 1]$ and $0 \leq R_1(\cdot), R_2(\cdot) \leq \bar{r}$. To complete the proof, recall from the discussion of Section 1.3.3 that $\|P^{(n)} - \lambda_0^{-1}\|_{\infty, [0,\bar{r}]} = O(L_n^{-1})$.

1.A.2 Proofs of Statistical Results

This subsection presents the proof of Lemmas 1.4.1-1.4.2 and Theorem 1.4.1. To simplify technicals details, I set $\gamma_J = \gamma_L = 1$.

1.A.2.1 Proof of Lemma 1.4.1

Before starting with the proof, an auxiliary lemma is stated, which obtains the uniform rates of convergence in probability for $\hat{\varphi}(x)$, $\hat{\pi}(\cdot|x)$, $\hat{G}(\cdot|\cdot, x)$, $\hat{g}(\cdot|\cdot, x)$, $\hat{b}_j(\cdot)$, and $\hat{b}_j'(\cdot)$.

Lemma 1.A.9. *Under Assumptions 1-2, the following statements hold for $j = 1, 2$:*

1. $|\hat{\varphi}(x) - \varphi(x)| = O_P\left([\log(N)/(Nh_G^d)]^{1/2} + h_G^{R+1}\right)$,

2. $|\hat{\pi}(i_j|x) - \pi(i_j|x)| = O_P\left([\log(N)/(Nh_G^d)]^{1/2} + h_G^{R+1}\right)$,

3. $\|\hat{G}(\cdot|i_j, x) - G(\cdot|i_j, x)\|_{\infty,[\underline{b},\bar{b}_j]} = O_P\left([\log(N)/(Nh_G^d)]^{1/2} + h_G^{R+1}\right)$, *and*

4. $\|\hat{g}(\cdot|i_j, x) - g(\cdot|i_j, x)\|_{\infty,(\underline{b}+h_g,\bar{b}_j-h_g)} = O_P\left([\log(N)/(Nh_g^{d+1})]^{1/2} + h_g^R\right)$.

Under Assumptions 1-3,

5. $|\hat{\underline{b}} - \underline{b}| = O_P(h_\partial)$ *and* $|\hat{\bar{b}}_j - \bar{b}_j| = O_P(h_\partial)$,

6. $\|\hat{b}_j(\cdot) - b_j(\cdot)\|_{\infty,[h_b,1-h_b]} = O_P(h_G^{R+1})$ *and* $\|\hat{b}_j(\cdot) - b_j(\cdot)\|_{\infty,[0,h_b)\cup(1-h_b,1]} = O_P(h_b)$,

7. $\underline{b} + h_g < \hat{b}_j(h_b) \le \hat{b}_j(1 - h_b) < \bar{b}_j - h_g$ *w.p.a.1,*

8. $\hat{g}[\hat{b}_j(\cdot)|i_j, x]$ *is bounded away from zero on* $[h_b, 1 - h_b]$ *w.p.a.1,*

9. $|\hat{\underline{g}}_j - g(\underline{b}|i_j, x)| = O_P(h_g)$ *and* $|\hat{\bar{g}}_j - g(\bar{b}_j|i_j, x)| = O_P(h_g)$, *and*

10. $\|\hat{b}_j'(\cdot) - b_j'(\cdot)\|_{\infty,[h_b,1-h_b]} = O_P(h_g^R)$ *as well as* $\|\hat{b}_j'(\cdot) - b_j'(\cdot)\|_{\infty,[0,h_b)\cup(1-h_b,1]} = O_P(h_b)$.

Proof of lemma: Before proceeding, define approximating functions of $b_j(\cdot)$ and $R_j(\cdot)$ as follows:

$$\begin{aligned}
\tilde{b}_j(\alpha) &= \frac{1}{h_b\bar{k}_b(\alpha)}\int_0^1 b_j(v)k_b\left(\frac{\alpha - v}{h_b}\right)dv, \\
\tilde{R}_j(\alpha) &= \frac{1}{h_b\bar{k}_b(\alpha)}\int_0^1 \frac{vb_j'(v)}{i_j - 1}k_b\left(\frac{\alpha - v}{h_b}\right)dv,
\end{aligned}$$

and $\Delta\tilde{b}(\alpha) \equiv \tilde{b}_2(\alpha) - \tilde{b}_1(\alpha) \ge 0$. Using these definitions, this proof is divided into two steps. The first one studies the approximation errors $\|\Delta\tilde{b}(\cdot) - \Delta b(\cdot)\|_{\infty,[0,1]}$ and $\|\tilde{R}_j(\cdot) - R_j(\cdot)\|_{\infty,[0,1]}$. The second step establishes uniform rates of convergence in probability for $\|\Delta\hat{b}(\cdot) - \Delta\tilde{b}(\cdot)\|_{\infty,[0,1]}$, as well as, $\|\hat{R}_j(\cdot) - \tilde{R}_j(\cdot)\|_{\infty,[0,1]}$.

Step 1: First, when $\alpha \in [h_b, 1 - h_b]$, it can be easily seen that $|\tilde{b}_j(\alpha) - b_j(\alpha)| \le h_b\|b_j'(\cdot)\|_{\infty,[0,1]}$, and similarly, $|\tilde{R}_j(\alpha) - R_j(\alpha)| \le h_b\|R_j'(\cdot)\|_{\infty,[0,1]}$. Just recall from

Lemma 1.A.2 that $b_j(\cdot)$ is twice continuously differentiable on $[0,1]$. Second, consider the case $\alpha \in [0, h_b)$. Since $b_1(0) = b_2(0) = \underline{b}$, we have that

$$\begin{aligned}
|\Delta\tilde{b}(\alpha) - \Delta b(\alpha)| &= \left| \int_{-\alpha/h_b}^{1} [b_2(\alpha + h_b v) - b_1(\alpha + h_b v)] k_b(v) dv - [b_2(\alpha) - b_1(\alpha)] \right| \\
&\le \int_{-\alpha/h_b}^{1} |b_2(\alpha + h_b v) - b_1(\alpha + h_b v)| k_b(v) dv + |b_2(\alpha) - b_1(\alpha)| \\
&\le \left(2h_b \int_{-1}^{1} |v| k_b(v) dv + 4h_b \right) \max\{\|b_1'(\cdot)\|_{\infty,[0,1]}, \|b_2'(\cdot)\|_{\infty,[0,1]}\},
\end{aligned}$$

which implies $\|\Delta\tilde{b}(\cdot) - \Delta b(\cdot)\|_{\infty,[0,h_b]} = O(h_b)$. Similarly,

$$\begin{aligned}
|\tilde{R}_j(\alpha) - R_j(\alpha)| &= \frac{1}{i_j - 1} \left| \int_{-\alpha/h_b}^{1} (\alpha + h_b v) b_j'(\alpha + h_b v) k_b(v) dv - \alpha b_j'(\alpha) \right| \\
&\le \frac{1}{i_j - 1} \int_{-\alpha/h_b}^{1} |(\alpha + h_b v) b_j'(\alpha + h_b v)| k_b(v) dv + |\alpha b_j'(\alpha)| \\
&\le \frac{1}{i_j - 1} \left(h_b \int_{-1}^{1} |v| k_b(v) dv + 2h_b \right) \|b_j'(\cdot)\|_{\infty,[0,1]}.
\end{aligned}$$

Third, when $\alpha \in (1 - h_b, 1]$, the desired result follows from the properties of boundary corrected kernels. As a result, $\|\Delta\tilde{b}(\cdot) - \Delta b(\cdot)\|_{\infty,[0,1]} = O(h_b)$ and $\|\tilde{R}_j(\cdot) - R_j(\cdot)\|_{\infty,[0,1]} = O(h_b)$

Step 2: To prove that $\|\Delta\hat{b}(\cdot) - \Delta\tilde{b}(\cdot)\|_{\infty,[0,1]} = O_P(h_b)$, it suffices to show

$$\|\max\{\hat{b}_2(\cdot) - \hat{b}_1(\cdot), 0\} - [b_2(\cdot) - b_1(\cdot)]\|_{\infty,[0,1]} = O_P(h_b).$$

For this purpose, just recall that $b_2(\alpha) \ge b_1(\alpha)$, so for any $\alpha \in [0,1]$ we have

$$|\max\{\hat{b}_2(\alpha) - \hat{b}_1(\alpha), 0\} - [b_2(\alpha) - b_1(\alpha)]| \le |[\hat{b}_2(\alpha) - \hat{b}_1(\alpha)] - [b_2(\alpha) - b_1(\alpha)]|. \quad (1.19)$$

After combining Lemma 1.A.9.6 with above expression, the desired result emerges. For the second part, observe that

$$\hat{R}_1(\alpha) - \tilde{R}_1(\alpha) = \frac{1}{i_1 - 1} \int_{-\alpha/h_b}^{1} (\alpha + h_b v) [\hat{b}_1'(\alpha + h_b v) - b_1'(\alpha + h_b v)] k_b(v) dv$$

for any $\alpha \in [0, 1 - h_b]$, and a similar result holds for $\alpha \in (1 - h_b, 1]$. Then, $\|\hat{R}_1(\cdot) - \tilde{R}_1(\cdot)\|_{\infty,[0,1]} = O_P(h_b)$ because both $b_j'(\cdot)$ and $\hat{b}_j'(\cdot) = 1/\hat{g}[\hat{b}_j(\cdot)|i_j, x]$ are uniformly bounded above w.p.a.1. Using arguments similar to the ones in (1.19), it can be shown that $\|\hat{R}_2(\cdot) - \tilde{R}_2(\cdot)\|_{\infty,[0,1]} = O_P(h_b)$.

1.A.2.2 Proof of Lemma 1.4.2

This proof is essentially based on Lemma 1.4.1, as well as, on the inequalities $\|f\|_{1,[0,\bar{r}]} \le \bar{r}\|f\|_{\infty,[0,\bar{r}]}$ and $|\|f\|_{1,[0,\bar{r}]} - \|g\|_{1,[0,\bar{r}]}| \le \|f-g\|_{1,[0,\bar{r}]}$ where f and g are integrable functions. Pick any arbitrary $[(A_t)_t, P] \in \Theta^{(N)}$. Since $0 \le A_0(\cdot) \le 1$, it can be easily seen that

$$\left|\int_0^{\hat{r}} |\hat{R}_1[A_0(u)] - u|du - \int_0^{\bar{r}} |R_1[A_0(u)] - u|du\right| \\ \le 2\bar{r}(\|\hat{R}_1(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + |\hat{r} - \bar{r}|)$$

w.p.a.1 because $\|\hat{R}_1(\cdot)\|_{\infty,[0,\bar{r}]} \le 2\bar{r}$ w.p.a.1. Similarly, for each $t \ge 1$, we have

$$\left|\int_0^{\hat{r}} |\hat{R}_1[A_t(u)] - \hat{R}_2[A_{t-1}(u)]|du - \int_0^{\bar{r}} |R_1[A_t(u)] - R_2[A_{t-1}(u)]|du\right| \\ \le 2\bar{r}(\|\hat{R}_1(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + \|\hat{R}_2(\cdot) - R_2(\cdot)\|_{\infty,[0,1]} + |\hat{r} - \bar{r}|).$$

w.p.a.1. Hence, $|\hat{Q}_1[(A_t)_t] - Q_1[(A_t)_t]|$is bounded above by

$$\left|\int_0^{\hat{r}} |\hat{R}_1[A_0(u)] - u|du - \int_0^{\bar{r}} |R_1[A_0(u)] - u|du\right| \\ + \sum_{t=1}^{K_N-1} \left|\int_0^{\hat{r}} |\hat{R}_1[A_t(u)] - \hat{R}_2[A_{t-1}(u)]|du - \int_0^{\bar{r}} |R_1[A_t(u)] - R_2[A_{t-1}(u)]|du\right| \\ \le 2K_N\bar{r}(\|\hat{R}_1(\cdot) - R_1(\cdot)\|_{\infty,[0,1]} + \|\hat{R}_2(\cdot) - R_2(\cdot)\|_{\infty,[0,1]} + |\hat{r} - \bar{r}|).$$

As $h_b \le (2\gamma_K/K_N)^2$ when N is sufficiently large (Assumption 4), the first desired result emerges from Lemma 1.4.1.

1.A.2.3 Proof of Theorem 1.4.1

Note first that $h_b^{1/2}/\gamma_K \ge 1/K_N$ due to Assumption 4. Given $\underline{K} > 0$ obtained in Proposition 1.3.1, by a standard triangular inequality and for any $\bar{K} > 0$, we have

$$\mathbb{P}\left[\|\nabla^r \hat{\lambda}^{-1} - \nabla^r \lambda_0^{-1}\|_{\infty,[0,\bar{r}]} \ge [4\bar{K}^2 h_b/(\underline{K}\gamma_K)^2]^{1/[2c(r)]}\right] \\ \le \mathbb{P}\left[\|\nabla^r \hat{\lambda}^{-1} - \nabla^r \lambda_0^{-1}\|_{\infty,[0,\bar{r}]}^{c(r)} \ge 2\bar{K}/(\underline{K}K_N)\right] \\ \le \mathbb{P}\left[Q[(\hat{A}_t)_t, \hat{\lambda}^{-1}] \ge 2\bar{K}/K_N\right] \\ \le \mathbb{P}\left[Q_1[(\hat{A}_t)_t] \ge \bar{K}/K_N\right] + \mathbb{P}\left[Q_2[(\hat{A}_t)_t, \hat{\lambda}^{-1}] \ge \bar{K}/K_N\right].$$

On the one hand, as $(\hat{A}_t)_t$ minimizes $\hat{Q}_1(\cdot)$, it follows

$$\begin{aligned}\mathbb{P}\left[Q_1[(\hat{A}_t)_t] \geq \bar{K}/K_N\right] &\leq \mathbb{P}\left[\left|Q_1[(\hat{A}_t)_t] - \hat{Q}_1[(\hat{A}_t)_t]\right| + \hat{Q}_1[(\hat{A}_t)_t] \geq \bar{K}/K_N\right] \\ &\leq \mathbb{P}\left[\|\hat{Q}_1 - Q_1\|_{\infty,\mathscr{A}^{(N)}} + \hat{Q}_1[(A_t^{(N)})_t] \geq \bar{K}/K_N\right] \\ &\leq \mathbb{P}\left[2\|\hat{Q}_1 - Q_1\|_{\infty,\mathscr{A}^{(N)}} + Q_1[(A_t^{(N)})_t] \geq \bar{K}/K_N\right].\end{aligned}$$

On the other hand, by the same arguments

$$\begin{aligned}\mathbb{P}\left[Q_2[(\hat{A}_t)_t, \hat{\lambda}^{-1}] \geq \bar{K}/K_N\right] &\leq \mathbb{P}\left[\|\hat{Q}_2 - Q_2\|_{\infty,\Theta^{(N)}} + \hat{Q}_2[(\hat{A}_t)_t, P^{(N)}] \geq \bar{K}/K_N\right] \\ &\leq \mathbb{P}\left[2\|\hat{Q}_2 - Q_2\|_{\infty,\Theta^{(N)}} + Q_2[(\hat{A}_t)_t, P^{(N)}] \geq \bar{K}/K_N\right].\end{aligned}$$

To complete the proof, in view of Proposition 1.3.2 and Lemma 1.4.2, just pick $\bar{K}$ large enough.

1.A.2.4 Proof of Proposition 1.5.1

The approach of this proof is similar to that of [Mat03], proof of Theorem 1. Pick any $y \in [0, \lambda_0^{-1}(\bar{r})]$. For the first part, on the one hand, if $0 \leq y \leq \hat{\lambda}^{-1}(\hat{r})$, then

$$\begin{aligned}|\hat{\lambda}(y) - \lambda_0(y)| &= \frac{1}{\nabla^1\lambda_0^{-1}(\lambda^*)}|\lambda_0^{-1}[\hat{\lambda}(y)] - \lambda_0^{-1}[\lambda_0(y)]| \\ &= \frac{1}{\nabla^1\lambda_0^{-1}(\lambda^*)}|\lambda_0^{-1}[\hat{\lambda}(y)] - \hat{\lambda}^{-1}[\hat{\lambda}(y)]| \\ &\leq \frac{1}{\nabla^1\lambda_0^{-1}(\lambda^*)}\left[\|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[0,\bar{r}]} + \|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[\bar{r},\hat{r}]}\right],\end{aligned}$$

where λ^* lies between $\hat{\lambda}(y)$ and $\lambda_0(y)$, while the notation here is $\|\cdot\|_{\infty,[\bar{r},\hat{r}]} \equiv 0$ if $\bar{r} > \hat{r}$. As a consequence, $\|\hat{\lambda}(\cdot) - \lambda_0(\cdot)\|_{\infty,[0,\hat{\lambda}^{-1}(\hat{r})]}$ is bounded above by

$$\frac{1}{\nabla^1\lambda_0^{-1}(\lambda^{**})}\left[\|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[0,\bar{r}]} + \|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[\bar{r},\hat{r}]}\right],$$

where $\lambda^{**} \in [0, \max\{\hat{r}, \bar{r}\}]$; therefore, $0 \leq \lambda^{**} \leq 2\bar{r}$ w.p.a.1. Recall that $0 \leq \nabla^1\hat{\lambda}^{-1}(\cdot) \leq 1$ on $[0, \bar{u}]$, and also, that $\bar{u}$ has been chosen so that $\bar{u} > \hat{r}$ w.p.a.1. As a result, it follows that $\|\hat{\lambda}(\cdot) - \lambda_0(\cdot)\|_{\infty,[0,\hat{\lambda}^{-1}(\hat{r})]} = O_P(h_b^{1/[2c(0)]})$. On the other hand, if $\hat{\lambda}^{-1}(\hat{r}) < y \le \lambda_0^{-1}(\bar{r})$, then $\lambda_0[\hat{\lambda}^{-1}(\hat{r})] < \lambda_0(y) \le \lambda_0[\lambda_0^{-1}(\bar{r})] = \bar{r}$ and $\hat{\lambda}(y) = \hat{r}$. As a result,

$$\begin{aligned}
|\hat{\lambda}(y) - \lambda_0(y)| &\le |\hat{r} - \bar{r}| + \bar{r} - \lambda_0(y) \\
&\le |\hat{r} - \bar{r}| + \lambda_0[\lambda_0^{-1}(\bar{r})] - \lambda_0[\hat{\lambda}^{-1}(\bar{r})] \\
&\le |\hat{r} - \bar{r}| + \left[\sup_{y \in [0, \lambda_0^{-1}(\bar{r})]} \lambda_0'(y)\right] \|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[0,\bar{r}]}.
\end{aligned}$$

The second part of the proposition follows by similar arguments. On the one hand, it can be shown that, there is a fixed constant $0 < \bar{K} < +\infty$ such that $\|\hat{U}(\cdot) - U(\cdot)\|_{\infty,[0,\hat{\lambda}^{-1}(\hat{r})]} \le \bar{K}\|\hat{\lambda}(\cdot) - \lambda_0(\cdot)\|_{\infty,[0,\hat{\lambda}^{-1}(\hat{r})]}$ w.p.a.1. On the other hand, if $\hat{\lambda}^{-1}(\hat{r}) < y \le \lambda_0^{-1}(\bar{r})$, then

$$\begin{aligned}
|\hat{U}(y) - U(y)| &= 1 - U(y) \\
&\le U[\lambda_0^{-1}(\bar{r})] - U[\hat{\lambda}^{-1}(\bar{r})] \\
&\le \left[\sup_{y \in [0, \lambda_0^{-1}(\bar{r})]} U'(y)\right] \|\hat{\lambda}^{-1}(\cdot) - \lambda_0^{-1}(\cdot)\|_{\infty,[0,\bar{r}]}.
\end{aligned}$$

Proceeding in a similar manner, these result can be extended to the derivatives of $\hat{U}(\cdot)$.

1.A.2.5 Proof of Proposition 1.5.2

1. After combining Lemma 1.4.1 with Theorem 1.4.1, we easily obtain

$$\|\tilde{\upsilon}_j(\cdot) - \upsilon(\cdot)\|_{\infty,[0,1]} = O_P\left(h_b^{1/[2c(0)]}\right).$$

For the rest of the proof, the approach is similar to that of [MS12]; specifically, see the proof of Lemma 1.(g)-(h) on pages 355-356.

2. From Lemma 1.(f) of [MS12], we have that

$$\|\tilde{g}'(\cdot|i_j, x) - g'(\cdot|i_j, x)\|_{\infty,\mathcal{B}} = O_P\left(\left[\frac{\log(N)}{N h_f^{d+3}}\right]^{1/2} + h_f^R\right).$$

where $\mathcal{B}$ is the range of the bidding function $s(\cdot;i_j)$ over $\mathcal{C}$, namely, $\mathcal{B} = s(\mathcal{C};i_j)$. In order to complete the proof, observe that each component of $\hat{f}_j(\cdot)$ converges uniformly at a rate of $h_b^{1/[2c(1)]}$, or faster.

3. Let $b = s(v;i_j)$ denote the optimal bid for the valuation $v \in (\underline{v}, \bar{v})$, or equivalently, $b = b_j[F(v)]$. From Lemma 2 of [MS12], the rest of the proof follows by standard arguments.

1.A.3 Proofs of Auxiliary Lemmas

Here, I provide detailed proofs of the auxiliary lemmas employed in subsections 1.A.1-1.A.2.

1.A.3.1 Proof of Lemma 1.A.1

Let $q \in [1, +\infty)$. From Theorem 1 of [Gab67],

$$\|\nabla^r f - \nabla^r g\|_{\infty,[0,\bar{r}]} \leq K_1 \left(\delta^{-r-(1/q)} \|f - g\|_{q,[0,\bar{r}]} + \delta^{R+1-r} 2\bar{H}\right)$$

for any $f, g \in \mathscr{H}_R$, $r \in \{0, 1, \ldots, R\}$, and $\delta \in (0, \bar{r})$; where $K_1 < +\infty$ is independent of δ, $\bar{r}$, f, and g. Since both $\|\nabla^{R+1} f\|_{\infty,[0,\bar{r}]}$, $\|\nabla^{R+1} g\|_{\infty,[0,\bar{r}]} < \bar{H}$ and $[0, \bar{r}]$ is bounded, there is a positive finite constant K_2 (which is independent of δ, f, and g) such that $\|f - g\|_{q,[0,\bar{r}]} < K_2$ for all $f, g \in \mathscr{H}_R$. Then, the desired result emerges after choosing $\delta = \bar{r}(\|f - g\|_{q,[0,\bar{r}]}/K_2)^{1/[R+1+(1/q)]} < \bar{r}$ and $q = 1$.

1.A.3.2 Proof of Lemma 1.A.2

First, recall from Definition 3 that $g_j(\cdot)$ has R continuous derivatives on its support $[\underline{b}, \bar{b}_j]$, therefore $b_j(\cdot)$ has $R+1$ continuous derivatives on $[0,1]$. Second, since $\nabla^{R+1} g_j(\cdot)$ is continuous on $(\underline{b}, \bar{b}_j]$, $\nabla^{R+2} b_j(\cdot)$ is also continuous on $(0,1]$. Third, as

$g_j(\cdot)$ is continuous and bounded away from zero, we have that

$$b_j'(\alpha) = \frac{1}{g_j[b_j(\alpha)]} \geq \frac{1}{\inf\{g_j(b) : b \in [\underline{b}, \bar{b}_j]\}} > 0$$

for any $\alpha \in [0,1]$.

The second statement is essentially derived from Definition 4, condition 2. Exploiting the compatibility conditions (a)-(b), define $\upsilon(\alpha) \equiv \xi_1[b_1(\alpha)] = \xi_2[b_2(\alpha)]$ for $\alpha \in [0,1]$. Observe that $\upsilon(0) = \underline{b}$, as well as, $\upsilon'(0) > 0$ because $b_j'(\cdot)$ is bounded away from zero and $\xi_j'(\underline{b}) = 1 + 1/[\lambda_0'(0)(i_j - 1)] > 0$. Since $\xi_j(\cdot)$ is strictly increasing, it follows that $b_j(\alpha) = \xi_j^{-1}[\upsilon(\alpha)]$, and therefore, $b_j'(0) = (\xi_j^{-1})'(\underline{b})\upsilon'(0)$. At the same time, we also have that $(\xi_j^{-1})'(\underline{b})$ equals $1/\xi_j'(\underline{b}) = \lambda_0'(0)(i_j - 1)/[\lambda_0'(0)(i_j - 1) + 1]$, which immediately implies

$$\frac{b_1'(0)}{i_1 - 1} = \frac{\lambda_0'(0)\upsilon'(0)}{\lambda_0'(0)(i_1 - 1) + 1} > \frac{\lambda_0'(0)\upsilon'(0)}{\lambda_0'(0)(i_2 - 1) + 1} = \frac{b_2'(0)}{i_2 - 1}.$$

The strict inequality is due to $\lambda_0'(0) > 0$ and $\upsilon'(0) > 0$.

1.A.3.3 Proof of Lemma 1.A.3

For the first part, considering the construction of $\bar{a}$, choose $\tilde{N}$ sufficiently large so that $\bar{L}K_{\tilde{N}}^{-1/2}$ is smaller than $\min\{b_2'(0)/[2(i_2-1)], \bar{\alpha}/4\}$, and consequently, smaller than $b_1'(0)/[2(i_1 - 1)]$ (see Lemma 1.A.2). Now choose any $n \geq \tilde{N}$ and $(j, \alpha) \in \{1,2\} \times (0, \bar{\alpha}/2]$, and then, consider k so that $\alpha \in (a_k^{(n)}, a_{k+1}^{(n)}]$, where $0 \leq k \leq \lceil K_n^{-1/2} \rceil$. Since $a_k^{(n)} < \bar{\alpha}/2$ and $\lceil K_n^{1/2} \rceil^{-1} \leq K_{\tilde{N}}^{-1/2} < \bar{\alpha}/4$, it follows that $a_{k+1}^{(n)} < \bar{\alpha}$ due to the length of $(a_k^{(n)}, a_{k+1}^{(n)}]$. By construction of $\bar{\alpha}$ and $\tilde{N}$, $\inf\{|R_j'(a)| : a \in (a_k^{(n)}, a_{k+1}^{(n)}]\} \geq b_j'(0)/[2(i_j - 1)] > \bar{L}K_{\tilde{N}}^{-1/2} \geq \bar{L}K_n^{-1/2}$, hence $R_j^{(n)}(\alpha) = R_j(\alpha)$.

For the second part, consider the case $j = 1$. Pick any $n \geq \tilde{N}$ and $\alpha \in [0,1]$ such that $\alpha \in (a_k^{(n)}, a_{k+1}^{(n)}]$ with $a_{k+1}^{(n)} > \bar{\alpha}/2$; otherwise, $a_{k+1}^{(n)} \leq \bar{\alpha}/2$, then $R_1^{(n)}(\alpha) = R_1(\alpha)$ due to the previous discussion. Suppose further that $\inf\{|R_1'(a)| : a \in (a_k^{(n)}, a_{k+1}^{(n)}]\} \leq \bar{L}K_n^{-1/2}$; otherwise, $R_1^{(n)}(\alpha) = R_1(\alpha)$ by construction of $R_1^{(n)}(\cdot)$.

Following standard arguments, we have that

$$\begin{aligned} |R_1^{(n)}(\alpha) - R_1(\alpha)| &= |R_1(a_k^{(n)}) + 2\bar{L}K_n^{-1/2}(\alpha - a_k^{(n)}) - R_1(\alpha)| \\ &\leq R_1'(\underline{a})|\alpha - a_k^{(n)}| + 2\bar{L}K_n^{-1/2}|\alpha - a_k^{(n)}| \leq 4\bar{L}K_n^{-1}, \end{aligned}$$

for some $\underline{a} \in (a_k^{(n)}, a_{k+1}^{(n)}]$, which satisfies $R_1'(\underline{a}) \leq 2\bar{L}K_n^{-1/2}$ because $a_k^{(n)} > \bar{\alpha}/5$ and $R_1'(\cdot)$ is Lipschitz continuous on $[\bar{\alpha}/5, 1]$ with constant $\bar{L}$. The first equality follows by definition of $R_1^{(n)}(\alpha)$. The second inequality follows by a standard triangular inequality and the fact that $R_1(\cdot)$ is continuously differentiable on $[0,1]$, and particularly, on $[a_k^{(n)}, a_{k+1}^{(n)}]$. The last inequality follows because $R_1'(\underline{a}) \leq 2\bar{L}K_n^{-1/2}$ and $|\alpha - a_k^{(n)}| \leq K_n^{-1/2}$. Symmetrically, we can show the same result for $j = 2$.

1.A.3.4 Proof of Lemma 1.A.4

1. We only prove that $R_1^{(n)}(\alpha) \geq R_1(\alpha)$, symmetrically, it follows that $R_2(\alpha) \geq R_2^{(n)}(\alpha)$. Consider again any $\alpha \in (a_k^{(n)}, a_{k+1}^{(n)}]$ and assume $a_{k+1}^{(n)} > \bar{\alpha}/2$ as well as $\inf\{|R_1'(a)| : a \in (a_k^{(n)}, a_{k+1}^{(n)}]\} \leq \bar{L}K_n^{-1/2}$; otherwise, $R_1^{(n)}(\alpha) = R_1(\alpha)$ by construction of $R_1^{(n)}(\cdot)$. After applying the mean value theorem on $R_1(\cdot)$, we obtain that $R_1(\alpha) = R_1(a_k^{(n)}) + R_1'(\bar{a})(\alpha - a_k^{(n)})$ for some $\bar{a} \in (a_k^{(n)}, a_{k+1}^{(n)}]$. Then, it must be $|R_1'(\bar{a})| \leq 2\bar{L}K_n^{-1/2}$ because $a_k^{(n)} > \bar{\alpha}/5$ and $R_1'(\cdot)$ is Lipschitz continuous on $[\alpha/5, 1]$ with constant $\bar{L}$. Therefore, it follows $R_1(\alpha) \leq R_1^{(n)}(\alpha)$ by definition of $R_1^{(n)}(\alpha)$.

2. There exists $a \in [0,1]$ that satisfies $R_1(a) = \bar{r}$ and $R_1'(a) \geq 0$.[7] As a consequence, $\bar{r} = R_1(a) \leq R_1^{(n)}(a) \leq R_1^{(n)}(a_k^{(n)})$ where k is taken to be $(a_{k-1}^{(n)}, a_k^{(n)}] \ni a$. Note that the last inequality follows because $R_1'(a) \geq 0$, and as a result, $R_1^{(n)}(\cdot)$ must be (strictly) increasing on $(a_{k-1}^{(n)}, a_k^{(n)}]$.

3. We immediately rule out $R_1^{(n)}(a_{l-1}^{(n)}) \geq R_1^{(n)}(a_l^{(n)})$ because l is a minimum. By definition, we have $u \leq R_1^{(n)}(a_l^{(n)})$, and again since l is a minimum, it must be the case $R_1^{(n)}(a_{l-1}^{(n)}) < u$.

[7]If $a = 1$, we simply define $R_1'(a) = \lim_{h\uparrow 0}[R_1(a+h) - R_1(a)]/h$.

4. Since both $R_1(\cdot)$ and $R_1^{(n)}(\cdot)$ are continuous on $(a_{k-1}^{(n)}, a_k^{(n)}]$, observe that

$$\lim_{\alpha \downarrow a_{k-1}^{(n)}} R_1^{(n)}(\alpha) = R_1(a_{k-1}^{(n)}) \leq R_1^{(n)}(a_{k-1}^{(n)}) < u \leq R_1^{(n)}(a_k^{(n)}),$$

where the first inequality follows by item 1. At the same time, there exists $\tilde{\alpha} \in (a_{k-1}, a_k]$ such that $R_1^{(n)}(\tilde{\alpha}) = u$. Moreover, since $R_1^{(n)}(\cdot)$ is strictly monotone on $(a_{k-1}, a_k]$ and $R_1^{(n)}(\tilde{\alpha}) = u \leq R_1^{(n)}(a_k^{(n)})$, $R_1^{(n)}(\cdot)$ must be strictly increasing on $(a_{k-1}, a_k]$.

5. By definition of l (third item), $u \in (R_1^{(n)}(a_{l-1}^{(n)}), R_1^{(n)}(a_l^{(n)})]$. By the fourth item, there is $\tilde{\alpha} \in (a_{l-1}^{(n)}, a_l^{(n)}]$ such that $R_1^{(n)}(a_{l-1}^{(n)}) < u = R_1^{(n)}(\tilde{\alpha}) \leq R_1^{(n)}(\alpha)$. We next show that $\tilde{\alpha} \leq \alpha$, by contradiction, assume $\tilde{\alpha} > \alpha$. In such case, we must have $\alpha \leq a_{l-1}^{(n)}$; otherwise, if $a_l^{(n)} \geq \tilde{\alpha} > \alpha > a_{l-1}^{(n)}$, then $R_1^{(n)}(\alpha) < R_1^{(n)}(\tilde{\alpha})$ because $R_1^{(n)}(\cdot)$ is strictly increasing on $(a_{l-1}^{(n)}, a_l^{(n)}]$. To complete the proof, consider the interval that contains α, that is $(a_{j-1}^{(n)}, a_j^{(n)}] \ni \alpha$ being $1 \leq j \leq l-1$. On the one hand, if $R_1^{(n)}(\cdot)$ is strictly increasing on $(a_{j-1}^{(n)}, a_j^{(n)}]$, it follows $u \leq R_1^{(n)}(\alpha) \leq R_1^{(n)}(a_j^{(n)})$. On the other hand, if $R_1^{(n)}(\cdot)$ is strictly decreasing on $(a_{j-1}^{(n)}, a_j^{(n)}]$, then $R_1(\cdot)$ must be necessarily decreasing on $[a_{j-1}^{(n)}, a_j^{(n)}]$, and as a result,

$$u \leq R_1^{(n)}(\alpha) = R_1(\alpha) \leq R_1(a_{j-1}^{(n)}) \leq R_1^{(n)}(a_{j-1}^{(n)}).$$

Both cases contradict the definition of l (third item) because $j \leq l-1$.

1.A.3.5 Proof of Lemma 1.A.5

For the first part, pick any (u, u') such that $0 \leq u < u' \leq \bar{r}$. Since $R_1^{(n)}[\alpha_0^{(n)}(u)] = u$ and $R_1^{(n)}[\alpha_0^{(n)}(u')] = u'$, we immediately rule out $\alpha_0^{(n)}(u) = \alpha_0^{(n)}(u')$. By contradiction, now suppose $\alpha_0^{(n)}(u') < \alpha_0^{(n)}(u)$. Then, $u \in (0, R_1^{(n)}[\alpha_0^{(n)}(u)]]$, and as a result of $u < u'$, $u \in (0, R_1^{(n)}[\alpha_0^{(n)}(u')]]$. By Lemma 1.A.4.5, there is $\tilde{\alpha}$ such that $0 < \tilde{\alpha} \leq \alpha_0^{(n)}(u') < \alpha_0^{(n)}(u)$ and $R_1^{(n)}(\tilde{\alpha}) = u$; this contradicts the definition of $\alpha_0^{(n)}(u)$.

For the second part, from Lemma 1.A.4.1, $R_2^{(n)}[\alpha_t^{(n)}(u)] < R_1^{(n)}[\alpha_t^{(n)}(u)]$. Then, by construction of $\alpha_{t+1}^{(n)}(u)$ and again Lemma 1.A.4.5, it follows $\alpha_{t+1}^{(n)}(u) < \alpha_t^{(n)}(u)$.

1.A.3.6 Proof of Lemma 1.A.6

This proof employs an inductive argument on $t \in \mathbb{N}_0$. We only consider cases in which $V_{(l_0)}$ or $V_{(l_0,\ldots,l_{t-1},l_t)}$ are nonempty, otherwise, the desired results follow trivially.

Case t =0

1. Since $(0,\bar{r})$ is an open interval, $V_{(l_0)}$ is also an open interval. Now pick any $u \in V_{(l_0)}$. Since $R_1^{(n)}(a_{l_0-1}^{(n)}) < R_1^{(n)}(a_{l_o}^{(n)})$ and $u \in (R_1^{(n)}(a_{l_0-1}^{(n)}), R_1^{(n)}(a_{l_0}^{(n)}))$, by Lemma 1.A.4.4, there is $\tilde{\alpha} \in (a_{l_0-1}^{(n)}, a_{l_o}^{(n)}]$ such that $R_1^{(n)}(\tilde{\alpha}) = u$. As $\alpha_0^{(n)}(u)$ is a minimum, we automatically rule out the case $\alpha_0^{(n)}(u) > a_{l_0}^{(n)}$, and as $u = R_1^{(n)}[\alpha_0^{(n)}(u)]$, it is obvious that $\alpha_0^{(n)}(u) \neq a_{l_0}^{(n)}$. By contradiction, suppose $\alpha_0^{(n)}(u) \in (a_{j-1}^{(n)}, a_j^{(n)}]$ for some $1 \leq j \leq l_0 - 1$ and recall that $R_1^{(n)}(\cdot)$ is strictly monotone on $(a_{j-1}^{(n)}, a_j^{(n)}]$. On the one hand, if $R_1^{(n)}(\cdot)$ is strictly increasing on $(a_{j-1}^{(n)}, a_j^{(n)}]$, it follows

$$u = R_1^{(n)}[\alpha_0^{(n)}(u)] \leq R_1^{(n)}(a_j^{(n)}) \leq \max_{j=0,\ldots,l_0-1} R_1^{(n)}(a_j^{(n)}).$$

On the other hand, if $R_1^{(n)}(\cdot)$ is strictly decreasing on $(a_{j-1}^{(n)}, a_j^{(n)}]$, then $R_1^{(n)}(\cdot) = R_1(\cdot)$ and therefore $R_1(\cdot)$ is decreasing on $[a_{j-1}^{(n)}, a_j^{(n)}]$ as well; as a result,

$$u = R_1^{(n)}[\alpha_0^{(n)}(u)] = R_1[\alpha_0^{(n)}(u)] \leq R_1(a_{j-1}^{(n)}) \leq R_1^{(n)}(a_{j-1}^{(n)}) \leq \max_{j=0,\ldots,l_0-1} R_1^{(n)}(a_j^{(n)}).$$

Both cases clearly contradict the definition of $V_{(l_0)}$.

2. For the first part, consider the function $f(\cdot,\cdot) \equiv R_1^{(n)}(\cdot) - \cdot$ defined over the open rectangle $(a_{l_0-1}^{(n)}, a_{l_0}^{(n)}) \times V_{(l_0)}$ and given by $f(\alpha, u) = R_1^{(n)}(\alpha) - u$ where

$(\alpha, u) \in (a_{l_0-1}^{(n)}, a_{l_0}^{(n)}) \times V_{(l_0)}$. From above result, $\alpha_0^{(n)}(\cdot)$ can be implicitly defined on $V_{(l_0)}$ as follows: $f[\alpha_0^{(n)}(\cdot), \cdot] = 0$. Since $f(\cdot, \cdot)$ is continuously differentiable on its domain $(a_{l_0-1}^{(n)}, a_{l_0}^{(n)}) \times V_{(l_0)}$, by the Implicit Function Theorem it follows immediately that $\alpha_0^{(n)}(\cdot)$ is continuously differentiable on $V_{(l_0)}$. Trivially, $R_2^{(n)}[\alpha_0^{(n)}(\cdot)]$ is also continuously differentiable on $V_{(l_0)}$ because $R_2^{(n)}(\cdot)$ is continuously differentiable on $(a_{l_0-1}^{(n)}, a_{l_0}^{(n)})$. Finally, the inequality $\alpha_0^{(n)\prime}(u) = 1/R_1^{(n)\prime}[\alpha_0^{(n)}(u)] \le K_n^{1/2}$ follows by construction of $R_1^{(n)}(\cdot)$ and because $\bar{L} \ge 1$.

3. By Lemma 1.A.4.2, there is j such that $R_1^{(n)}(a_j^{(n)}) > \bar{r}$, so

$$(0, \bar{r}) = \Big(\cup_{1 \le l \le \lceil K_n^{1/2} \rceil} V_l\Big) \dot{\bigcup} \Big(\cup_{0 \le l \le \lceil K_n^{1/2} \rceil} \{u \in (0, \bar{r}) : u = R_1^{(n)}(a_l^{(n)})\}\Big).$$

To complete the proof, note that the first union $\cup_{1 \le l \le \lceil K_n^{1/2} \rceil} V_l$ is disjoint, while the second one $\cup_{0 \le l \le \lceil K_n^{1/2} \rceil} \{u \in (0, \bar{r}) : u = R_1^{(n)}(a_l^{(n)})\}$ is in fact a finite set, so its Lebesgue measure equals zero.

Suppose that the statement holds for $t-1$, being $t \in \mathbb{N}$

1. Using an inductive argument, $V_{(l_0, \ldots, l_t)}$ is an open interval because $V_{(l_0, \ldots, l_{t-1})}$ is an open interval and $R_2^{(n)}[\alpha_{t-1}^{(n)}(\cdot)]$ is strictly monotone and continuous on $V_{(l_0, \ldots, l_{t-1})}$. For the rest of the proof, just follow the same steps as in the case $t = 0$.

2. For the first part, now consider the function $R_1^{(n)}(\cdot) - R_2^{(n)}[\alpha_{t-1}^{(n)}(\cdot)]$ defined over the open rectangle $(a_{l_t-1}^{(n)}, a_{l_t}^{(n)}) \times V_{(l_0, \ldots, l_t)}$. For the second part, when $u \in V_{(l_0, \ldots, l_{t-1}, l_t)}$, just note

$$\alpha_t^{(n)\prime}(u) = \frac{R_2^{(n)\prime}[\alpha_{t-1}^{(n)}(u)]\alpha_{t-1}^{(n)'}(u)}{R_1^{(n)\prime}[\alpha_t^{(n)}(u)]} \le K_n^{(t/2)+(1/2)}$$

because $V_{(l_0, \ldots, l_{t-1}, l_t)} \subseteq V_{(l_0, \ldots, l_{t-1})}$ and $|\alpha_{t-1}^{(n)'}(u)| \le K_n^{t/2}$ (inductive argument).

3. Observe that

$$\sum_{(l_0,l_1,\dots,l_t)} |V_{(l_0,\dots,l_{t-1},l_t)}| = \sum_{(l_0,l_1,\dots,l_{t-1})} \sum_{l_t=1}^{\lceil K_n^{1/2}\rceil} |V_{(l_0,\dots,l_{t-1},l_t)}|,$$

where the support of the sum on the left-hand side runs over $\{(l_0,l_1,\dots,l_t) \in \mathbb{N}^{t+1} : l_j \le \lceil K_n^{1/2}\rceil, j = 0,\dots,t\}$, while the support of the first sum on right-hand side runs over $\{(l_0,l_1,\dots,l_{t-1}) \in \mathbb{N}^t : l_j \le \lceil K_n^{1/2}\rceil, j = 0,\dots,t-1\}$. It can be easily seen that

$$\sum_{l_t=1}^{\lceil K_n^{1/2}\rceil} |V_{(l_0,\dots,l_{t-1},l_t)}| = |V_{(l_0,\dots,l_{t-1})}|,$$

then the desired result follows from the inductive argument $\sum_{(l_0,l_1,\dots,l_{t-1})} |V_{(l_0,\dots,l_{t-1})}| = \bar{r}$.

1.A.3.7 Proof of Lemma 1.A.7

Consider any $t \in \mathbb{N}_0$ and $h \in \mathbb{R}$ such that $0 < |h| \le c\lceil K_n^{1/2}\rceil^{-(t+5)}$. From Lemma 1.A.6.3, and since $\alpha_t^{(n)}(\cdot)$ is bounded above by one, we have

$$\begin{aligned}
\int_{\mathbb{R}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du &\le \int_{-|h|}^{\bar{r}+|h|} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \\
&\le 4|h| + \int_0^{\bar{r}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \\
&\le 4|h| + \sum_{(l_0,l_1,\dots,l_t)} \int_{V_{(l_0,\dots,l_t)}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du,
\end{aligned}$$

where the support of the sum runs over $\{(l_0,l_1,\dots,l_t) \in \mathbb{N}^{t+1} : l_j \le \lceil K_n^{1/2}\rceil, j = 0,1,\dots,t\}$. Then, for each summand, it can be easily shown that

$$\begin{aligned}
&\int_{V_{(l_0,\dots,l_t)}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \le \\
&\int_{\underline{v}_{(l_0,l_1,\dots,l_t)}}^{\underline{v}_{(l_0,l_1,\dots,l_t)}+|h|} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \\
&+ \int_{\underline{v}_{(l_0,l_1,\dots,l_t)}+|h|}^{\bar{v}_{(l_0,l_1,\dots,l_t)}-|h|} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \\
&+ \int_{\bar{v}_{(l_0,l_1,\dots,l_t)}-|h|}^{\bar{v}_{(l_0,l_1,\dots,l_t)}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du,
\end{aligned}$$

where $\underline{v}_{(l_0,l_1,\dots,l_t)}$ and $\bar{v}_{(l_0,l_1,\dots,l_t)}$ denote the infimum and supremum of the open interval $V_{(l_0,l_1,\dots,l_t)}$, respectively; here, the notation for the second integral is $\int_a^b \cdot = 0$

whenever $a \geq b$. Since $\alpha_t^{(n)}(\cdot)$ is bounded above by one and continuously differentiable on $V_{(l_0,l_1,\ldots,l_t)}$, by the second part of Lemma 1.A.6, we have that

$$\int_{V_{(l_0,\ldots,l_t)}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du \leq 4|h| + |V_{(l_0,\ldots,l_t)}||h|K_n^{(t+1)/2},$$

and then from Lemma 1.A.6.3, it follows

$$\begin{aligned}\int_{\mathbb{R}} |\alpha_t^{(n)}(u+h) - \alpha_t^{(n)}(u)|du &\leq 8|h| + \lceil K_n^{1/2}\rceil^{t+1}4|h| + \bar{r}|h|K_n^{(t+1)/2} \\ &\leq 8|h| + (c4 + c\bar{r})K_n^{-2} \\ &\leq c(2^4 + \bar{r})K_n^{-2}.\end{aligned}$$

1.A.3.8 Proof of Lemma 1.A.8

For $n > \tilde{N}$ and $\alpha \in [0,1]$, define the functions $\delta^{(n)}(\alpha) \equiv R_1^{(n)}(\alpha) - R_2^{(n)}(\alpha)$ and $\delta(\alpha) = R_1(\alpha) - R_2(\alpha) \geq 0$. Notice that $\delta^{(n)}(\alpha) \geq \delta(\alpha)$ for all $\alpha \in [0,1]$, as well as, $\delta^{(n)}(\alpha) = \delta(\alpha) = 0$ if and only if $\alpha = 0$; see Lemma 1.A.4. When $\alpha \in [0, \bar{\alpha}/2]$, by Lemma 1.A.3 we have that

$$\delta^{(n)}(\alpha) = \delta(\alpha) = R_1(\alpha) - R_2(\alpha) = \left[\frac{b_1'(\alpha)}{i_1 - 1} - \frac{b_2'(\alpha)}{i_2 - 1}\right]\alpha \equiv \tilde{\delta}(\alpha)\alpha, \tag{1.20}$$

thus, $\delta'(\alpha) = \tilde{\delta}'(\alpha)\alpha + \tilde{\delta}(\alpha)$. From Lemma 1.A.2 and as $\tilde{\delta}'(\cdot)$ is bounded, pick first $\tilde{a} \in (0, \bar{\alpha}/2)$ sufficiently small so that $\min\{\tilde{\delta}(\cdot), \delta'(\cdot)\} > 0$ on $[0, \tilde{a}]$; obviously, $\delta(\cdot)$ becomes strictly increasing on $[0, \tilde{a}]$. Second, note that $\min_{\alpha \in [\tilde{a},1]} \delta(\alpha)$ is strictly greater than zero and then choose $a^* \in (0, \tilde{a}]$ such that $\delta(a^*) < \min_{\alpha \in [\tilde{a},1]} \delta(\alpha)$. Third, considering that $(T_m)_m$ is divergent and also $\min_{\alpha \in [a^*,1]} \delta(\alpha) > 0$, pick $N^* > \tilde{N}$ such that $T_{N^*} \min_{\alpha \in [a^*,1]} \delta(\alpha) > \bar{r}$.

Before proceeding, we remark that both $\tilde{a}$ and a^* are independent of n because $R_j^{(n)}(\cdot) = R_j(\cdot)$ on $[0, \bar{\alpha}/2]$. Note also that $\min_{\alpha \in [0,a^*]} \tilde{\delta}(\alpha) > 0$ is independent of n because $a^* < \tilde{a} < \bar{\alpha}/2$. Next we show that $T_m \|\alpha_{0,T_m}^{(n)}(\cdot)\|_{\infty,[0,\bar{r}]} \leq \bar{K}$ for all $m, n \geq N^*$, where $\bar{K}$ equals $\bar{r}/\min_{\alpha \in [0,a^*]} \tilde{\delta}(\alpha)$. In order to do so, choose any arbitrary $m, n \geq N^*$ and $u \in [0, \bar{r}]$.

Observe first that $\alpha_{0,T_{N^*}}^{(n)}(u) < \alpha^*$. Suppose not, that is, $\alpha_{0,T_{N^*}}^{(n)}(u) \geq a^*$. On the one hand, $\alpha_t^{(n)}(u) \geq \alpha^*$ for all $t \leq T_{N^*}$ because $\alpha_t^{(n)}(u)$ is decreasing in t when both n and u are fixed; see Lemma 1.A.5. On the other hand,

$$\begin{aligned}
\bar{r} &\geq \tilde{R}_1^{(n)}[\alpha_0^{(n)}(u)] - \tilde{R}_2^{(n)}[\alpha_{T_{N^*}}^{(n)}(u)] \\
&= \{\tilde{R}_1^{(n)}[\alpha_0^{(n)}(u)] - \tilde{R}_2^{(n)}[\alpha_0^{(n)}(u)]\} + \sum_{t=1}^{T_{N^*}} \{\tilde{R}_1^{(n)}[\alpha_t^{(n)}(u)] - \tilde{R}_2^{(n)}[\alpha_t^{(n)}(u)]\} \\
&\geq \sum_{t=0}^{T_{N^*}} \left\{R_1[\alpha_t^{(n)}(u)] - R_2[\alpha_t^{(n)}(u)]\right\} \geq (T_{N^*}+1) \min_{\alpha \in [a^*,1]} \delta^{(n)}(\alpha) \\
&\geq T_{N^*} \min_{\alpha \in [a^*,1]} \delta(\alpha),
\end{aligned}$$

which contradicts the construction of N^*. As $R_1^{(n)}[\alpha_0^{(n)}(u)] = u \leq \bar{r}$ and $\tilde{R}_2^{(n)}(\cdot) \geq 0$ (Lemma 1.A.3), the first inequality is trivial. The second equality is the key step and follows inductively from $R_1^{(n)}[\alpha_t^{(n)}(u)] - R_2^{(n)}[\alpha_{t-1}^{(n)}(u)] = 0$ for all $t \in \mathbb{N}$. The third inequality is due to Lemma 1.A.4.1, while the fourth one can be derived from $\alpha_t^{(n)}(u) \geq a^*$ for all $t \leq T_{N^*}$.

Before proceeding, note that $\alpha_t^{(n)}(u) \leq \alpha_{T_{N^*}}^{(n)}(u) < a^*$ for all $t \geq T_{N^*}$. Since $(T_m)_m$ is increasing and divergent, when $m \geq N^*$, we have that

$$\begin{aligned}
\bar{r} &\geq R_1^{(n)}[\alpha_0^{(n)}(u)] - R_2^{(n)}[\alpha_{T_m}^{(n)}(u)] \\
&= \sum_{t=0}^{T_m} \delta^{(n)}[\alpha_t^{(n)}(u)] \\
&\geq T_m \delta[\alpha_{T_m}^{(n)}(u)] \\
&= T_m \tilde{\delta}[\alpha_{T_m}^{(n)}(u)] \alpha_{T_m}^{(n)}(u) \geq \left[\min_{\alpha \in [0,a^*]} \tilde{\delta}(\alpha)\right] T_m \alpha_{T_m}^{(n)}(u).
\end{aligned}$$

The second equality follows by definition of $\delta^{(n)}(\cdot)$ and previous arguments, that is, $R_1^{(n)}[\alpha_t^{(n)}(u)] = R_2^{(n)}[\alpha_{t-1}^{(n)}(u)]$. In order to establish the third inequality, we next consider two possible cases: $\alpha_t^{(n)}(u) \geq \tilde{a}$ and $\alpha_t^{(n)}(u) < \tilde{a}$. On the one hand, if $\alpha_t^{(n)}(u) \geq \tilde{a}$, then

$$\delta^{(n)}[\alpha_t^{(n)}(u)] \geq \min_{\alpha \in [\tilde{a},1]} \delta^{(n)}(\alpha) \geq \min_{\alpha \in [\tilde{a},1]} \delta(\alpha) > \delta(a^*) \geq \delta[\alpha_{T_{N^*}}^{(n)}(u)] \geq \delta[\alpha_{T_m}^{(n)}(u)]$$

because $\delta(\cdot)$ is strictly increasing on $[0,\tilde{a}]$, and at the same time, $\tilde{a} > a^* > \alpha_{T_{N^*}}^{(n)}(u) \geq \alpha_{T_m}^{(n)}(u)$; on the other hand, if $\alpha_t^{(n)}(u) < \tilde{a} < \bar{\alpha}/2$, it follows easily that

$\delta^{(n)}[\alpha^{(n)}_{T_m}(u)] = \delta[\alpha^{(n)}_{T_m}(u)] \leq \delta[\alpha^{(n)}_t(u)]$ due to the support of the sum $0 \leq t \leq T_m$. The fourth equality follows simply by the definition in (1.20). The last inequality is a consequence of $\alpha_{T_m}(u) \leq \alpha_{T_{N^*}}(u) < a^*$, and to complete the proof, recall that $\min_{\alpha\in[0,a^*]} \tilde{\delta}(\alpha)$ is strictly positive and independent of n.

1.A.3.9 Proof of Lemma 1.A.9

Items 1-4 are well-known results, hence their proofs are omitted; see [GPV00], as well as, [MS12]. The fifth item is simply a particular case of Proposition 2 in [GPV00], so in what follows, we focus only on 6-8. Before proceeding, note that under Assumptions 1-3, in items 1-3 the rate of convergence inside $O_P(\cdot)$ becomes h_G^{R+1}, whereas in item 4 the corresponding rate is h_g^R.

6. The first part is based on the proof of Lemma 1.(d) in [MS12]. Under Assumption 3, we have $\|\hat{G}(\cdot|i_j,x) - G(\cdot|i_j,x)\|_{\infty,[\underline{b},\bar{b}_j]} = O_P(h_G^{R+1})$, and as a result,

$$\mathbb{P}[\hat{b}_j(h_b) \leq \underline{b}] = \mathbb{P}\left[\inf_{b\in\mathbb{R}_{\geq 0}} \{\hat{G}(b|i_j,x) \geq h_b\} \leq \underline{b}\right] = \mathbb{P}[\hat{G}(\underline{b}|i_j,x) \geq h_b] = o(1);$$

the last equality follows from item 3 and the fact $h_b/h_G^{R+1} \to +\infty$ as $N \to +\infty$. Symmetrically, $\mathbb{P}[\hat{b}_j(1-h_b) \geq \bar{b}_j] = o(1)$, and as a result, $\underline{b} < \hat{b}_j(h_b) \leq \hat{b}_j(1-h_b) < \bar{b}_j$ w.p.a.1. The rest of the proof follows exactly by the same arguments of eqs. (40)-(48) in [MS12]. The second part follows immediately from item 5 because $b_j(\cdot)$ is Lipschitz continuous on $[0,1]$, and $h_\partial < h_b$ when N is large enough.

7. I first show that $\underline{b} + h_g < \hat{b}_j(h_b)$ w.p.a.1. On the one hand, for N sufficiently large

$$\begin{aligned}
\mathbb{P}[|\hat{b}_j(h_b) - b_j(h_b)| \leq h_g] &= \mathbb{P}[|\underline{b} + h_b b_j'(\tilde{h}) - \hat{b}_j(h_b)| \leq h_g] \\
&\leq \mathbb{P}\left[\underline{b} + h_b\left(\inf_{\alpha\in[0,1]} b_j'(\alpha)\right) - h_g \leq \hat{b}_j(h_b)\right] \\
&\leq \mathbb{P}[\underline{b} + h_g < \hat{b}_j(h_b)],
\end{aligned}$$

where $\tilde{h} \in [0, h_b]$ and the last inequality follows from $h_b/h_g \to +\infty$ as $N \to +\infty$ (Assumption 3). On the other hand, we already know from the previous item that the left hand side converges to one, hence $\mathbb{P}[\underline{b} + h_g < \hat{b}_j(h_b)] \to 1$. Symmetrically, it can be shown that $\mathbb{P}[\hat{b}_j(1 - h_b) < \bar{b}_j - h_g] \to 1$, and as a result, $\underline{b} + h_g < \hat{b}_j(h_b) \leq \hat{b}_j(1 - h_b) < \bar{b}_j - h_g$ w.p.a.1.

8. Recall that $g(\cdot|i_j, x)$ is bounded away from zero, so from items 4 and 7, it follows immediately that $g[\hat{b}_j(\cdot)|i_j, x]$ is bounded away from zero on $[h_b, 1 - h_b]$ w.p.a.1.

9. By standard arguments; see [CGP11].

10. From items 4, 7, and 9 above.

CHAPTER 2

On Tacit Collusion among Asymmetric Firms in Bertrand Competition

2.1 Introduction

The model of repeated Bertrand competition explains how firms may be able to collude and sustain a high price even when they produce identical goods. Thus it resolves so called "Bertrand paradox," which would arise in one-shot interaction, that firms lose any monopoly power and make no profit as soon as two firms are present in the market; see for example, [Tir88]. Since it is a simple and very convenient model, it has been used in numerous applied works.[1]

However, we still do not fully understand when and how collusion can be sustained except for the very special case where firms are symmetric. This assumption of symmetric firms is of course very strong and unrealistic; firms in general differ in various dimensions. What we think is particularly strong is the assumption of equal discounting. There are at least two reason to believe that future profit is discounted differently by different firms. First, some firms may be subject to a less favorable interest rate than others due to some kind of credit market imperfection. Second, even if the time preference is the same across firms, the time preferences of the managers who run those firms can be different. Some manager may discount future heavily if she expects to retire or be fired soon.

[1] There are many other ways to resolve "Bertrand paradox" such as introducing capacity constraints or differentiated demands.

Some manager's preference may be more in line with the preference of the firm if she may own more stocks (and stock options) of the firm.

The goal of this paper is to understand the nature of collusion in the repeated Bertrand competition model when firms are asymmetric, especially when different firms discount future profits in different way.[2]

We have two main results. First we identify the critical level of discount factor such that a collusive outcome can be sustained if and only if the *average discount factor within the lowest cost firms* is above the critical level. More generally, we show that the necessary and sufficient condition for sustaining a collusion at a certain price (or more) is that the average discount factor of all the firms whose marginal cost is below the price must be larger than $(n' - 1)/n'$, where n' is the number of such firms. A more patient firm is willing to give up more market shares to more impatient firms, whose incentive constraints are then relaxed. So the distribution of discounting rates matters in general. In our simple setting with homogeneous good, the mean of discounting rates among colluding firms determines the possibility of collusion.

Our second result is a characterization of all efficient (profit-maximizing) collusive equilibria when firms differ only in their discounting rates. In efficient equilibria, more impatient firms gain a larger share of the market at an earlier stage and more patient firms gain a larger share at a later stage. Such an intertemporal substitution of the market share is subjective to the incentive constraint: we cannot assign 0% share forever even to the most impatient firm. Hence the equilibrium outcome is not the first best.

Our characterization provides a totally new picture of collusion, which is radically different from the one among symmetric firms. First, the equilibrium market

[2] We assume that heterogeneous discounting rates are given exogenously. Of course, it would be interesting to think about a model in which they are endogenously determined for a variety of reasons. We think that our model with fixed heterogeneous discounting rates would open a possibility of building such a model.

share in any efficient collusive equilibrium changes over time. More specifically, the market share dynamics of each firm can be described by three phases. In the first phase, a firm has no share of the market, leaving the market to more impatient firms. In the second phase, the firm enters the market and gains all the rest after leaving more impatient firms the minimum amount of stationary market share, which correspond to the worst stationary collusive equilibrium market share for them. The final phase starts when a more patient firm enters the market. In the final phase, the firm's marker share drops to the level that corresponds to its worst stationary collusive equilibrium market share and stays there forever.

Secondly, our results deliver the unique prediction in the long run. As described above, the equilibrium market share for each firm, except for the most patient firm, converges to its worst stationary collusive equilibrium market share in any efficient collusive equilibrium. More precisely, every efficient collusive equilibrium converges to the unique stationary collusive equilibrium within finite time.[3]

We know that, with symmetric firms, there are many efficient stationary equilibria with different market shares because how to share the market is irrelevant for efficiency. With asymmetric discounting, however, efficiency imposes a sharp restriction on how the market should be allocated intertemporally. As a consequence, even though there are many efficient equilibria, the long run market share must be the same across all efficient equilibria.

From a more theoretical perspective, our results deliver new insights into the theory of repeated games with differential discounting. As reviewed briefly next, the major results for repeated games with differential discounting are restricted to asymptotic results (i.e. firms are infinitely patient) and the two-player case. In our setting, we characterize all the efficient equilibria with n players for a fixed discount factor, possibly due to some special structure of Bertrand competition

[3] The time to reach the efficient stationary collusive equilibrium is bounded across all efficient collusive equilibria for a given profile of discounting rates.

game.

2.1.1 Related Literature

It is not without reason that previous works have focused on the symmetric model. First, there is the issue of equilibrium selection as mentioned. There are always many equilibria - hence there is always the issue of equilibrium selection - in repeated games. The model of dynamic Bertrand competition is no exception. For symmetric models, it might make sense to focus on the symmetric (and efficient) equilibrium, possibly as a focal point. However, it is not clear which equilibrium would be a focal point when firms are asymmetric. Secondly, the theory of repeated games with differential discounting is still at its development stage. For these reasons, there are not many works that study collusion among heterogeneous firms. In our view, this fact limits the scope of applications of the repeated Bertrand competition model.

One notable exception is [Har89]. It shows that a stationary collusive equilibrium can be sustained with differential discounting if and only if the average discount factor exceeds some critical level. Our first result builds and improves on this result. We provide a more complete characterization regarding the possibility of collusion by considering all equilibria including nonstationary ones.[4] Clearly it is important to consider nonstationary equilibria because almost all stationary equilibria are not efficient with differential discounting as our second result shows. Another difference between our paper and [Har89] is that we obtain the unique equilibrium in the long run. To cope with the issue of multiple stationary equilibria, [Har89] uses a bargaining solution to select one equilibrium. On the other hand, we show that the long run equilibrium behavior is the same across all effi-

[4]Since a collusive outcome can be sustained by a stationary equilibrium when the average discount factor exceeds the critical level, the crucial step for our result is to show that no nonstationary collusive equilibrium exists when the average discount factor is below the same critical level.

cient equilibria. Thus we do not need to rely on any equilibrium selection criterion other than efficiency as long as we are concerned with the long-run outcome.

The seminal contribution in the theory of repeated game with differential discounting is [LP99]. It studies a general two-player repeated game with differential discounting and shows that the set of feasible payoffs is larger than the convex hull of the underlying stage game payoffs because players can mutually benefit from trading payoffs across time. They also characterize the limit equilibrium payoff set as discount factors go to 1 keeping their ratio fixed. In particular, they show that there is some individually rational and feasible payoff that cannot be sustained in equilibrium no matter how patient the players are.

There are some recent contributions in the theory of repeated games with differential discounting. [Che08] and [GLT11] study stage games with one dimensional payoffs. [Sug13] proves a folk theorem for repeated games with imperfect monitoring and with differential discounting. [FS09] studies repeated prisoner's dilemma games with differential discounting and with side payments. This paper seems to be particularly related to our paper because we use market share as a way to transfer utility.

This paper is organized as follows. We describe the model in detail in the next section. In section 2.3, we prove our first result regarding the critical average discount factor. In section 2.4, we characterize efficient equilibria. We conclude and discuss potential extensions of our results in the last section. Most of the proofs are relegated to the appendix.

2.2 Model of Repeated Bertrand Competition with Differential Discounting

This section describes the basic structure of our model, an infinitely repeated Bertrand game. In what follows, we first define the stage game, then construct the infinitely repeated game.

The main features of the stage game are the followings. The players are $n \geq 2$ firms represented by the numbers $\mathcal{I} = \{1, 2, ..., n\}$. They offer a homogeneous product whose market demand is characterized by continuous function $D : \mathbb{R}_+ \to \mathbb{R}_+$. Each firm has a linear cost function $C_i : \mathbb{R}_+ \to \mathbb{R}_+$ given by $C_i(q_i) = c_i q_i$, where $i \in \mathcal{I}$, $c_i \geq 0$ is the marginal cost, and q_i indicates the quantity produced by firm i. We suppose that $c_1 \leq c_2 \leq .... \leq c_n$ without loss of generality and denote $\mathcal{I}_* = \{i \in \mathcal{I} : c_i = c_1\}$ and $n_* = \#(\mathcal{I}_*)$. We assume that $n_* \geq 2$. Hence, in one-shot Bertrand competition, the market price would be c_1 and no firm would make any profit. It is assumed that the demand function satisfies the following regularity conditions: D is decreasing on $(0, \infty)$; there exists the monopoly price for each firm: $p_i^m > c$ for firm i that maximizes $p\left(D(p) - c_i\right)$. We assume that the marginal costs are not very different: even the highest cost c_n is less than p_1^m. This implies that $p_i^m > c_j$ for any $i, j \in \mathcal{I}$.[5]

At the beginning of a stage game, firms make price decisions and suggest how to allocate output quotas in case of a draw in prices. If a firm charges a price that is higher than a price charged by another firm, then the firm's market share is 0. The firm that charges the lowest price must produce enough output to satisfy the market demand. In case there are more than one firm that charges the lowest price, the market is allocated among those firms according to their suggestions. Formally, firm i's pure action is given by a 2-tuple $a_i = (p_i, r_i) \in A_i$, where p_i is the price choice, r_i reflects firm i's request of market share in case of tie. Hence $A_i = \mathbb{R}_+ \times [0, 1]$ is the set of pure actions available for player i. The set of pure action profiles is $A = \prod_{i \in \mathcal{I}} A_i$. Firm i's profit function $\pi_i : A \to \mathbb{R}$ can be written

[5] If the marginal of some firm is too high, it is likely that the presence of such a firm is irrelevant for our analysis.

as

$$\pi_i[a] = \begin{cases} D(p_i)(p_i - c_i) & \text{if } p_i < p^*_{-i}, \\ \frac{r_i}{R^*} D(p_i)(p_i - c_i) & \text{if } p_i = p^*_{-i} \text{ and } R^* \neq 0, \\ \frac{1}{|\widehat{\mathcal{I}}|} D(p_i)(p_i - c_i) & \text{if } p_i = p^*_{-i} \text{ and } R^* = 0, \\ 0 & \text{if } p_i > p^*_{-i}, \end{cases}$$

where $p^*_{-i} = \min_{j \neq i} p_j$, $\widehat{\mathcal{I}} = \{i \in \mathcal{I} : p_i = \min_{j \in \mathcal{I}} p_j\}$, and $R^* = \sum_{j \in \widehat{\mathcal{I}}} r_j$.

Given the stage game described above, we now define the infinitely repeated game. Basically, we adopt a discrete time model in which the previous stage game is played in each of the periods $t \in \mathbb{N}$. The distinguishing feature of our dynamic Bertrand competition model is that the players have different discount factors given by $\delta_i \in (0, 1), i \in \mathcal{I}$.

The set of possible histories in period t is given by $H^t = A^{t-1}$, where A^0 indicates the empty set, and A^t denotes the t-fold product of A. A period t-history is thus a list of $t - 1$ action profiles. We suppose perfect monitoring throughout, i.e., at the end of each period, all players observe the action profile chosen in all the previous periods. Setting $H = \cup_{t \in \mathbb{N}} H^t$, a pure strategy for firm i is defined as a mapping $\sigma_i : H \to A_i$, and consequently, a strategy profile is given by $\sigma = (\sigma_i)_{i \in \mathcal{I}}$.

Each strategy profile σ induces an infinite sequence of action profiles $\mathbf{a}(\sigma) = (a^t(\sigma))_{t \in \mathbb{N}} \in A^\infty$, where $a^t(\sigma) \in A$ denotes the action profile induced by σ in period t. We call the sequence $\mathbf{a}(\sigma)$ outcome path (or more simply, outcome) generated by a strategy profile σ. Finally, for a given strategy profile σ, and its corresponding outcome path $\mathbf{a}(\sigma) = (a^t(\sigma))_{t \in \mathbb{N}}$, the time-average repeated game payoff for firm i at time t is

$$U_{i,t}[\mathbf{a}(\sigma)] = (1 - \delta_i) \sum_{\tau = t}^{\infty} \delta_i^{\tau - t} \pi_i[a^\tau(\sigma)].$$

In the following sections, we will just focus on subgame perfect equilibrium solutions, and we will limit our attention to pure strategy equilibria.

2.3 Critical Average Discount Factor for Collusion

In this section, we derive a necessary and sufficient condition to sustain a collusive equilibrium outcome. We say that the firms are colluding when there is at least one period in which the equilibrium outcome is not a competitive one, i.e. when there is at least one firm that makes positive profit in some period. We formalize this as follows.

Definition 7. *An outcome* $\mathbf{a} = (a^t)_{t\in\mathbb{N}}$ *is considered a* ***collusive outcome*** *if and only if there exists* $t' \in \mathbb{N}$ *such that* $\pi_i(a^{t'}) > 0$ *for some* $i \in \mathcal{I}$*. A* ***collusive equilibrium*** *is a subgame perfect equilibrium that generates a collusive outcome.*

Then we can obtain the following sharp characterization, which says that a collusive outcome can be sustained if and only if the average discount factor among the lowest cost firms is above some threshold.

Theorem 2.3.1. *There exists a collusive equilibrium if and only if*

$$\frac{\sum_{i\in\mathcal{I}_*}\delta_i}{n_*} \geq \frac{n_* - 1}{n_*}.$$

Proof. See the appendix. □

When the firms are symmetric, there exists a collusive equilibrium if and only if $\delta \geq \frac{n-1}{n}$. Thus our result is a substantial generalization of this well-known result to the case with heterogeneous discounting and costs.

It follows from the result in [Har89] that $\frac{n_*-1}{n_*}$ is the critical threshold to support a collusive outcome by a stationary collusive equilibrium, i.e. an equilibrium in which each firm keeps a certain level of market share every period and the price is always the same. Take any price p strictly between the minimum cost $c^* = \min_{i\in\mathcal{I}} c_i$

and the next smallest cost. There exists a stationary collusive equilibrium by the lowest cost firms in which the market price is always p and firm $i(\in \mathcal{I}_*)$ gains share $\alpha_i \in [0,1]$ of the joint profit π in every period if the following inequalities are satisfied for all $i \in \mathcal{I}_*$.

$$(1-\delta_i)\pi \leq \alpha_i \pi$$

By dividing both sides by π and summing up these inequalities across the firms, it can be shown that such $\alpha_i, i \in \mathcal{I}_*$ exists if and only if the average discount factor among the lowest cost firms is larger than or equal to $\frac{n^*-1}{n*}$.

A more difficult part of the proof is to show that collusion is impossible when the average discount factor is less than $\frac{n^*-1}{n*}$, even if nonstationary equilibria are considered. In nonstationary equilibrium, it is possible to transfer market shares over time to generate larger continuation profits in the future, which may enable the firms to sustain collusion. It turns out that such transfer does not work. To improve efficiency, it is necessary to let less patient firms to gain more market shares first and let more patient firms to gain more shares later. Intuitively, such an arrangement is in conflict with less patient firms' incentive constraints in later periods.

Here is a sketch of our formal proof. We assume that the marginal cost is the same across all firms to simplify our exposition. Firm i's incentive constraint in period t is given by the equality

$$U_{i,t} = (1-\delta_i)\,\pi_i\left(a^t\right) + \delta_i U_{i,t+1} = (1-\delta_i)\,\pi^*\left(a^t\right) + \eta_{i,t}$$

where a^t is the action profile in period t, $\pi^*\left(a^t\right) = \sum_i \pi_i\left(a^t\right)$ is the joint profit in period t, $U_{i,t+1}$ is firm i's continuation profit from period $t+1$ on, and $\eta_{i,t} \geq 0$ is a slack variable (firm i's incentive constraint is binding in period t if and only if $\eta_{i,t} = 0$). Note that each firm gains the same equilibrium join profit by price-cutting because the cost is assumed to be the same. Since this equality holds in every period, we can replace $U_{i,t+1}$ with $(1-\delta_i)\,\pi^*\left(a^{t+1}\right) + \eta_{i,t+1}$ and divide both sides by $1 - \delta_i$ to obtain

$$\pi_i\left(a^t\right) + \delta_i \pi^*\left(a^{t+1}\right) = \pi^*\left(a^t\right) + \frac{\eta_{i,t} - \delta_i \eta_{i,t+1}}{1 - \delta_i}.$$

Summing up these equalities across the firms, we have the following equation regarding $\pi^*\left(a^t\right)$:

$$\pi^*\left(a^{t+1}\right) = \frac{n-1}{\sum_{i\in\mathcal{I}} \delta_i}\pi^*\left(a^t\right) + \frac{1}{\sum_{i\in\mathcal{I}} \delta_i}\sum_{i\in\mathcal{I}} u_{i,t},$$

where $u_{i,t} = \frac{\eta_{i,t} - \delta_i \eta_{i,t+1}}{1-\delta_i}$.

The coefficient of $\pi^*\left(a^t\right)$ is larger than 1 if and only if the average discount factor is less than $\frac{n-1}{n}$. In fact, we can show that, when the joint profit is strictly positive in some period, the sequence $\pi^*\left(a^t\right), t = 1, 2, ..$ must diverge to infinity, which is a contradiction. To prove this formally, however, we need to examine carefully the behavior of $\sum_{i\in\mathcal{I}} u_{i,t}, t = 1, 2, 3, ...$

A collusive equilibrium we construct uses a price between the lowest cost and the second lowest cost, so it is not very profitable when this difference between them is small. In such a case, the lowest cost firms would prefer to include the second lowest cost firm(s) in their coalition to raise the equilibrium price. Our result can be easily generalized to accommodate such possibility. Let p be any price. Let $\mathcal{I}(p)$ be the set of firms such that $c_i \leq p$ if and only if $i \in \mathcal{I}(p)$ and $|\mathcal{I}(p)| = n_*(p)$. Call a subgame perfect equilibrium p-collusive equilibrium if the equilibrium price is always at least as large as p. We can prove the following generalization of the above result.

Theorem 2.3.2. *For any $0 < p \leq p^m$, there exists a p-collusive equilibrium if and only if*

$$\frac{\sum_{i\in\mathcal{I}(p)_*} \delta_i}{n_*(p)} \geq \frac{n_*(p) - 1}{n_*(p)}.$$

The proof is almost the same, hence omitted. We remark that when $\mathcal{I}_* = 1$, i.e. there is the unique lowest cost firm, Theorem 2 still holds. But we need to rely on a less natural punishment. The assumption $\mathcal{I}_* \geq 2$ guarantees that any deviation from a collusive outcome is punished by Nash reversion with 0 profit forever. If $\mathcal{I}_* = 1$, the 0 profit equilibrium requires that there are at least two firms charging c_1, but firm 1 serves the whole market ($r_1 = 1, r_i = 0$ for all $i \neq 1$).

2.4 Characterization of Efficient Collusive Equilibria

In this section, we characterize efficient collusive equilibria with differential discounting rates. We assume that the marginal cost is the same across firms and normalize it to 0. Then the monopoly price can be determined without any ambiguity. Let p^m be the monopoly price and π^m be the monopoly profit. We also assume that $0 < \delta_1 < \delta_2 < .... < \delta_{n-1} < \delta_n < 1$ for the sake of simplicity. The result can be easily extended to the case where the discounting factors of some firms are the same.

Let $\pi_{i,t}, i = 1, ..., n, t \in \mathbb{N}$ be a sequence of profits associated with any collusive equilibrium. By definition, they satisfy the following incentive compatibility condition in every period:

$$(1 - \delta_i)\, \pi_t \leq U_{i,t}$$

where $U_{i,t}$ is firm i's equilibrium continuation profit in the beginning of period t and $\pi_t = \sum_i \pi_{i,t}$. On the other hand, it is clear that any sequence of profit profiles that satisfy those conditions are generated by a collusive equilibrium. Hence we use such a sequence of profit profiles to describe any collusive equilibrium.

A collusive equilibrium is efficient if there is no subgame perfect equilibrium that makes every firm better off weakly and some strictly. Observe that π_t is always in $(0, \pi^m]$ for any efficient collusive equilibrium. π_t cannot exceed the monopoly profit by definition. If $\pi_t < 0$, then we can construct a more efficient equilibrium by just dropping period t.

We know that there exists a stationary collusive equilibrium with monopoly price if and only if $\frac{\sum_{i=1}^{n}\delta_i}{n} \geq \frac{n-1}{n}$. When the average discount factor is strictly larger than $\frac{n-1}{n}$, there is a range of market shares that can be supported by stationary collusive equilibrium. Let $\widehat{\pi}_i$ be firm i's per period profit in the worst stationary collusive equilibrium profit for firm i. Note that $\widehat{\pi}_i = (1-\delta_i)\,\pi^m$ by the incentive compatibility condition. We assume $\frac{\sum_{i=1}^{n}\delta_i}{n} > \frac{n-1}{n}$ for the rest of this section.

We first prove that, in any efficient collusive equilibrium, the joint profit must be strictly increasing until it reaches the monopoly profit and stays there forever.

We start with the following lemma.

Lemma 2.4.1. *Consider any efficient collusive equilibrium where, for some $t \geq 1$, $\pi_{t+1} < \pi^m$ and there is a firm i' such that $U_{i',t+1} > (1-\delta_{i'})\pi_{t+1}$ and $\pi_{i',t+1} > 0$. Then $\pi_{t+1} \geq \frac{\pi_t}{\delta_n}$.*

Proof. Define $\widetilde{\mathcal{I}}_{t+1} = \{i \in \mathcal{I} : U_{i,t+1} = (1-\delta_i)\pi_{t+1}\}$, which is not empty (otherwise the joint profit can be increased to improve efficiency). Suppose that $\pi_{t+1} < \frac{\pi_t}{\delta_n}$. Then $\pi_{i,t} \geq \pi_t - \delta_i\pi_{t+1} \geq \pi_t - \delta_n\pi_{t+1} > 0$ for all $i \in \widetilde{\mathcal{I}}_{t+1}$. Consequently, the profits can be perturbed as follows: $\pi'_{i,t} = \pi_{i,t} - \delta_i\varepsilon$ and $\pi'_{i,t+1} = \pi_{i,t+1} + \varepsilon$, for $i \in \widetilde{\mathcal{I}}_{t+1}$; whereas $\pi'_{i',t} = \pi_{i',t} + \sum_{i\in\widetilde{\mathcal{I}}_{t+1}}\delta_i\varepsilon$ and $\pi'_{i',t+1} = \pi_{i',t+1} - (|\widetilde{\mathcal{I}}_{t+1}|-1)\varepsilon$. Since $\pi_{t+1} < \pi^m$ and $\pi_{i',t+1} > 0$, this new allocation is feasible and incentive compatible for $\varepsilon > 0$ small enough. Moreover, as $\sum_{i\in\widetilde{\mathcal{I}}_{t+1}}\delta_i > |\widetilde{\mathcal{I}}_{t+1}|-1$, it also Pareto-dominates the initial one. This is a contradiction. □

The next theorem proves a strong monotonicity property for efficient collusive equilibria.

Theorem 2.4.1. *For any efficient collusive equilibrium, there exists T such that $\pi_t < \pi_{t+1}$ for $t = 1, ..., T-1$ and $\pi_t = \pi^m$ for any $t \geq T$. Furthermore, this T is bounded across all efficient collusive equilibria.*

Proof. Take any efficient collusive equilibrium. Let $\pi_t \in (0, \pi_m]$ be a joint profit in any period t. We assume that $\pi_t > \delta_n\pi_{t+1}$ and $\pi_{t+1} < \pi^m$, and derive a contradiction. If those two conditions are satisfied, then it must be the case that $\pi_{t+1} = \sum_{i\in\widetilde{\mathcal{I}}_{t+1}} \pi_{i,t+1}$ by Lemma 2.4.1. Therefore, there is $j \in \widetilde{\mathcal{I}}_{t+1}$ such that $\pi_{j,t+1} > (1-\delta_j)\pi_{t+1}$, otherwise,

$$\pi_{t+1} = \sum_{i\in\widetilde{\mathcal{I}}_{t+1}} \pi_{i,t+1} \le \pi_{t+1} \sum_{i\in\widetilde{\mathcal{I}}_{t+1}} (1-\delta_i) = \pi_{t+1}(|\widetilde{\mathcal{I}}_{t+1}| - \sum_{i\in\widetilde{\mathcal{I}}_{t+1}} \delta_i),$$

but $\sum_{i\in\widetilde{\mathcal{I}}_{t+1}} \delta_i > |\widetilde{\mathcal{I}}_{t+1}| - 1$.

As a result, $(1-\delta_j)\pi_{t+2} \le U_{j,t+2} < (1-\delta_j)\pi_{t+1}$. The first inequality is derived from the incentive constraint in period $t+2$, whereas the second one from the fact that $\pi_{j,t+1} > (1-\delta_j)\pi_{t+1}$ and $U_{j,t+1} = (1-\delta_j)\pi_{t+1}$. Then, $\pi_{t+1} > \pi_{t+2}$. We can proceed in a similar manner to obtain $\pi_{t+k} > \pi_{t+k+1}$ for every $k \ge 1$, which contradicts the efficiency assumption. Hence it must be the case that either $\pi_t \le \delta_n \pi_{t+1}$ or and $\pi_{t+1} = \pi^m$. Clearly this implies that there is T such that $\pi_t < \pi_{t+1}$ for $t = 1, ..., T-1$ and $\pi_t = \pi^m$ for any $t \ge T$.

Finally we prove that this T is bounded across all efficient equilibria. For any given T, each firm's profit per period is at most $\delta_n^{T'-1}\pi^m$ for the first $T-T'$ periods for any $T' \le T$. If T is large, then firm i's payoff is less than $\widehat{\pi}$. Such payoff profile is Pareto-dominated by any stationary collusive equilibrium. □

Next we provide an (almost) complete characterization of efficient collusive equilibria. Consider any efficient collusive equilibrium where firm i's equilibrium profit exceeds $\widehat{\pi}_i$. Then every firm's incentive constraint is not binding in the first period, hence the equilibrium joint profit must be π^m in the first period. Given our monotonicity result, this implies that the equilibrium price is always p^m for this class of efficient collusive equilibria. We call such collusive equilibrium p^m-efficient collusive equilibrium.

The next theorem characterizes the structure of p^m-efficient collusive equilibrium. Observe that this characterization is a complete characterization of the asymptotic behavior of all efficient collusive equilibria, because every efficient collusive equilibrium converges to some p^m-efficient collusive equilibrium eventually within finite time by our previous result.

In p^m-efficient collusive equilibrium, more patient firms "lend" the market share initially to more impatient firms. However, the ability of impatient firms to "pay back" the market share is limited by the requirement that each firm's profit cannot be lower than its worst stationary equilibrium profit $\widehat{\pi}_i$.

Theorem 2.4.2. *Every p^m-efficient collusive equilibrium has the following structure: there exists $t_1 \leq t_2 \leq ..., \leq t_{n-1}$ such that, for every i,*

1. $\pi_{i,t} = 0$ *for every* $t < t_{i-1}$
2. $\pi_{i,t} \in \left[0, \pi^m - \sum_{h=1}^{i-1} \widehat{\pi}_h\right]$ *for* $t = t_{i-1}$
3. $\pi_{i,t} = \pi^m - \sum_{h=1}^{i-1} \widehat{\pi}_h$ *for* $t = t_{i-1} + 1, ..., t_i - 1$
4. $\pi_{i,t} \in \left[\widehat{\pi}_i, \pi^m - \sum_{h=1}^{i-1} \widehat{\pi}_h\right]$ *for* $t = t_i$
5. $\pi_{i,t} = \widehat{\pi}_i$ *for* $t > t_i$
6. ***Incentive Constraints in the first period***

$$\begin{aligned} &\delta_i^{t_{i-1}-1}\left[(1-\delta_i)\,\pi_{i,t_{i-1}} + \left(\delta_i - \delta_i^{t_i - t_{i-1}}\right)\left\{\pi^m - \sum_{h=1}^{i-1}\widehat{\pi}_h\right\}\right] \\ &+\delta_i^{t_{i-1}-1}\left[(1-\delta_i)\,\delta_i^{t_i-t_{i-1}}\widehat{\pi}_{i,t_i} + \delta_i^{t_i-t_{i-1}+1}\widehat{\pi}_i\right] \\ &\quad \geq (1-\delta_i)\,\pi^m \end{aligned}$$

Furthermore, if there exist $(t_1, t_2, ..., t_{n-1})$ and a sequence of profit profiles $\pi_{i,t}$ that satisfy the above conditions, then there exists a corresponding p^m-efficient collusive equilibrium that generates them.

Proof. See the appendix. □

In words, every p^m-efficient collusive equilibrium has the following properties:

- From period 1 to period $t_1 - 1$, firm 1 gets the whole share.
- In period t_1, firm 1 and 2 shares the market where $\pi_{i,t_1} \geq \widehat{\pi}_1$. After this period, firm 1's share is going to be always $\widehat{\pi}_1$.
- From period $t_1 + 1$ to period $t_2 - 1$, firm 2 gets $\pi^m - \widehat{\pi}_1$.
- In period t_2, firm 2 and 3 shares the market where $\pi_{i,t_2} \geq \widehat{\pi}_2$. After this period, firm 2's share is going to be always $\widehat{\pi}_2$.
- From period $t_2 + 1$ to period $t_3 - 1$, firm 3 gets $\pi^m - \widehat{\pi}_1 - \widehat{\pi}_2$.
- ...
- After period t_{n-1}, firm n gets $\pi^m - \sum_{h=1}^{n-1} \widehat{\pi}_h$ and firm $h < n$ gets $\widehat{\pi}_h$ forever.

There are two critical periods for firm i: t_{i-1} and t_i. Up to t_{i-1}, firm i's market share is 0. The periods between t_{i-1} and t_i is the pay back time when firm i gets all the market share subject to the constraint that each less patient firm $h < i$ gains $\widehat{\pi}_h$. After t_i, firm i's profit is reduced to $\widehat{\pi}_i$ and stay there forever. It may be the case that there is some overlap: $t_k = t_{k+1} =, ..., = t_m = t'$ for some $m > k$. Note that $\pi_{i,t'} \geq \widehat{\pi}_i$ for $i = t_k, t_{k+1}, ..., t_{m-1}$ in such a case.

Several comments are noteworthy. First, one implication of our theorem is that there exists the unique efficient stationary collusive equilibrium, to which every efficient collusive equilibrium converges. This is the stationary collusive equilibrium where the price is p^m, firm i's market share is $\widehat{\pi}_i$ for $i = 1, ..., n-1$ and firm n's market share is $\pi^m - \sum_{i=1,...,n-1} \widehat{\pi}_i$ in every period, which corresponds to the worst stationary collusive equilibrium for firm $i = 1, ..., n-1$ (and the best one for firm n). All the other efficient collusive equilibria must be nonstationary.

Second, our result delivers the unique prediction in the long run without any equilibrium selection. This is not the case if we focus on stationary collusive equilibria. Third, when $\delta_i = \delta_{i+1}$ for some i, their market share is characterized by similar conditions: their market share is 0 before t_{i-1}, $\pi_{i,t} = \pi_{i-1,t} = \pi_i (= \pi_{i+1})$ after t_i, and can be somewhat arbitrary between t_i and t_{i-1} (but we can assume that their market shares are constant within these periods without loss of generality).

2.5 Conclusion and Discussion

In the context of Bertrand price competition in an infinitely repeated game, this paper studies collusive behavior among n firms with asymmetric discount factors and asymmetric marginal costs.

We find that it is possible to sustain a collusion if and only if the average discount factor of the lowest cost firm is at least as large as $(n_* - 1)/n_*$, where n_* is the number of the lowest cost firms.

This paper also studies efficient collusive equilibria among n firms with differential discounting when the marginal cost is the same across firms. We succeed in characterizing the structure of efficient collusive equilibria. More specifically, we show the followings results:

- In any efficient collusive equilibrium, the joint profit must be strictly increasing over time until it reaches the monopoly profit level within finite time and stay there forever.

- Every efficient collusive equilibrium where no firm's payoff is not too low must be a collusive equilibrium with the monopoly price in every period ("p^m-efficient collusive equilibrium"). In every p^m-efficient collusive equilibrium, a firm's market share is 0 initially, reaches a peak, then converges to the market share that corresponds to the worst stationary collusive equilibrium with the monopoly price (except for the most patient firm).

- Every efficient collusive equilibrium converges to the unique efficient stationary collusive equilibrium in the long run, where the equilibrium price is

p^m, firm i's profit per period is $\widehat{\pi}_i$ for $i = 1, ..., n-1$ and $1 - \sum_{i=1,...,n-1} \widehat{\pi}_i$ for $i = n$ in every period.

2.A Appendix

2.A.1 Proof of Theorem 2.3.1

We already discussed that there exists a collusive stationary subgame perfect equilibrium when the inequality is satisfied. Thus we just need to show that there is no collusive subgame perfect equilibrium when $\frac{\sum_{i \in \mathcal{I}_*} \delta_i}{n_*} < \frac{n_* - 1}{n_*}$.

By contradiction, begin by assuming that $\tilde{\mathbf{a}} = (\tilde{a}^t)_{t \in \mathbb{N}}$ is a collusive equilibrium outcome, and without loss of generality, assume that $\pi_*(\tilde{a}^1) = \sum_{i \in \mathcal{I}} \pi_(\tilde{a}^1) > 0$.

Note first that for each $i \in \mathcal{I}_*$, there exists a bounded nonnegative sequence $\{\eta_i^{(t)} : t \in \mathbb{N}\}$ defined by $\eta_i^{(t)} = U_{i,t}(\tilde{\mathbf{a}}) - (1 - \delta_i)\pi_*(\tilde{a}^t)$. Moreover, since $U_{i,t}(\tilde{\mathbf{a}}) = (1 - \delta_i)\pi_i(\tilde{a}^t) + \delta_i U_{i,t}(\tilde{\mathbf{a}})$, we have that

$$(1 - \delta_i)\pi_i(\tilde{a}^t) + \delta_i U_{i,t}(\tilde{\mathbf{a}}) = (1 - \delta_i)\pi_*(\tilde{a}^t) + \eta_i^{(t)},$$

and therefore

$$(1 - \delta_i)\pi_i(\tilde{a}^t) + \delta_i[(1 - \delta_i)\pi_*(\tilde{a}^{t+1}) + \eta_i^{(t+1)}] = (1 - \delta_i)\pi_*(\tilde{a}^t) + \eta_i^{(t)},$$

or equivalently,

$$\pi_i(\tilde{a}^t) = \left[\pi_*(\tilde{a}^t) + \frac{\eta_i^{(t)}}{(1 - \delta_i)}\right] - \delta_i \left[\pi_*(\tilde{a}^{t+1}) + \frac{\eta_i^{(t+1)}}{(1 - \delta_i)}\right].$$

Adding up this inequality across $i \in \mathcal{I}_*$ and denoting $s_* = \sum_{i \in \mathcal{I}_*} \delta_i$, we obtain

$$\pi_*(\tilde{a}^t) = n_* \pi_*(\tilde{a}^t) - s_* \pi_*(\tilde{a}^{t+1}) + \sum_{i \in \mathcal{I}_*} \frac{\eta_i^{(t)} - \delta_i \eta_i^{(t+1)}}{(1 - \delta_i)},$$

or more shortly,

$$\pi_*(\tilde{a}^{t+1}) = \gamma \pi_*(\tilde{a}^t) + \frac{1}{s_*} \sum_{i \in \mathcal{I}_*} u_{i,t},$$

where $\gamma = (n_* - 1)/s_*$ and $u_{i,t} = (\eta_i^{(t)} - \delta_i \eta_i^{(t+1)})/(1-\delta_i)$.

Before proceeding, it is useful to note that $\gamma > 1$ and therefore

$$\begin{aligned}
\pi_*(\tilde{a}^{t+1}) &\geq \pi_*(\tilde{a}^t) + \frac{1}{s_*}\sum_{i\in\mathcal{I}_*} u_{i,t} \\
&\geq \pi_*(\tilde{a}^1) + \frac{1}{s_*}\sum_{i\in\mathcal{I}_*}\sum_{j=1}^{t} u_{i,j},
\end{aligned}$$

for every $t \in \mathbb{N}$.

Now consider the series $\sum_{j=1}^t u_{i,j}$ for $i \in \mathcal{I}_*$, and observe that it maybe written as

$$\sum_{j=1}^{t} u_{i,j} = \frac{\eta_i^{(1)}}{(1-\delta_i)} + \sum_{j=2}^{t} \eta_i^{(j)} - \frac{\delta_i \eta_i^{(t+1)}}{(1-\delta_i)}.$$

Since the equilibrium profit is bounded above for each firm by assumption, the equilibrium profit is bounded below as well for each firm; otherwise the average discounted profit is negative. This implies that there exists M^* such that $0 \leq \eta_i^{(j)} \leq M^*$ for all $j \in \mathbb{N}$ and $i \in \mathcal{I}_*$. Observe that this implies that the series $\sum_{j=2}^t \eta_i^{(j)}$ must be either unbounded above or converging to a finite (nonnegative) real number.

Suppose first that $\sum_{j=2}^\infty \eta_i^{(j)}$ is unbounded above for some $i \in \mathcal{I}_*$. On the one hand, we know that $\sum_{i\in\mathcal{I}_*}(\sum_{j=1}^t u_{i,t})$ is unbounded above, too. On the other hand, we have that $\pi_*(\tilde{a}^{t+1}) \geq \pi_*(\tilde{a}^1) + (1/s_*)\sum_{i\in\mathcal{I}_*}\sum_{j=1}^t u_{i,t}$, which is a contradiction because the sequence $\{\pi_*(\tilde{a}^t) : i \in \mathbb{N}\}$ is bounded above.

Suppose now that $\sum_{j=2}^\infty \eta_i^{(j)}$ is finite for all $i \in \mathcal{I}_*$. Then we have that $\eta_i^{(t)}$ as well as $u_{i,t}$ converge to zero for all $i \in \mathcal{I}_*$. If $\eta_i^{(j)} = 0$ for all $i \in \mathcal{I}_*$ and $j \in \mathbb{N}$, it follows immediately that $\sum_{i\in\mathcal{I}_*}(\sum_{j=1}^t u_{i,t}) \geq 0$. On the other hand, if $\eta_i^{(t_i)} = c_i > 0$ for some $i \in \mathcal{I}_*$ and $t_i \in \mathbb{N}$, there exists $T_i > t_i$ such that $|\eta_i^{(t)}| < c_i(1-\delta_i)/(2\delta_i)$ for all $t \geq T_i$. As a result, we have that

$$\sum_{j=1}^{t} u_{i,j} \geq \frac{\eta_i^{(1)}}{(1-\delta_i)} + c_i - \frac{c_i}{2},$$

when $t > T_i$. As $\mathcal{I}_*$ is a finite set, there is $T \in \mathbb{N}$ (independent of i) such that

$\sum_{i\in\mathcal{I}_*}(\sum_{j=1}^{t} u_{i,t}) \geq 0$ for all $t \geq T$, and consequently, $\pi_*(\tilde{a}^t) \geq \pi_*(\tilde{a}^1)$ as long as $t \geq T$.

Before proceeding, observe first that there is $\tilde{t} \in \mathbb{N}$ such that $\gamma^{\tilde{t}}\pi_*(a^1) > 2M$. Secondly, since $\mathcal{I}_*$ is a finite set and $u_{i,t}$ converge to zero for each $i \in \mathcal{I}_*$, there exists $\tilde{T} \in \mathbb{N}$ (independent of i) such that $\tilde{T} > T$ and $|u_{i,t}| < (s^* M^*)/(n^* \tilde{t}\gamma^{\tilde{t}})$ for all $i \in \mathcal{I}_*$ and $t \geq \tilde{T}$.

The following inequality is a straightforward implication:

$$\begin{aligned}
\pi_*(\tilde{a}^{\tilde{T}+t}) &= \gamma\pi_*(\tilde{a}^{\tilde{T}+t-1}) + \frac{1}{s_*}\sum_{i\in\mathcal{I}_*} u_{i,\tilde{T}+t-1} \\
&\geq \gamma\pi_*(\tilde{a}^{\tilde{T}+t-1}) - \frac{M}{\tilde{t}\gamma^{\tilde{t}}},
\end{aligned}$$

and by induction, we can prove that

$$\pi_*(\tilde{a}^{\tilde{T}+t}) \geq \gamma^t\pi_*(\tilde{a}^{\tilde{T}}) - \sum_{j=0}^{t-1}\frac{M^*}{\tilde{t}\gamma^{\tilde{t}-j}},$$

for every $t \in \mathbb{N}$. Finally, after replacing $t = \tilde{t}$ in the previous inequality, we obtain the desired result:

$$\begin{aligned}
\pi_*(\tilde{a}^{\tilde{T}+\tilde{t}}) &\geq \gamma^{\tilde{t}}\pi_*(\tilde{a}^{\tilde{T}}) - \sum_{j=0}^{\tilde{t}-1}\frac{M^*}{\tilde{t}\gamma^{\tilde{t}-j}} \\
&> \gamma^{\tilde{t}}\pi_*(\tilde{a}^1) - \sum_{j=0}^{\tilde{t}-1}\frac{M^*}{\tilde{t}} \\
&> 2M^* - M^*.
\end{aligned}$$

The second inequality follows by $\pi_*(\tilde{a}^{\tilde{T}}) \geq \pi_*(\tilde{a}^1)$ and $\gamma > 1$, whereas the last one by $\gamma^{\tilde{t}}\pi_*(\tilde{a}^1) > 2M^*$. And obviously, this is a contradiction because $\pi_*(\tilde{a}^{\tilde{T}+\tilde{t}}) \leq M^*$.

2.A.2 Proof of Theorem 2.4.2

We prove the theorem through a series of lemmata.

Lemma 2.A.1. *For any efficient subgame perfect equilibrium, if firm i's incentive constraint is not binding in period $t > 1$, then $\pi_{j,t-1} = 0$ for every $j > i$.*

Proof. Suppose not, i.e. there exists some monopoly-price efficient SPE where firm i's incentive constraint is not binding in period $t > 1$ and $\pi_{j,t-1} > 0$ for some $j > i$. Then there is a period $t' > t$ such that firm i's incentive constraint is not binding for $t, t+1, ..., t'$ and $\pi_{i,t'} > 0$. We can find such t', otherwise $\pi_{i,t+1} = \pi_{i,t+2} = ... = 0$ (if $\pi_{i,t+1} = 0$, then $U_{i,t+2} > \widehat{\pi}_i$ hence i's incentive constraint in period $t+2$ is not binding. If $\pi_{i,t+2} = 0$, then $U_{i,t+2}...$). Such a path is not sustainable.

Now perturb the profit of firm i and j as follows.

$$\begin{aligned}
\pi'_{i,t} &= \pi_{i,t} + \varepsilon, \\
\pi'_{j,t} &= \pi_{j,t} - \varepsilon, \\
\pi'_{i,t'} &= \pi_{i,t'} - \varepsilon', \\
\pi'_{j,t'} &= \pi_{j,t'} + \varepsilon',
\end{aligned}$$

We are basically exchanging firm j's market share in period t with firm i's market share in period t',keeping every other firm's profit at the same level. Since $\delta_i < \delta_j$, $\pi_{j,t} > 0$ and $\pi_{i,t'} > 0$, we can pick $\varepsilon, \varepsilon' > 0$ so that firm j's continuation payoff in every period from t to t' increases and firm i's continuation payoff in period t increases. So this allocation Pareto-dominates the original SPE allocation. Firm j's incentive constraints are not affected at all. Firm i's incentive constraints in period t is satisfied by construction. Finally, we can take $\varepsilon, \varepsilon' > 0$ small enough so that firm i's incentive constraint from period $t+1$ to t' is still not binding. So we can construct a more efficient SPE in this case, a contradiction. □

Lemma 2.A.2. *For any monopoly-price efficient subgame perfect equilibrium, if* $\pi_{i,t} < \widehat{\pi}_i$, *then* $\pi_{j,t} = 0$ *for every* $j > i$.

Proof. If $\pi_{i,t} < \widehat{\pi}_i$, then $U_{i,t+1} > \widehat{\pi}_i$. Hence firm i's incentive constraint is not binding in period $t+1$. Then $\pi_{j,t} = 0$ for every $j > i$ by Lemma 1. □

Lemma 2.A.3. *For any efficient subgame perfect equilibrium with* $\pi > \widehat{\pi}$, *(1)*$\pi_{1,t} \geq$

$\widehat{\pi}_1$ *for every* $t \geq 1$ *and (2)* $\pi_{1,t'+k} = \widehat{\pi}_1$ *for any* $k = 0, 1, ...$ *if firm* 1*'s incentive constraint is binding in period* t'*.*

Proof. If $\pi_{1,t} < \widehat{\pi}_1$ for any t, then $\pi_{j,t} = 0$ for every $j = 2, 3, ..., n$ by Lemma 2. This contradicts to $\sum_i \pi_{i,t} = \pi^m$.

Firm 1's incentive constraint is binding in period t' if and only if $U_{1,t'} = \widehat{\pi}_1$. Clearly this holds if and only if $\pi_{1,t'+k} = \widehat{\pi}_1$ for $k = 0, 1, 2.....$

□

By induction, a similar property holds for every firm.

Lemma 2.A.4. *For any efficient subgame perfect equilibrium with* $\pi > \widehat{\pi}$*, suppose that* $\pi_{h,t+k} = \widehat{\pi}_h$ *for every* $k = 0, 1, 2, ...$ *and every* $h = 1, 2, ..., i$ *for some* t *and some* $i \in \mathcal{I}$*. Then (1)* $\pi_{i+1,t+k} \geq \widehat{\pi}_{i+1}$ *for every* $k = 0, 1, 2, ...$ *and (2)* $\pi_{i+1,t'+k} = \widehat{\pi}_{i+1}$ *for every* $k = 0, 1, 2...$ *if firm* $i + 1$*'s incentive constraint is binding in period* $t' \geq t$*.*

Proof. Suppose that $\pi_{i+1,t+k} < \widehat{\pi}_{i+1}$ for any k. Then $U_{i+1,t+k+1} > \widehat{\pi}_{i+1}$. Hence firm $i + 1$'s incentive constraint is not binding in period $t + k + 1$. Then $\pi_{j,t+k+1} = 0$ for every $j > i + 1$ by Lemma 1. However, $\sum_h \pi_{h,t+k+1} = \sum_{h=1}^{i+1} \pi_{h,t+k+1} = \sum_{h=1}^{i} \widehat{\pi}_h + \pi_{i+1,t+k+1} < \sum_{h=1}^{i+1} \widehat{\pi}_h \leq \pi^*$, which is a contradiction. This proves (1).

As for (2), firm $i + 1$'s incentive constraint is binding in period $t' \geq t$ if and only if $U_{i,t'} = \widehat{\pi}_i$. By Lemma 4, this holds if and only if $\pi_{i+1,t'+k} = \widehat{\pi}_{i+1}$ for every $k = 0, 1, 2...$ □

Now we can prove Theorem 2.4.2.

Proof. In period 1, we have π_1 such that (1) $\sum_{h=1}^{n} \pi_{h,1} = \pi^m$ and (2) $\pi_{h,1} \geq \widehat{\pi}_h$ for $h = 1, 2, ..., h_1 - 1$, $\pi_{h_1,1} \in \left[0, \pi^m - \sum_{h=1}^{h_1-1} \widehat{\pi}_h\right]$, and $\pi_{h,1} = 0$ for $h > h_1$ for some $h_1 \geq 1$ by.Lemma 2.

By Lemma 1, the incentive constraint must be binding for $h = 1, 2, ..., h_1 - 1$ in period 2. By Lemma 3 and Lemma 4, $\pi_{h,1+k} = \widehat{\pi}_h$ for $h = 1, 2, ..., h_1 - 1$ for the rest of the game ($k = 1, 2, ...$).

In period 2, we have π_2 such that (1) $\sum_{h=1}^{n} \pi_{h,2} = \pi^m$, (2) $\pi_{h,2} = \widehat{\pi}_h$ for $h = 1, 2, ..., h_1 - 1$ (by the previous step), (3) $\pi_{h,2} \geq \widehat{\pi}_h$ for $h = h_1, h_1 + 1, ..., h_2 - 1$, $\pi_{h_2,2} \in \left[0, \pi^m - \sum_{h=1}^{h_2-1} \widehat{\pi}_h\right]$, and $\pi_{h,2} = 0$ for $h > h_2$ for h_2 for some $h_2 \geq h_1$ by.Lemma 4.

By Lemma 1, the incentive constraint must be binding for $h = h_1, ..., h_2 - 1$ in period 3. By Lemma 3 and Lemma 4, $\pi_{h,2+k} = \widehat{\pi}_h$ for $h = h_1, ..., h_2 - 1$ for the rest of the game ($k = 1, 2, ...$) and so on... This proves 1-6 in the statement of the theorem.

On the other hand, suppose that there exist $(t_1, t_2, ..., t_{n-1})$ and a sequence of profit profiles $\pi_{i,t}$ that satisfy 1-6. It is clear that this corresponds to some monopoly-price SPE. So we just show that it is an efficient equilibrium. Suppose not. Then there exists a Pareto-superior monopoly-price efficient SPE, which of course satisfies 1-6. Let $(\widetilde{t}_1, \widetilde{t}_2, ..., \widetilde{t}_{n-1})$ be the corresponding critical periods and $\widetilde{\pi}_{i,t}$ be the associated sequence of profit profiles. Since this equilibrium is more efficient than the former one, it must be the case that either (1) $t_1 < \widetilde{t}_1$ or (2) $t_1 = \widetilde{t}_1$ and $\pi_{1,t_1} \leq \widetilde{\pi}_{1,t_1}$. In either case, it must be the case that, for firm 2, either (1) $t_2 < \widetilde{t}_2$ or (2) $t_2 = \widetilde{t}_2$ and $\pi_{1,t_2} \leq \widetilde{\pi}_{1,t_2}$. By induction, either (1) or (2) holds up to firm $n - 1$. Then firm n's average profit given $\pi_{n,t}$ is higher than firm n's average profit given $\widetilde{\pi}_{n,t}$. This is a contradiction to the assumption that the latter equilibrium is more efficient. □

CHAPTER 3

Semiparametric Estimation of Regression Functions with Continuous and Discrete Covarites

3.1 Introduction

A well-known model in semiparametric econometrics is the single index model. This model can be characterized as follows: $E(y|x) = G(x'\beta)$, where y is a scalar dependent variable, x is a vector of explanatory variables, β is a vector of coefficients, and G is a real function. The distinguishing feature of this model is that the conditional expectation $E(y|x)$ depends on x only through $x'\beta$.

A standard econometric problem is to estimate β when $G(\cdot)$ is unknown. [IL91] and [Ich93] have proposed different estimators for β without imposing a parametric restriction on $G(\cdot)$. These estimators may be difficult to compute because it is required to solve nonlinear optimization problems whose objective function may have many local minima. Nevertheless, when x is a continuous random vector, one can avoid this problem by employing average derivative methods such as [PSS89]. In this case, the computation is simple and do not rely on iterative procedures.

Following a fully nonparametric approach, the econometric literature has focused on estimating $E(y|x)$ when x contains discrete regressors. [RL04]has proposed a kernel smoothing method for nonparametric regression which admits continuous and categorical data; the distinguishing feature of their approach is the use of kernel smoothing for both the continuous and the discrete covariates. [LRW08] uses kernel smoothing in discrete and continuous covariates to estimate treatment effects for Swan-Ganz catherization (right heart catheterization) for critically ill patients admitted to the intensive care unit. [LRW09] proposes an estimator to recover the average treatment effects in the presence of mixed categorical and continuous covariates. [OLR09] considers the problem of estimating a nonparametric regression containing discrete regressors only.

Following a semiparametric approach, [HH96] proposed an alternative non-iterative estimator for (α, β) in the context $E(y|x^c, x^d) = G(x^{c\prime}\beta + x^{d\prime}\alpha)$, where x^c and x^d are continuous and discrete random vectors, respectively. Unfortunately, the identification of (α, β) depends on non-trivial monotonicity conditions and the proposed estimator also depends on them. Furthermore, this estimator does not allow for interaction among discrete and continuous regressors, and therefore, the scope of such estimator may be limited to a small number of applications.

This paper develops a non-iterative weighted average derivative estimator for β in the model $E(y|x^c, x^d) = G(x^{c\prime}\beta, x^d)$. More specifically, I develop an estimator for

$$\delta = E\left[f(x^c, x^d)\frac{\partial g(x^c, x^d)}{\partial x^c}\right], \tag{3.1}$$

where $f(\cdot,\cdot)$ is the joint density of (x^c, x^d) and $g(\cdot,\cdot)$ denotes the conditional expectation of y given (x^c, x^d), i.e., $E(y|x^c, x^d) = g(x^c, x^d)$. Under standard regularity conditions, it can be shown that

$$\delta = -2E\left[y\frac{\partial f(x^c, x^d)}{\partial x^c}\right], \tag{3.2}$$

where $\partial f(x^c, x^d)/\partial x^c$ denotes the derivative of $f(\cdot,\cdot)$ with respect to x^c. Then, if we suppose that $g(x^c, x^d) = G(x^{c\prime}\beta, x^d)$, it follows immediately that $\delta = \beta E[f(x^c, x^d)G_1(x^{c\prime}\beta, x^d)]$ where $G_1(\cdot,\cdot)$ denotes the partial derivative of $G(\cdot,\cdot)$ with respect to its first argument. Thus, it is straightforward to estimate β using the proposed estimator for δ, and therefore, this paper focus on estimating δ.

In order to do so, I combine two approaches. First, I employ kernel techniques for mixed data to estimate $f(\cdot,\cdot)$, as well as, its partial derivative $\partial f(x^c,x^d)/\partial x^c$. Second, I use a standard sample analogues to estimate the expectation term $E\{y[\partial f(x^c,x^d)/\partial x^c]\}$. Combining both approaches, then I propose an estimator for δ that is non-iterative, $\sqrt{N}$-consistent, and asymptotically normal, where N denotes the sample size.

The paper is organized as follows. In the next section, I introduce the basic framework. In section 3.3, I propose the estimator. In section 3.4, I study its asymptotic properties. In section 3.5, I report Monte Carlo experiment results. In the last section, I present the conclusions and suggest further topics for future research.

3.2 Assumptions and Technical Background

3.2.1 The Basic Framework

We consider a model in which Y denotes a dependent variable whereas X is a vector of independent variables. The main feature of this model is that some components of the random vector X are discrete. Formally, we suppose that X takes values on the set $\mathbb{R}^{k_c} \times \mathcal{D}$, where $\mathcal{D} \subseteq \mathbb{R}^{k_d}$ is a finite set, and $k_c, k_d \in \mathbb{N}$. Since we will work with a single equation model, we assume that Y takes values on $\mathbb{R}$. In what follows, in order to distinguish between the continuous and the discrete components of X, we use the notation $X = (X^c, X^d)$, where X^c takes values on $\mathbb{R}^{k_c}$ and X^d on $\mathcal{D}$. Moreover, (Y,X) will denote a $(k+1)\times 1$ random vector, with $k = k_c + k_d$.

Before proceeding, we establish $(\mathbb{R}^{k_c+1} \times \mathcal{D}, \mathcal{M}, \nu)$ as the underlying measure space of (Y,X). We assume that ν is a (σ-finite) product measure which can be written as $\nu = \nu_y \times \nu_c \times \nu_d$, where ν_y a σ-finite measure over $\mathbb{R}$, ν_c is the Lebesgue measure on $\mathbb{R}^{k_c}$, and ν_d is the counting measure over $\mathcal{D}$.

The previous structure may be summarized in the following assumption.

Assumption 5. *The* $(k+1)\times 1$ *random vector* (Y,X) *is a measurable function defined over some probability space* $(\Gamma,\mathcal{G},\mathbb{P})$ *and it takes values on the measure space* $(\mathbb{R}^{k_c+1}\times\mathcal{D},\mathcal{M},\nu)$*. The distribution of* (Y,X) *is absolutely continuous with respect to* ν *with Radon-Nikodyn derivative denoted by* F*.*

In addition to this structural assumption, we also impose the subsequent regularity conditions on the density function and on the conditional expectation of Y given X:

Assumption 6. *Let* $\Omega_d \subseteq \mathcal{D}$ *be the support of* X^d*, let* $f:\mathbb{R}^{k_c}\times\mathcal{D}\to\mathbb{R}$ *be the marginal density of* X *and* $f(\cdot|x^d)$ *the conditional density of* X^c *given* $X^d = x^d$*. For each* $x^d\in\Omega_d$*, the support* $\Omega(x^d)$ *of* $f(\cdot|x^d)$ *is a convex subset of* $\mathbb{R}^{k_c}$*,* $f(\cdot|x^d)$ *is continuous on* $\mathbb{R}^{k_c}$*, and it is also continuously differentiable on the interior of* $\Omega(x^d)$ *(denoted by* $\Omega^\circ(x^d)$*).*

Assumption 7. *Let* $g:\mathbb{R}^{k_c}\times\mathcal{D}\to\mathbb{R}$ *be the conditional expectation of* Y *given* X*, i.e.,* $g(x)=E(Y|X=x)$*. For any* $x^d\in\Omega_d$*,* $g(\cdot,x^d)$ *is continuously differentiable on the closure of* $\Omega(x^d)$ *(denoted by* $\bar{\Omega}(x^d)$*). Moreover,* $\bar{\Omega}(x^d)$ *differs from* $\Omega^\circ(x^d)$ *by a set of measure of zero w.r.t.* ν_c*.*

Assumption 8. *For every* $x^d\in\Omega_d$*, each component of the random vector* $\partial g/\partial x^c$ *and the random matrix* $[\partial f(x^c|x^d)/\partial x^c][(y,x^{c\prime},x^{d\prime})]$ *has finite second moment with respect to the density* $f(\cdot|x^d)$*. Furthermore, for each* $x^d\in\Omega_d$*,* $\partial f(\cdot|x^d)/\partial x^c$ *and* $\partial[g(\cdot,x^d)f(\cdot|x^d)]/\partial x^c$ *satisfy the following Lipschitz conditions: for some function* m_{x^d} *such that* $E\{[(1+|y|+\|x^c\|)m_{x^d}(x^c)]^2|x^d\}<+\infty$*,*

$$\left\|\frac{\partial f(x^c+v|x^d)}{\partial x^c}-\frac{\partial f(x^c|x^d)}{\partial x^c}\right\| < m_{x^d}(x^c)\,\|v\|\,,\ \text{and}$$

$$\left\|\frac{\partial g(x^c+v,x^d)f(x^c|x^d)}{\partial x^c}-\frac{\partial g(x^c,x^d)f(x^c|x^d)}{\partial x^c}\right\| < m_{x^d}(x^c)\,\|v\|\,.$$

Finally, it is also established that $v(x^c,x^d)=E(y^2|x^c,x^d)$ *is continuous in* x^c*.*

After establishing the previous assumptions, we can state the following lemma:

Lemma 3.2.1. *Given Assumptions 5-8, for every* $x^d \in \Omega_d$*,*

$$\int \frac{\partial g(x^c, x^d)}{\partial x^c} f(x^c|x^d)^2 dx^c = -2 \int g(x^c, x^d) \frac{\partial f(x^c|x^d)}{\partial x^c} f(x^c|x^d) dx^c,$$

or equivalently,

$$E\left[f(x^c|x^d) \frac{\partial g(x^c, x^d)}{\partial x^c} \middle| x^d \right] = -2E\left[y \frac{\partial f(x^c|x^d)}{\partial x^c} \middle| x^d \right].$$

Exploiting this result, in the next sections, we will propose an estimator for the following parameter

$$\delta = E\left[f(x^c, x^d) \frac{\partial g(x^c, x^d)}{\partial x^c} \right] = -2E\left[y \frac{\partial f(x^c, x^d)}{\partial x^c} \right], \tag{3.3}$$

where the second equality follows immediately from the law of iterated expectations applied to Lemma 3.2.1.

Given this representation for δ, we will combine two approaches to estimate it. First, we will employ kernel techniques for mixed data to estimate f, and its corresponding partial derivative $\partial f(x^c, x^d)/\partial x^c$. Second, we will employ a standard sample analogue approach to estimate the expectation term $E\{y[\partial f(x^c, x^d)/\partial x^c]\}$.

3.2.2 Kernel Estimators

To construct an estimator for δ, we consider $\{(y_i, x_i) : i = 1, 2, ..., N\}$ as a sample of $(k+1) \times 1$ random vectors. We suppose that they are independent and identically distributed as (Y, X).

Firstly, a standard kernel estimator of the mixed density f is given by

$$\hat{f}(x^c, x^d) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{h} \right)^{k_c} L_\lambda(x^d - x_i^d) K\left(\frac{x^c - x_i^c}{h} \right), \tag{3.4}$$

where the bandwidth h is a smoothing scalar parameter that depends on N, whereas K and L_λ are kernel functions that satisfy the following requirements:

Assumption 9. *The kernel function $K : \mathbb{R}^{k_c} \to \mathbb{R}$ satisfy the following conditions. Its support $\Omega_K \subseteq \mathbb{R}^{k_c}$ is convex and contains the origin as an interior point. Besides, K is a bounded symmetric differentiable function such that $\int K(u)du = 1$, $\int uK(u)du = 0$, and $K(u) = 0$ for all $u \in \partial\Omega_K$, where $\partial\Omega_K$ denotes the boundary of Ω_K.*

Assumption 10. *Let λ_s be s-th component of the vector $\lambda \in \mathbb{R}_+^{k_d}$, and also let $\Omega_{d,s} \subseteq \mathbb{R}$ denote the support of X_s^d, s-th component of X^d. The kernel $L_\lambda : \mathbb{R}^{k_d} \to \mathbb{R}$ can be written as $L_\lambda(u) = \Pi_{s=1}^{k_d}[\lambda_s/(c_s-1)]^{1-\mathbb{1}(u_s)}[1-\lambda_s]^{\mathbb{1}(u_s)}$, where $c_s = \#(\Omega_{d,s}) > 2$, $\lambda_s \in [0,(c_s-1)/c_s]$, and $\mathbb{1}(u_s)$ is an indicator function that equals one when $u_s = 0$, and zero otherwise.*

Under the above conditions, a kernel estimator for $\partial f(x^c, x^d)/\partial x^c$ may then be obtained by differentiating (3.4) with respect to x^c as follows

$$\frac{\hat{f}(x^c, x^d)}{\partial x^c} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{h}\right)^{k_c+1} L_\lambda(x^d - x_i^d)K'\left(\frac{x^c - x_i^c}{h}\right). \tag{3.5}$$

3.3 The Estimator

Following closely[PSS89], this section proposes an estimator for δ. Exploiting expression (3.3), we set $\hat{\delta}$ as the sample analog estimator of δ. Specifically, we define

$$\hat{\delta} = -\frac{2}{N}\sum_{i=1}^{N}\frac{\partial \hat{f}_i(x_i^c, x_i^d)}{\partial x^c} y_i, \tag{3.6}$$

where $\hat{f}_{x,i}(x_i^c, x_i^d)$ is the kernel density estimator

$$\hat{f}_i(x^c, x^d) = \frac{1}{N-1}\sum_{j\neq i}\left(\frac{1}{h}\right)^{k_c} L_\lambda(x^d - x_j^d)K\left(\frac{x^c - x_j^c}{h}\right), \tag{3.7}$$

and $\hat{f}_{x,i}(x_i^c, x_i^d)/\partial x^c$ denotes its partial derivative with respect to x^c, that is

$$\frac{\partial \hat{f}_i(x^c, x^d)}{\partial x^c} = \frac{1}{N-1}\sum_{j\neq i}\left(\frac{1}{h}\right)^{k_c+1} L_\lambda(x^d - x_j^d)K'\left(\frac{x^c - x_j^c}{h}\right). \tag{3.8}$$

After replacing (3.7) in (3.6), $\hat{\delta}_N$ may now be written as

$$\hat{\delta} = \frac{-2}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \left(\frac{1}{h}\right)^{k_c+1} L_\lambda(x_i^d - x_j^d) K' \left(\frac{x_i^c - x_j^c}{h}\right) y_i. \tag{3.9}$$

Furthermore, after defining $z_i = (y_i, x_i^c, x_i^d)$ and using standard U-statistic notation, an alternative representation of $\hat{\delta}$ is

$$\hat{\delta} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} p_N(z_i, z_j), \tag{3.10}$$

with $p_N(z_i, z_j)$ defined as

$$p_N(z_i, z_j) = -\left(\frac{1}{h}\right)^{k_c+1} L_\lambda(x_i^d - x_j^d) K' \left(\frac{x_i^c - x_j^c}{h}\right) (y_i - y_j). \tag{3.11}$$

3.4 Asymptotic Properties

In this section, we first show that the asymptotic bias of $\hat{\delta}$ vanishes at rate $\sqrt{N}$. Second, we prove that $\sqrt{N}[\hat{\delta} - E(\delta)]$ has a limiting normal distribution with zero mean. Exploiting these results, we conclude that $\hat{\delta}$ is a consistent estimator for δ, and also, we provide an explicit formula for its asymptotic covariance matrix.

3.4.1 Asymptotic Bias

To prove that the asymptotic bias of $\hat{\delta}$ vanishes at rate $\sqrt{N}$, we need to impose further restrictions on the kernel K as well as on the rate of convergence of the bandwidths.

Before proceeding, it is useful to express $E(\hat{\delta}_N)$ as a function of h, that is

$$E(\hat{\delta}) = -2 \sum_{x_1^d, x_2^d} L_\lambda(x_1^d - x_2^d) \int \int K(u) g(x^c, x_1^d) f(x^c, x_1^d) \frac{\partial f(x^c + hu, x_2^d)}{\partial x^c} dx^c du, \tag{3.12}$$

where the support of the sum is $(x_1^d, x_2^d) \in \Omega^2$.

As in [PSS89], in order to derive a Taylor expansion for the previous formula, it is sufficient to impose the following restrictions on f and K:

Assumption 11. *Let* $P = (k_c + 4)/2$ *if* k_c *is even and* $P = (k_c + 3)/2$ *if* k_c *is odd. For every* $x^d \in \Omega_d$*, all partial derivatives of* $f(\cdot, x^d)$ *of order* $P + 1$ *exist, and the expectation* $E\{y[\partial f(x^c, x^d)/\partial x^c_{l_1} ... \partial x^c_{l_\rho}]\}$ *exists for all* $\rho < P + 1$*. All moments of* K *of order* P *also exist.*

The expectation $E(\hat{\delta}_N)$ can now be expanded as a Taylor series in h, around $h = 0$:

$$E(\hat{\delta}) = b_0 + b_1 h + b_2 h^2 + ... + b_{P-1} h^{P-1} + O(h^P), \tag{3.13}$$

with b_0 and the lth component of b_p $(p = 1, ..., P-1)$ defined as

$$b_0 = -2 \sum_{x_1^d, x_2^d} L_\lambda(x_1^d - x_2^d) \int g(x^c, x_1^d) f(x^c, x_1^d) \frac{\partial f(x^c, x_2^d)}{\partial x^c} dx^c, \tag{3.14}$$

$$b_{p,l} = \frac{-2}{p!} \sum_{x_1^d, x_2^d} L_\lambda(x_1^d - x_2^d) \tilde{b}_{p,l}(x_1^d, x_2^d), \text{ and}$$

$$\tilde{b}_{p,l}(x_1^d, x_2^d) = \sum_{l_1, ..., l_p = 1}^{k_c} \int \int u_{l_1} ... u_{l_p} K(u) g(x^c, x_1^d) f(x^c, x_1^d) \frac{\partial^{p+1} f(x^c, x_2^d)}{\partial x^c_{l_1} ... \partial x^c_{l_p} \partial x^c_l} dx^c du.$$

After subtracting δ from both sides and multiplying them by $\sqrt{N}$, we obtain

$$\sqrt{N}[E(\hat{\delta}) - \delta] = \sqrt{N}[b_0 - \delta] + b_1 \sqrt{N} h + ... + b_{P-1} \sqrt{N} h^{P-1} + O(\sqrt{N} h^P). \tag{3.15}$$

Hence, with the aim of proving that $\sqrt{N}[E(\hat{\delta}) - \delta]$ converges to zero, we impose the additional restrictions on the kernel K, and on the rate of convergence of the bandwidths h and λ.

Assumption 12. *The kernel function* K *obeys*

$$\int K(u) du = 1, \text{ and}$$

$$\int u_1^{l_1} ... u_{k_c}^{l_p} K(u) du = 0 \text{ for } l_1 + ... + l_p < P.$$

Assumption 13. $N h^{2P} \to 0$ *and* $N \lambda_s^2 \to 0$ *for* $s \in \{1, ..., k_d\}$*, as* $N \to \infty$*.*

Given these assumptions, first, $\sqrt{N}[b_0 - \delta]$ goes to zero due to the fact $\lambda_s = o(\sqrt{N})$. Second, the terms $b_1, ..., b_{P-1}$ vanish because K is a higher order kernel, and finally, since $h^P = o(\sqrt{N})$, the last term $O(\sqrt{N}h^P)$ goes to zero as well.

Formally, we summarize the above discussion in the following theorem:

Theorem 3.4.1. *Under Assumptions 5-13,* $\sqrt{N}[E(\hat{\delta}) - \delta]$ *converges to zero.*

3.4.2 Asymptotic Normality

In order to characterize the asymptotic distribution of $\sqrt{N}(\hat{\delta} - \delta)$, we first show that $\sqrt{N}[\hat{\delta} - E(\hat{\delta})]$ has a limiting normal distribution. Then, we use Theorem 3.4.1 to conclude that $\sqrt{N}(\hat{\delta} - \delta)$ and $\sqrt{N}[\hat{\delta} - E(\hat{\delta})]$ share the same asymptotic distribution, and as a result, $\hat{\delta}$ is a consistent estimator for δ.

To characterize the asymptotic distribution of $\hat{\delta}_N$, we define

$$\begin{aligned} r(z_i) &= f(x_i^c, x_i^d)\frac{\partial g(x_i^c, x_i^d)}{\partial x^c} - [y_i - g(x_i^c, x_i^d)]\frac{\partial f(x_i^c, x_i^d)}{\partial x^c}, \text{ and} \\ r_N(z_i) &= E[p_N(z_i, z_j)|z_i]. \end{aligned} \tag{3.16}$$

Then, we begin by setting an additional assumption and stating a useful lemma:

Assumption 14. $Nh^{k_c+2} \to \infty$ *as* $N \to \infty$.

Lemma 3.4.1. *Given Assumptions 5-14, it follows that* $N^{-1}E[\|p_N(z_i, z_j)\|^2]$ *converges to zero.*

Before proving the asymptotic normality of $\hat{\delta}_N$, it is useful to state a result regarding the asymptotic equivalence between two general second order U-statistics. We establish this result in the subsequent lemma:

Lemma 3.4.2. *Suppose that* $N^{-1}E[\|p_N(z_i, z_j)\|^2]$ *converges to zero, and define*

$$\hat{U} = E[r_N(z_i)] + \frac{2}{N}\sum_{i=1}^{N}\{r_N(z_i) - E[r_N(z_i)]\}.$$

Then, $\sqrt{N}(\hat{\delta} - \hat{U})$ *converges in probability to zero.*

Due to this lemma and since $E(\hat{\delta}_N) = E[r_N(z_i)]$, it is equivalent to characterize the asymptotic distribution of $\sqrt{N}[\hat{\delta}_N - E(\hat{\delta}_N)]$ or $(2/\sqrt{N})\sum_i\{r_N(z_i) - E[r_N(z_i)]\}$.

Lemma 3.4.3. *Under Assumptions 5-14, it follows that*

$$\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\{r_N(z_i) - E[r_N(z_i)]\} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where $\Sigma = 4E[r(z_i)r(z_i)' - \delta\delta']$.

Combining together previous lemmas, we can state the next concluding theorem:

Theorem 3.4.2. *Given Assumptions 5-14,* $\sqrt{N}(\hat{\delta} - \delta)$ *has a limiting normal distribution with mean zero and covariance matrix* Σ*, and consequently,* $\hat{\delta}$ *is a consistent estimator for* δ*.*

3.4.3 Estimation of the Asymptotic Covariance Matrix

In addition to establishing asymptotic normality, Theorem 3.4.2 suggest a natural estimator for the asymptotic covariance matrix Σ. In particular, using its sample analogue, this matrix can be consistently estimated as follows

$$\hat{\Sigma} = \frac{4}{N}\sum_{i=1}^{N}[\hat{r}(z_i)\hat{r}(z_i)' - \hat{\delta}\hat{\delta}'], \tag{3.17}$$

where $\hat{r}(z_i) = \frac{-1}{N-1}\sum_{j \neq i} p_N(z_i, z_j)$.

Essentially, the consistency of $\hat{\Sigma}$ is established in the lemma stated below:

Lemma 3.4.4. *Under Assumptions 5-14,* $\hat{\Sigma}$ *converges in probability to* Σ*.*

Hypothesis tests can now be performed with standard Wald statistics using $\hat{\delta}$ and $\hat{\Sigma}$. Basically, if R is a $k_1 \times k$ full rank matrix with $k_1 \leq k$, and $R\delta = \delta_\circ$ represents the coefficient restrictions involved in the null, then the limiting distribution of $N(R\hat{\delta} - \delta_\circ)'(R\hat{\Sigma}R')^{-1}(R\hat{\delta} - \delta_\circ)$ is χ^2 with k_1 degrees of freedom.

Table 3.1: Monte Carlo Experiments

True Values				Mean (Std. Deviation)		
β_2	α	E_0	E_1	$\hat{\beta}_2$	$\hat{E}_0$	$\hat{E}_1$
0.00	0.00	2.718	2.718	0.802	7.441	7.202
				(1.168)	(3.199)	(2.856)
0.00	0.50	2.718	4.482	-1.766	13.411	23.875
				(56.405)	(113.759)	(113.794)
0.00	1.00	2.718	7.389	0.682	26.109	151.562
				(1.451)	(287.831)	(1961.234)
0.50	0.00	4.482	4.482	0.039	11.211	11.015
				(13.869)	(17.101)	(16.130)
0.50	0.50	4.482	9.488	0.996	12.655	33.448
				(2.573)	(11.605)	(60.490)
0.50	1.00	4.482	20.086	0.913	21.551	111.536
				(0.889)	(64.401)	(468.691)

3.5 Monte Carlo Experiments

In order to evaluate the finite sample behavior of the estimator presented herein, this section reports the results of Monte Carlo investigations.

In the simulations, $k_c = 2$, $k_d = 1$, and $N = 250$. The dependent variable y_i is generated from a nonlinear specification as

$$y_i = \exp[(\beta_1 x^c_{1,i} + \beta_2 x^c_{2,i})(1 + \gamma x^d_i)] + \varepsilon_i, \tag{3.18}$$

where $\{(x^c_{1,i}, x^c_{2,i}, x^d_i, \varepsilon_i) : i = 1, ..., N\}$ is a sample of independent and identically distributed pseudo-random vectors. The components of $(x^c_{1,i}, x^c_{2,i}, x^d_i, \varepsilon_i)$ are independent of each other. Each component of $(x^c_{1,i}, x^c_{2,i})$ is distributed as $N(1, 2)$, x^d_i equals zero or one with probability 1/2, whereas the error term ε_i is distributed as $N(0,1)$. The first coefficient β_1 is equal to one by scale normalization and it is held constant across designs. We assume that ε_i is not observed, thus, the estimator is based only on the sample $\{(y_i, x^c_{1,i}, x^c_{2,i}, x^d_i) : i = 1, ..., N\}$. Given this framework, we investigate the finite sample performance of our estimator, and also, we study its effect on the estimation of $E(y|x^c_1, x^c_2, x^d)$, i.e., the conditional expectation of y given (x^c_1, x^c_2, x^d).

Using first $\hat{\delta}$ from expression (3.9), we have estimated $\delta = \beta E\{f(x^c_1, x^c_2, x^d)(1 + \gamma x^d_i)\exp[(\beta_1 x^c_{1,i} + \beta_2 x^c_{2,i})(1 + \gamma x^d_i)]\}$ where $\beta = (\beta_1, \beta_2)'$. Following closely [HH96], we used the fourth-order kernel $K(u_1, u_2) = k(u_1)k(u_2)$ with $k(u) = (105/64)(1 - 5u^2 + 7u^4 - 3u^6)\mathbb{1}(|u| \leq 1)$. In the absence of a theoretical guidance, we employed a simple selection bandwidth procedure that satisfies Assumptions 13 and 14, specifically, $h = 2n^{-1/(kc+3.5)}$ and $\lambda = n^{-2}$. Once obtained $\hat{\delta}$, we have then estimated β_2 by setting $\hat{\beta}_2 = \hat{\delta}_2/\hat{\delta}_1$.

Second, noting that $E(y|x^c_1, x^c_2, x^d)$ depends on (x^c_1, x^c_2) only through $(x^c_1 + \beta_2 x^c_2)$, we estimated $E(y|x^c_1, x^c_2, x^d)$ for $(x^c_1, x^c_2, x^d) = (1,1,0)$ and $(1,1,1)$ as follows

$$\hat{E}_{x^d} \overset{\text{def}}{=} \hat{E}(y|1,1,x^d) = \frac{\sum_{i=1}^N y_i k_e\{[(1+\hat{\beta}_2) - (x^c_{1,i} + \hat{\beta}_2 x^c_{2,i})]/h_e\}\lambda_e^{[1-\mathbb{1}(x^d - x^d_i)]}}{\sum_{i=1}^N \{k_e[(1+\hat{\beta}_2) - (x^c_{1,i} + \hat{\beta}_2 x^c_{2,i})]/h_e\}\lambda_e^{[1-\mathbb{1}(x^d - x^d_i)]}}, \quad (3.19)$$

where k_e denotes the standard normal kernel, and $\mathbb{1}(u)$ is an indicator function which equals one when $u = 0$, and zero otherwise. The bandwidths were $h_e = \sqrt{2}n^{-1/5}$ and $\lambda_e = n^{-2/5}$. For further references about kernel estimation with mixed data, see [LR07].

There were 500 replications in each experiment. Table 3.1 shows the empirical means and standard deviations of $\hat{\beta}_2$, $\hat{E}_0$ and $\hat{E}_1$. First, regarding the estimation of β_2, the estimator $\hat{\beta}_2$ has performed quite well in the contexts studied herein. Except for the cases $(\beta_2, \alpha) = (0.00, 0.50)$ and $(0.50, 0.00)$, we observe a small positive bias and standard deviations are not too large. Second, the proposed estimator for the conditional expectation, derived in (3.19), has had a poor perform. In particular, results have shown a very large positive bias.

3.6 Concluding Remarks

In order to estimate β in the model $E(y|x^c, x^d) = G(x^{c\prime}\beta, x^d)$, I have constructed an estimator for $\delta = E[f(x^c, x^d)\partial g(x^{c\prime}\beta, x^d)/\partial x^c]$ where $g(x^c, x^d) = E(y|x^c, x^d)$. Under standard regularity conditions, the resulting estimator is $\sqrt{N}$-consistent and asymptotically normal. Moreover, it can be computed directly from the data and it does not require to solve difficult nonlinear optimization problems.

The proposed estimator can be useful for estimating binary response models, which include discrete regressors, and for recovering the average treatment effect using a semiparametric approach. Monte Carlo experiments have shown that the estimator may perform well in small samples. A shortcoming of this paper is the absence of theory regarding the bandwidth choice, so as a topic for further research, I suggest addressing future investigations in that direction.

References

[AH07] S. Athey and P. A. Haile. "Nonparametric approaches to auctions." In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 6A, chapter 60, pp. 3847–3965. Elsevier, 2007.

[AHS11] D. Ackerberg, K. Hirano, and Q. Shahriar. "Identification of time and risk preferences in buy price auctions." Working Paper, University of Arizona, 2011.

[ALS11] S. Athey, J. Levin, and E. Seira. "Comparing open and sealed bid auctions: evidence from timber auctions." *Quarterly Journal of Economics*, **126**:207–257, 2011.

[BCK07] R. Blundell, X. Chen, and D. Kristensen. "Semi-nonparametric IV estimation of shape-invariant Engel curves." *Econometrica*, **75**(6):1613–1669, 2007.

[BS12] H. J. Bierens and H. Song. "Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method." *Journal of Econometrics*, **168**:108–119, 2012.

[CE07] A. Cohen and L. Einav. "Estimating risk preferences from deductible choice." *American Economic Review*, **97**(3):745–788, 2007.

[CGP11] S. Campo, E. Guerre, I. Perrigne, and Q. Vuong. "Semiparametric estimation of first-price auctions with risk-averse bidders." *Review of Economic Studies*, **78**:112–147, 2011.

[Che07] X. Chen. "Large sample sieve estimation of semi-nonparametric models." In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 76, pp. 5549–5632. Elsevier, 2007.

[Che08] B. Chen. "On effective minimax payoffs and unequal discounting." *Economics Letters*, **100**(1):105–107, 2008.

[CHT07] V. Chernozhukov, H. Hong, and E. Tammer. "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, **75**(5):1243–1284, 2007.

[CS98] X. Chen and X. Shen. "Sieve extremum estimates for weakly dependent data." *Econometrica*, **66**(2):289–314, 1998.

[Del08] M. R. Delgado. "Understanding overbidding: using the neural circuitry of reward to design economic auctions." *Science*, **321**:1849–1852, 2008.

[DL93] R. A. DeVore and G. G. Lorentz. *Constructive Approximation.* Springer-Verlag, 1993.

[DP97] L. Dechevsky and S. Penev. "On shape-preserving probabilistic wavelet approximators." *Stochastic Analysis and Applications*, **15**(2):187–215, 1997.

[FS09] Y. Fong and J. Surti. "The optimal degree of cooperation in the repeated Prisoners' dilemma with side payments." *Games and Economic Behavior*, **67**(1):277–291, 2009.

[Gab67] V. N. Gabushin. "Inequalities for the norms of a function and its derivatives in metric L_p." *Matematicheskie Zametki*, **1**(3):291–298, 1967.

[GLT11] Y. Guéron, T. Lamadon, and C. D. Thomas. "On the folk theorem with one-dimensional payoffs and different discount factors." *Games and Economic Behavior*, **73**:287295, 2011.

[GPV00] E. Guerre, I. Perrigne, and Q. Vuong. "Optimal nonparametric estimation of first-price auctions." *Econometrica*, **68**(3):525–574, 2000.

[GPV09] E. Guerre, I. Perrigne, and Q. Vuong. "Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions." *Econometrica*, **77**(4):1193–1227, 2009.

[GS12] E. Guerre and C. Sabbah. "Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function." *Econometric Theory*, **28**:87–129, 2012.

[Har89] J. Harrington. "Collusion among asymmetric firms: The case of different discount factors." *International Journal of Industrial Organization*, **7**(2):289–307, 1989.

[HH96] J. L. Horowitz and W. Hardle. "Direct semiparametric estimation of single-index models with discrete covariates." *Journal of the American Statistical Association*, **91**(436):1632–1640, 1996.

[HMZ10] A. Hu, S. A. Matthews, and L. Zou. "Risk aversion and optimal reserve prices in first- and second-price auctions." *Journal of Economic Theory*, **145**:1188–1202, 2010.

[HP95] K. Hendricks and H. J. Paarsch. "A survey of recent empirical work concerning auctions." *Canadian Journal of Economics*, **28**(2):403–426, 1995.

[HP07] K. Hendricks and R. Porter. "An Empirical Perspective on Auctions." In M Armstrong and R. Porter, editors, *Handbook of Industrial Organization*, volume 3, chapter 32, pp. 2073–2143. Elsevier, 2007.

[Ich93] H. Ichimura. "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models." *Journal of Econometrics*, **58**(1):71–120, 1993.

[IL91] H. Ichimura and L. Lee. "Semiparametric least squares estimation of multiple index models: single equation estimation." In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 3–49. Cambridge University Press, 1991.

[Laf97] J. Laffont. "Game theory and empirical economics: the case of auction data." *European Economic Review*, **41**(1):1–35, 1997.

[LP99] E. Lehrer and A. Pauzner. "Repeated games with differential time preferences." *Econometrica*, **67**(2):393–412, 1999.

[LP08] J. Lu and I. Perrigne. "Estimating risk aversion from ascending and sealed-bid auctions: the case of timber." *Journal of Applied Econometrics*, **23**:871–896, 2008.

[LPV03] T. Li, I. Perrigne, and Q. Vuong. "Semiparametric estimation of the optimal reserve price in first-price auctions." *Journal of Business and Economic Statistics*, **21**(1):53–64, 2003.

[LR07] Q. Li and J. S. Racine. *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, 2007.

[LRW08] Q. Li, J. S. Racine, and J. M. Wooldridge. "Estimating average treatment effects with continuous and discrete covariates: The case of swan-ganz catheterization." *American Economic Review: Papers & Proceedings*, **98**(2):357–362, 2008.

[LRW09] Q. Li, J. S. Racine, and J. M. Wooldridge. "Efficient estimation of average treatment effects with mixed categorical and continuous data." *Journal of Business & Economic Statistics*, **27**(2), 2009.

[Lu04] J. Lu. "Stochastic private values in auctions: identification and estimation." Working Paper, University of Southern California, 2004.

[Mat94] R. L. Matzkin. "Restrictions of Economic Theory in Nonparametric Methods." In R. F. Engle and D.L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 42, pp. 2523–2558. Elsevier, 1994.

[Mat03] R. L. Matzkin. "Nonparametric estimation of nonadditive random functions." *Econometrica*, **71**(5):1339–1375, 2003.

[MR84] E. Maskin and J. Riley. "Optimal auctions with risk averse buyers." *Econometrica*, **52**(6):1473–1518, 1984.

[MS12] V. Marmer and A. Shneyerov. “Quantile-based nonparametric inference for first-price auctions.” *Journal of Econometrics*, **167**:345–357, 2012.

[OLR09] D. Ouyang, Q. Li, and J. S. Racine. “Nonparametric estimation of regression functions with discrete regressors.” *Econometric Theory*, **25**(1):1–42, 2009.

[PHH06] H. J. Paarsch, H. Hong, and M. R. Haley. *An Introduction to the Structural Econometrics of Auction Data*. MIT Press, 2006.

[PSS89] J. L. Powell, J. H. Stock, and T. M. Stoker. “Semiparametric estimation of index coefficients.” *Econometrica*, **57**(6):1403–1430, 1989.

[PV99] I. Perrigne and Q. Vuong. “Structural econometrics of first-price auctions: a survey of methods.” *Canadian Journal of Agricultural Economics*, **47**(3):203–223, 1999.

[RL04] J. Racine and Q. Li. “Nonparametric estimation of regression functions with both categorical and continuous data.” *Journal of Econometrics*, **119**(1):99–130, 2004.

[Sto82] C. J. Stone. “Optimal Global Rates of Convergence for Nonparametric Regression.” *Annals of Statistics*, **10**(2):1040–1053, 1982.

[Sug13] T. Sugaya. “Characterizing the limit set of PPE payoffs with unequal discounting.” Working Paper, Stanford GSB, 2013.

[Tan11] X. Tang. “Bounds on revenue distributions in counterfactual auctions with reserves prices.” *RAND Journal of Economics*, **42**(1):175–203, 2011.

[Tir88] J. Tirole. *The Theory of Industrial Organization*. MIT Press, 1988.