The Pennsylvania State University
The Graduate School
College of the Liberal Arts

STRUCTURAL
ANALYSIS OF AUCTION MODELS

A Dissertation in
Economics
by
María Florencia Gabrielli

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009

The dissertation of María Florencia Gabrielli was reviewed and approved* by the following:

Joris Pinkse
Professor of Economics
Dissertation Adviser, Chair of Committee

Herman J. Bierens
Professor of Economics

Sung Jae Jun
Assistant Professor of Economics

Runze Li
Professor of Statistics

Vijay Krishna
Distinguished Professor of Economics
Director of Graduate Studies

*Signatures are on file in the Graduate School.

Abstract

The analysis of auctions is an active area of research for both theoretical and empirical economists. Concentrating on a particular auction format, namely the First–Price Sealed–Bid auction, this dissertation contributes to the analysis of auction both from a methodological viewpoint and from a more applied perspective.

The first chapter of this dissertation provides and overview of the literature relevant to this thesis. There, we review the theoretical background to auction models, collusion and the econometric literature on nonparametric estimators.

Chapter 2 proposes a semiparametric estimator for the distribution of private values within the class of indirect methods that have been analyzed using the so–called structural approach to estimate auction models. The proposed estimator is shown to have desirable statistical properties namely, it is consistent and has an asymptotic normal distribution. Moreover, the estimator attains the parametric rate of convergence.

Chapter 3 concentrates on the study of collusion in auctions. The main objective of chapter 3 is twofold. First, to provide a methodology to detect collusion using a structural approach, and second to apply the methodology to field data on procurement auctions for highway construction in California.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am extremely indebted to my adviser Prof. Joris Pinkse. Expressing my gratitude towards him would take longer than this entire thesis. Not only has he been a great guidance, but also a great source of inspiration. Prof. Pinkse has provided invaluable comments, continuous advice and illuminating points which have been of greatest help for me while preparing this dissertation. I am also deeply grateful to Prof. Herman J. Bierens and Asst. Prof. Sung Jae Jun for their insightful guidance, constant encouragement and for always be willing to help me. I would like to thank Prof. Runze Li for all the help and agreeing to be on the committee. I am thankful to Prof. Isabelle Perrigne, and Prof. Quang Vuong for their help in the second chapter of this dissertation. I especially thank Prof. Quang Vuong for always finding time to work on this chapter. It would have been an impossible task without him.

I am also grateful to Dr. George McCandless, for encouraging me to pursue graduate studies in the U.S. and Prof. Edward Green for his constant support throughout all these years.

Special thanks go to the Dean family for being my second family here in the U.S. I would like to thank my friends here at Penn State, for all the support and encouragement. In particular, Merih Danali, Camilo Rubbini, Vanessa Anichiárico, Alonso Alfaro–Ureña, Luz Monge, Manaswini Bhalla, Kerem Çosar, Hernán Bejarano, Çan Uz, Teague and Liz Ruder and Gaurab Aryal have been extremely generous with me. Also, my friends from Argentina have always been with me through the highs and lows throughout these years. Thank you to Evelina Bertranou, Eugenia Yañez, M. Pia Olaguer–Feliú, Gabriela Scalia, Josefina Rouillet and Cecilia Todesca–Bocco.

Last, but by no means the least, I would like to thank my family for their love and support. In particular to my grandparents who always taught me by example. My parents, Rolo and Cristina who are always there for me and have been my primary source of admiration. My sisters Eugenia, Lucía and Alicia, and my brother Federico, who are also my best friends. I would like to thank my uncle Jorge and my aunts Susana and Gladys for their confidence in me. Finally, I thank my husband, Carlos, for his immense patience and faith in my abilities. Also Norma and Carlos, my parents in law for their constant support. Financial support for this research is provided by the College of Liberal Arts, the Pennsylvania State University.

Dedication

To my dearest husband, Carlos, who always believed in me.

Chapter 1

Introduction

1.1 Overview

The analysis of auctions is an active area of research for both theoretical and empirical economists. In this dissertation we study a particular auction format, namely the First–Price Sealed–Bid auction. Chapter 2 proposes a semiparametric estimator for the distribution of private values within the class of indirect methods that have been analyzed using the so–called structural approach to estimate auction models. Therefore the main contribution of this chapter is methodological. The proposed estimator is shown to have desirable statistical properties; it is consistent and has an asymptotic normal distribution. Moreover, the estimator attains the parametric rate of convergence. Furthermore, it is not subject to a common problem encountered in the nonparametric literature referred as to the "curse of dimensionality". At a more technical level, the use of a Local Polynomial Estimator (LPE) in the estimation procedure prevents boundary effects (another problem with nonparametric estimation) from affecting the convergence rate of the estimator of interest and therefore it is not necessary to trim out observations. This is a remarkable advantage with respect to kernel methods.

Although the benchmark theoretical model in chapter 2 is one of the simplest auction models in the literature, i.e. the symmetric Independent Private Value (IPV) model, the methodology extends to a broader class of models. This is a direct consequence of the use of indirect methods to recover the distribution of private values.

Chapter 3 concentrates on the study of collusion in auctions. The main objective of chapter 3 is twofold. First, to provide a methodology to detect collusion using a structural approach, and second to apply the methodology to field data on procurement auctions for highway construction in California.

Building on previous theoretical work, we use an asymmetric IPV model to characterize the bidding behavior of firms competing for construction projects in a procurement–auction setup. we consider two possible scenarios. In the first model, firms are engaged in a competitive game. The second model incorporates the possibility that a subset of firms collude. The goal is then to identify which model best describes the behavior of bidding firms in the data. Relying on the assumption of exogenous participation, we are able to differentiate between the two models. The key observation is that under the correct specification, the underlying distributions of private costs should not change as the number of bidders change.

1.2 Auction Theory

At a theoretical level the most general paradigm identified in the literature to model auctions is the Affiliated Value (AV) model which is defined by the pair $[U(\cdot), F(\cdot)]$, where $U(\cdot)$ represents the utility function and $F(\cdot)$ denotes the distribution of the information bidders have. This information could be private to each bidder or common to all bidders. More precisely, $U_i = U(x_i, v)$ is the utility of a potential bidder i, $i = 1, \ldots, N$ for the object where x_i denotes the ith player's private signal or information and v represents a common component or value affecting all utilities. The utility, $U(\cdot)$, is increasing in both arguments. The vector $(x_1, \ldots, x_n, v)$ is a realization of a random vector whose $(N+1)$–dimensional cumulative distribution function is $F(\cdot)$. The latter is assumed to be affiliated with a support $[\underline{x}, \overline{x}]^n \times [v, v]$, $\underline{x} \geq 0$, and a density $f(\cdot)$.[1] Each bidder i knows the value of his signal, x_i, but neither the other signals x_j nor the common value v. On the other hand, the number N and the functions $U(\cdot)$ and $F(\cdot)$ are common knowledge.

Depending on the nature of the utility function and that of the information held by bidders, a further subclassification produces the Affiliated Private Value (APV) paradigm in which $U(x, v) = x$, and the General Common Value paradigm, where $U(x, v) = v$. Each model in turn is said to be symmetric if the function $F(\cdot)$ is symmetric in it first N arguments, otherwise the models are called asymmetric. A special case of the APV model is the Independent Private Value (IPV) model. In the case of symmetric bidders, this model takes the form $U_i = x_i$ with X_i independently and identically distributed as

[1]Roughly speaking, affiliation is a strong form of positive correlation. In an auction context, private signals $X_1, \ldots, X_N$ are affiliated if when a subset of the X_i's are all large, then this makes it more likely that the remaining X_j's are also large. For a formal definition see Krishna (2002).

$F(\cdot)$. For the asymmetric IPV model, the utility function is still given by $U_i = x_i$, but (in its simplest form) one bidder draws his valuation from a distribution that differs from that of other bidders. In this dissertation we mainly focus on the IPV case in both the symmetric and asymmetric versions.

1.2.1 The Symmetric IPV Model

A single and indivisible object is offered for sale to N potential buyers who bid in an auction where the highest bidder gets the object and pays the amount of his bid. Each bidder i assigns a value X_i to the object, the maximum amount the bidder is willing to pay for the object. It is assumed that for each i, $i = 1, \ldots, N$, X_i is distributed according to the increasing distribution function $F(\cdot)$ with support on $[\underline{x}, \overline{x}]$. $F(\cdot)$ is common knowledge and admits a continuous density $f(\cdot)$ on $[\underline{x}, \overline{x}]$. X_i is the private information of bidder i. In other words, bidder i observes a realization x_i of X_i and only knows that other bidders' valuations are independent draws from $F(\cdot)$. The fact that for all i, $F_i(\cdot) = F(\cdot)$ is referred to as a situation involving symmetric bidders. Let $s_i : [\underline{x}, \overline{x}] \to I\!R_+$ denote bidder i's strategy, which determines his bid for any private value. Given that bidders are symmetric, it is natural to concentrate on symmetric equilibria, i.e. $s_i = s$ for all i.

For a given bid, b_i, the payoffs for the ith bidder are given by

$$\Pi_i = \begin{cases} (x_i - b_i) & \text{if } b_i > \max_{j \neq i} b_j \\ 0 & \text{if } b_i < \max_{j \neq i} b_j. \end{cases} \tag{1.1}$$

where in case of a tie among two or more bidders it is assumed that the object goes to each bidder with equal probability. From the expected profit function (1.1) above it can be seen that no bidder would bid an amount equal to his value since in this case the payoff would be zero. Thus, in equilibrium bidders shade their valuations. Notice also that a bidder faces a trade–off at any bid holding constant the behavior of his rivals. Indeed, by increasing one's bid, the probability of winning also increases, but at the same time there is a reduction in the gains from winning. More formally, the optimization problem for each bidder is

$$\max_{b_i} (x_i - b_i) F(s^{-1}(b_i))^{N-1}$$

where $s^{-1}(\cdot)$ denotes the inverse strategy function and $F(s^{-1}(b_i))^{N-1}$ is the probability of winning.

Riley and Samuelson (1981) among others, have characterized the unique symmetric differentiable Bayesian Nash equilibrium. In particular, for $N \geq 2$ the equilibrium bid B_i is $s(X_i)$ with

$$s(x_i) = x_i - \frac{1}{F(x_i)^{N-1}} \int_{\underline{x}}^{x_i} F(v)^{N-1} dv, \tag{1.2}$$

for any $x_i \in [\underline{x}, \overline{x}]$ subject to the boundary condition $s(\underline{x}) = \underline{x}$. The expression in (1.2) shows that the equilibrium bid is less than the private value. Moreover, the degree of "shading" (the amount by which the bid is less than the value) depends on the number of competing bidders. In particular, the larger the number of competing bidders, the smaller the difference between bids and private values (i.e the degree of shading is less in auctions with more participating bidders).

1.2.2 The Asymmetric IPV Model

In the third chapter of this dissertation we study potential collusive behavior by a subset of firms competing in a procurement–auction setting. The underlying model considers a situation in which bidders are ex ante asymmetric. In this section we present the relevant theoretical background. To make the presentation of the model more transparent we consider the two bidder case here. However in chapter 3 the model allows for the more realistic case of more types of bidders. Although this section explains the theoretical framework for an auction model, in the empirical application of chapter 3 bidders compete for the right to sell their services and thus the winner is the bidder with the lowest bid. The main features of the auction model outlined here are the same for the procurement setting discussed later with the appropriate sign changes.

Asymmetries introduce numerous complications in first–price auctions. In particular, and despite the fact that an equilibrium of this game exists under some regular conditions, there is no closed form expression for the bidding strategies. This further complicates the econometric analysis of this kind of auctions. We explain in section 1.4 how the use of indirect methods of estimation are especially useful in this context. Another feature of the asymmetric first–price auction model is that the resulting allocation is not necessarily efficient; that is, the object may not end up in the hands of the bidder who values it the most.[2]

Let X_1 and X_2 be the private values for bidders 1 and 2, respectively. Each value is

[2] As is well known in the auction theory literature, the Revenue Equivalence Principle holds under the assumption of symmetry (see Krishna (2002) for further details).

independently distributed according to $F_1(\cdot)$ and $F_2(\cdot)$ on the common support $[0, \overline{x}]$.[3] These two distributions are common knowledge with corresponding densities $f_1(\cdot)$ and $f_2(\cdot)$ which are assumed to be continuously differentiable and bounded away from zero on their support.[4] Let $s_1(\cdot)$ and $s_2(\cdot)$ denote the equilibrium strategies for each bidder. Moreover assume that these are increasing and differentiable with inverses $\xi_1 = s_1^{-1}$ and $\xi_2 = s_2^{-1}$. The expected payoff of bidder 1, say, when bidder 2 follows the strategy $s_2(\cdot)$ is

$$\begin{aligned} \Pi_1(b, x_1) &= F_2(\xi_2(b))(x_1 - b) \\ &= G_2(b)(x_1 - b) \end{aligned} \tag{1.3}$$

where $G_2(\cdot) \equiv F_2(\xi_2(\cdot))$ denotes the distribution of bidder 2's bid.

Observe that $s_1(0) = 0 = s_2(0)$ since it would not be optimal for a bidder to bid more than the value. Moreover, $s_1(\overline{x}) = s_2(\overline{x})$ since otherwise if, say, $s_1(\overline{x}) > s_2(\overline{x})$ then bidder 1 would win with probability one when his value is $\overline{x}$ and would pay more than he needs to. Let $\overline{b} = s_1(\overline{x}) = s_2(\overline{x})$. These constitute the boundary conditions. Among others, Lebrun (1996, 1999), and Maskin and Riley (2000a,b, 2003) have studied the existence and uniqueness of the Bayesian–Nash equilibrium in asymmetric first–price, sealed–bid auctions. Differentiating (1.3) yields

$$g_2(b)(\xi_1(b) - b) = G_2(b) \tag{1.4}$$

with $g_2(b) \equiv G_2'(b) = f_2(\xi_2(b))\xi_2'(b)$ the density of bidder 2's bids. Thus, (1.4) can be written as

$$\xi_2'(b) = \frac{F_2(\xi_2(b))}{f_2(\xi_2(b))} \frac{1}{[\xi_1(b) - b]} \tag{1.5}$$

The solution to the system of differential equations (1.5) (i.e one equation per bidder) along with the boundary conditions constitute and equilibrium. As mentioned above, there is no explicit solution except for some special cases.[5]

---

[3] The assumption of a common support is for simplicity.

[4] This set of regularity conditions guarantee the existence and uniqueness of the equilibrium. See, e.g Lebrun (1996, 1999), and Maskin and Riley (2000a,b, 2003).

[5] To see an example which uses Uniform distributions see Krishna (2002).

1.3 Collusion

Although collusion is an illegal activity, it is a pervasive problem in auction markets. In this section we briefly discuss the relevant theoretical and empirical work that has been done in an attempt to better understand and detect this practice. As emphasized by Baldwin, Marshall, and Richard (1997), the profitability and prevalence of bid rigging call for the incorporation of the possibility of collusive behavior into empirical models.

Theoretical work on bidder collusion at auctions is extensive. The following description is by no means exhaustive. Robinson (1985) analyzes the relative propensity of the different auction formats, second–price, first–price, and English auctions, to collusion. In particular, the author shows the relative nonsusceptibility of first–price auctions to bidder collusion. The analysis of collusion in second–price auctions was initiated by Graham and Marshall (1987) in a symmetric IPV model. Graham, Marshall, and Richard (1990) extend previous results to the case of distributionally heterogeneous bidders. McAfee and McMillan (1992) analyze collusion in first price auctions by an all–inclusive coalition. They also study a pre–auction knockout mechanism used by the cartel. As is well–known, asymmetric first–price auction models yield an equilibrium with no closed form.[6] Marshall, Meurer, Richard, and Stromquist (1994) look at less than all–inclusive cartels at first–price auctions and propose numerical techniques to solve for the equilibrium of the underlying asymmetric game. From a more policy–oriented view, Marshall and Meurer (2004) argue that the relative lack of attention given to bidder collusion is based on the mistaken belief that the economics of bidder collusion and that of price fixing are essentially equivalent. The authors illustrate the differences between standard industry posted–price cartels and collusion by bidders at auctions or procurements by means of several models and examples. Moreover, they propose policy recommendations that apply specifically to bidder collusion.

The empirical literature on bidder collusion is more limited. Hendricks and Porter's (1989) survey paper discusses mechanisms that are likely to facilitate collusion in auctions and propose some tests in order to detect bid rigging by analyzing two commonly used data sets within both the IPV framework and the Common Value (CV) framework. Porter and Zona (1993, 1999) and Pesendorfer (2000) concentrate on collusion in auction markets given that it is known that bid–rigging has taken place. The objective of these papers is basically to find empirical facts in collusive markets.

Another set of empirical papers proposes methods to detect collusion. Porter and

[6] Except for some special cases (see footnote 5).

Zona (1993), Baldwin, Marshall, and Richard (1997) and Bajari and Ye (2003) study collusion in IPV settings. The paper by Asker (2008) seeks to better understand the functioning of an operating cartel. Within the structural approach, the author examines a first–price knockout auction mechanism used by a cartel of stamp dealers in the 1990s.

Porter and Zona (1993) argue that detection of collusion is possible because of limited participation in the collusive setup. Accordingly, they attempt to detect differences in behavior between ring members and non–members. The authors have detailed information of the operation of a cartel and its bidding practices. In particular they study the bidding behavior of firms competing for highway construction projects on Long Island in the early 1980s. They propose two types of analysis. The first one is based on the level of bids and the second one on the ranking of bids. Accordingly, Porter and Zona (1993) argue that the evidence of collusive behavior relies on the fact that the lowest noncartel bidder's behavior is not statistically different from that of other noncartel firms, while the determinants of the low cartel bid differ from those of higher cartel bids. By knowing the identities of cartel members Porter and Zona (1993) estimate two models for each subgroup of bidders and test the null hypothesis of absence of collusion by testing the equality of parameter values in the models. The starting point for each (econometric) model is given by the first order conditions for an equilibrium strategy, namely

$$\varphi_{i\ell} + (b_{i\ell} - c_{i\ell})\varphi'_{i\ell}(b_{i\ell}) = 0 \tag{1.6}$$

where $b_{i\ell}$ is the submitted bid for firm i in project ℓ, $c_{i\ell}$ is the corresponding cost, and $\varphi_{i\ell}$ is the probability of winning.

Regarding the analysis based on the level of bids, Porter and Zona (1993) exploit the characterization of the equilibrium bid given by (1.6) and assume that equilibrium behavior satisfies the log–linear bidding rule

$$\log(b_{i\ell}) = \alpha_\ell + \beta' X_{i\ell} + \epsilon_{i\ell} \tag{1.7}$$

where α_ℓ is an auction–specific effect, $X_{i\ell}$ is a vector of observable variables affecting firm i's probability of winning object ℓ. In the empirical application the authors include the utilization rate, the firm's backlog and capacity and a dummy variable regarding the location of the firm. The error term, $\epsilon_{i\ell}$, represents private information for firm i on project ℓ. It is assumed to have zero expectation and an auction–specific variance, σ_ℓ^2.

By estimating the auction–specific variance using the auction mean–squared residual, the authors implement a feasible generalized least squares (GLS) estimator to obtain estimates for the parameters in (1.7). The reported results are given for three subsets of data: bids from all firms, bids from competitive firms and bids from cartel firms only. The authors conjecture that if all bids were competitive, the three subsets of data should give the same underlying parameters. On the other hand, if cartel bids were not competitive, then the model would be misspecified, and only the estimators based on competitive data would be consistent.

The two main conclusions from this analysis indicate that the model fits the competitive data reasonably well according to a Wald test and that bids from cartel firms statistically differ from those of competitive firms. The authors claim that the analysis based on the ranking of bids (i.e. the second kind of test proposed by the authors) sheds light on the reasons for this discrepancy.

To perform an analysis based on the ranking of bids, the main argument used by Porter and Zona (1993) states that fundamental differences may exist between the ordering of competitive and cartel bids conditional on observed data. This observation relies on the fact that firms submitting phantom bids know that a designated firm will submit a lower bid (recall that Porter and Zona (1993) study a procurement–auction). Thus, complementary bids have no probability of winning by design. The rationale for phantom bidding is just to create the appearance of competition. However, the designated cartel bid must bid competitively like the remaining noncartel firms. The authors do not explain how the designated bidder is selected.

From this observation, Porter and Zona (1993) propose a rank–based test intended to detect differences in the ordering of higher bids, as opposed to the determinants of the probability of being the lowest bid, for each set of firms. The conjecture in this case is that if cartel bids were indeed competitive, their ordering should reflect observable cost differences.

To implement the rank test the authors use equation (1.7) to characterize the probability of winning by approximating it with a multinomial logit (MNL) model as follows

$$\ln \mathrm{P}[b_{i\ell} < \min_{j \neq i} b_{ij}] = \theta_\ell + \beta' \frac{X_{i\ell}}{\sigma_t} \sqrt{\frac{\pi}{6}}$$

Let $-X_{i\ell}/\sigma_t \equiv Z_{i\ell}$. The MNL model giving the log probability that firm i will win auction ℓ is,

$$\ln \mathrm{P}[b_{i\ell} < \min_{j \neq i} b_{ij}] = \alpha_\ell + \beta' Z_{i\ell} \qquad (1.8)$$

By exponentiating the log probabilities, equation (1.8) can be expressed as

$$\mathrm{P}[b_{i\ell} < \min_{j \neq i} b_{ij}] = \frac{\exp(\beta' Z_{i\ell})}{\sum_j \exp(\beta' Z_{j\ell})}$$

Given the MNL specification chosen by the authors, the probability of observing any particular ranking of bids on a project can be expressed as the product of individual choice probabilities. If n_ℓ bids are submitted on job ℓ, $\ell = 1, \ldots L$, the likelihood of observing the rankings of the data from all auctions in the sample is

$$L(\beta) = \prod_{t=1}^{L} \prod_{i=1}^{n_\ell} \frac{\exp(\beta' Z_{r_i \ell})}{\sum_j \exp(\beta' Z_{r_j \ell})}$$

where r_m denotes the index of the firm with bid ranked m (see Porter and Zona (1993) for further details).

The model is estimated using standard maximum likelihood (ML) estimation for MNL. If the model is correctly specified, the parameters can be estimated from any subset of the data. To test the hypothesis of no phantom bidding the authors use a Likelihood Ratio (LR) test. In other words, the null hypothesis states that the parameters estimated using only the lowest cartel ranks and those estimated from higher cartel ranks should be the same. As pointed out in the paper, rejection of the null hypothesis could be because of two reasons. First, the model may be misspecified for some reason other than phantom bidding. However, the authors argue that if the test did not reject the null when applied to competitive data, then it is less likely to have a specification problem. The second reason leading to rejection could be due to an effect that is common to non–winning cartel bids but not non–winning competitive bids. Porter and Zona (1993) conclude that under the assumptions of the model, the rejection is likely to be the result of phantom bidding.

The main conclusion drawn from the analysis of competitive bid rank data states that these bids are generated by the same process whether or not they are low. In other words, the estimates are stable over ranks and the LR test does not lead to rejection of the null hypothesis of no model misspecification. The second analysis based on cartel bid rank data yields the opposite conclusion, namely that cartel bids are generated by a different process depending on whether or not they are low.

Finally, the authors conclude that they have found evidence supporting cartel activity in the sample since they do not have reason to believe that the difference between cartel and competitive bidding is structural.

Bajari and Ye (2003) propose a model in which bidders are asymmetric in a procurement first–price–auction setup. The authors derive two conditions that must hold under competitive bidding, namely, conditional independence of bids and exchangeability of bid distributions. They also propose a third test based on Bayesian techniques which requires inside information from the industry. Bajari and Ye (2003) apply their tests to a data set on seal coat contracts in the Midwest.

In chapter 3 we propose a model for collusive behavior and compare it to a model of competitive behavior. Since there is no information about the identities of potential cartel members we use the first two tests proposed in Bajari and Ye (2003) to identify pairs of firms which might constitute a cartel. We now explain this set of conditions in more detail.

Let $Z = (Z_1, Z_2, \ldots, Z_N)$ denote a set of covariates that is observable to all firms. Let $G_i(\cdot; Z)$ be the cumulative distribution function of firm i's bid given covariates. Observe that the distribution of bids depends on the entire vector Z.

Conditional on $Z = z$, firm i's bid and firm j's bid are independently distributed. As a result

$$G(b_1, \ldots, b_N; z) = \prod_{i=1}^{N} G_i(b_i; z), \tag{1.9}$$

where $G(b_1, \ldots, b_N; z)$ is the joint distribution of bids. As mentioned in Bajari and Ye (2003), there is more than one way of testing this condition. Ideally each side of (1.9) can be estimated nonparametrically and compared. However with limited data this becomes less attractive. Alternatively, regression–based methods can be used. That is the marginal distribution, $G_i(b_i; z)$, can be estimated using a regression (see also Porter and Zona (1993, 1999)), and then the residuals are tested to assess if they are independent.

The second condition that must hold in equilibrium when bidding is competitive is referred to as exchangeability of the distribution of bids. More formally, let π be a permutation, that is, a one–to–one mapping from the set $\{1, \ldots, N\}$ onto itself. Then exchangeability is defined as follows: for any permutation π and any index i the following equality must hold

$$G_i(b; z_1, z_2, z_3, \ldots, z_N) = G_{\pi(i)}(b; z_{\pi(1)}, z_{\pi(2)}, z_{\pi(3)}, \ldots, z_{\pi(N)})$$

Like for conditional independence, regression–based methods can be used to test this condition. This is the approach taken in Bajari and Ye (2003).

The papers by Baldwin, Marshall, and Richard (1997) and Asker (2008), are related to the third chapter in this dissertation since both of them use a structural approach to analyze auction data.[7] There are, nevertheless, important differences with the approach taken in chapter 3 of this dissertation. Baldwin, Marshall, and Richard's (1997) data set comes from oral ascending auctions. Therefore, the authors concentrate on this auction format to derive the econometric models used in the application. Moreover, one of the maintained assumptions in that paper is distributional homogeneity across bidders' valuations (i.e. bidders are symmetric). At a more technical level, the empirical model is fully parametric. The underlying theoretical model in chapter 3 deals with an asymmetric first–price sealed–bid auction within the IPV framework and the resulting econometric model is nonparametric. On the other hand, the work by Asker (2008) considers asymmetries across bidders. However, its main objective is to analyze the functioning of a cartel as opposed to study the main auction. The knockout auction is conducted using a sealed–bid format. The author focuses in modeling the pre–auction knockout mechanism used by the ring to designate the serious bidder at the main auction. The econometric procedure used is fully nonparametric.

Baldwin, Marshall, and Richard (1997) formulate various empirical models using the structural approach allowing for both bidder collusion and supply effects in order to analyze auctions for timber in the Pacific Northwest. The main objective is to determine whether price variations, conditional on demand characteristics, are better explained by collusion or, alternative, variations in timber supply conditions. The authors provide some evidence revealing the similarity between the winning bid and the reserve price in timber auctions during the late 1970s and early 1980s. This information motivates the following observation stated in the paper. *"Although effective bidder collusion will produce winning bids that are low relative to the predictions of a suitable model of noncooperative behavior, it clearly would not be reasonable to conclude that bidders are colluding solely on the basis of the observation of relatively low winning bids"*. Accordingly, five models are estimated: the noncooperative model with no supply effects, the collusive model with no supply effects, the noncooperative model with supply effects and two nested models that contain both collusion and supply effects.

We concentrate here on the collusive model with unit supply since it is the one most closely related to the work presented in chapter 3. The underlying theoretical

[7]In the following section we explain in more details the structural approach used in the literature as opposed to reduced form approach.

model that leads to the empirical model is based on Graham and Marshall's (1987) collusive mechanism. In this pre–auction knockout, colluding bidders find participation individually rational. The effective coalition size is denoted by K_i. Conditional on K_i, the price of the object is given by the $K_i + 1$st order statistic of the private values. Two important assumptions behind the model are that bidder collusion is a (symmetric) bidder–specific decision and that there is only one coalition. Thus all nonring bidders act noncooperatively. Another important element of the model is the probability of joining the coalition, p_i. Conditional on $Z_i = z_i$ the natural logarithm of private values is assumed to be normally distributed with mean $\beta' z_i$ and variance σ^2, where z_i is a vector of covariates. The standardized price, $U_i = (\ln(B_i/v_i) - \beta' z_i)/\sigma$, is a mixed random variable with $\Pr(U_i = t_i) = p_i^{n_i}$ and density $h_c(\cdot|z_i)$, where B_i is the winning bid at auction i, v_i is the volume in mbf and $t_i = (\ln(r_i/v_i) - \beta' z_i)/\sigma$, with r_i denoting the reserve price.[8] The authors specify the likelihood of the collusive model as well as a parametric expression for p_i. Let C_i be the coalition participating at auction i and $y_i = (\ln[1.055(r_i/v_i)] - \beta' z_i)/\sigma$. The likelihood function is

$$L(\beta, \gamma, \sigma; d) = \left[\prod_{i \in I_a} \frac{v_i}{\sigma b_i} h_c(u_i|z_i)\right] \left[\prod_{i \in I_b} H_c(y_i|z_i)\right]$$

where $I = I_a \cup I_b$ denotes the total number of observations. I_b is the set containing 13 observations in which the winning bid is within 5.5% of the reserve price, the remaining observations belong to I_b.[9] γ is a parameter from the expression for p_i and d denotes the data set.

The model is estimated using standard ML techniques. The main conclusion from this analysis is that the collusive model outperforms the noncooperative model. Moreover, the authors highlight that both models pass the Kolmogorov–Smirnov test. Thus, lognormality is not rejected. This leads the authors to the further conclusion that the increase in the log likelihood function observed in the collusive case is not due to misspecification of the private value distribution. With respect to the model containing supply effects and the nesting models, the authors conclude that as soon as collusion is taken into account, supply does not add explanatory power. Overall, the collusive model emerges as the preferred model.

[8] In their paper Baldwin, Marshall, and Richard (1997) provide the explicit forms for the density and cumulative distribution functions of U_i.

[9] The 13 observations in the set I_b are considered as outliers. However, Baldwin, Marshall, and Richard (1997) argue that it would be inappropriate to discard them from the estimation. See section VI.C and Appendix D of that paper for a detailed discussion.

1.4 Econometrics of Auctions

Chapters 2 and 3 of this dissertation contribute to the empirical analysis of auction data within the structural approach. Here we explain more in detail the relevant literature on this approach.

The availability of numerous data sets and the well–defined game forms associated with auctions mechanisms makes "econometrics of auctions" a particularly interesting field within economics. There are three main approaches to analyze auction data. The experimental approach, aimed to test the predictions of game–theoretic models, uses experimental data for controlling the underlying elements of the model. The reduced–form approach uses field data mostly relying on linear regressions of the logarithm of bids on a set of observed variables. These first two approaches were used almost exclusively until the late 1980s constituting the first step towards the formulation of empirical auction models. Most recently, a so–called structural approach has been developed. A detailed discussion and references to this literature is given by Perrigne and Vuong's (1999, 2008) survey papers. In particular these authors concentrate on first–price auctions within the private value paradigm.[10]

1.4.1 Structural Approach to Analyze Auction Data

Assuming that observed bids are the Bayesian Nash equilibria of the game–theoretical model under consideration, the structural approach provides a rich framework in which the theoretical model and its empirical counterpart are closely related. The main objective of this approach is to recover the structural elements of the auction model. This line of research has attracted considerable attention over the last fifteen years.

A first classification of methods for structurally estimating auction models distinguishes between direct methods and indirect methods.

Relying on parametric econometric models, direct methods were first developed in the literature. The starting point is to specify the underlying distribution of private values in order to estimate the parameter vector characterizing such a distribution. Within this class of methods, there are two major estimation procedures. The first methodology introduced by Paarsch (1992) and Donald and Paarsch (1993) is a fully parametric setup that uses ML–based estimation procedures requiring the computation of the equilibrium strategy. This in turn could be highly computationally demanding, as recognized by Donald and Paarsch (1993), and thus only very simple distributions are considered in

[10] See also Paarsch and Hong (2006) for an extensive survey on structural estimation of auction models within the IPV paradigm.

practice. In particular, because the upper bound of the bid distribution depends on the parameter(s) of the underlying distribution, the ML estimator has a nonstandard limiting distribution. In view of this, Donald and Paarsch (1993) develop a so–called piecewise pseudo ML estimator requiring the computation of the equilibrium strategy that can be obtained using specific parametric distribution(s). Marshall, Meurer, Richard, and Stromquist (1994) propose a set of numerical algorithms to solve for the equilibrium strategy of an asymmetric first–price auction allowing for arbitrary distributions of the private values.

Laffont, Ossard, and Vuong (1995) introduced a second methodology, which is more computationally convenient. Relying on the revenue equivalence theorem, the authors propose a simulation–based method that avoids computation of the equilibrium strategy and therefore allows for more general parametric specifications for the private value distribution. This method is called simulated nonlinear least squares (SNLLS) and is related to the methods proposed by McFadden (1989) and Pakes and Pollard (1989). The SNLLS estimator of Laffont, Ossard, and Vuong (1995) uses only the winning bid since the authors considered a descending oral auction of eggplants in their application. Li and Vuong (1997) extended the SNLLS estimator to the case in which all bids are observed, such as in sealed–bid auctions.

The use of indirect methods to structurally estimate auction models was introduced in Guerre, Perrigne, and Vuong (2000). The authors develop a fully nonparametric procedure for the structural estimation of auction models. This alternative methodology relies on a simple but crucial observation, namely that each private value can be expressed as a function of the corresponding bid, the distribution of observed bids and its density using the first–order condition of the bidder's optimization problem. Thus, in contrast to direct methods, the starting point of indirect procedures is the distribution of (observed) bids, which is used to recover the distribution of (unobserved) private values without computing the Bayesian Nash equilibrium strategy or its inverse explicitly

Based on the equation that defines the inverse of the equilibrium strategy, the authors show that the model is nonparametrically identified in an IPV framework. Other papers by the same authors and others consider other auction models in a similar fashion. Including the affiliated private value model, models with asymmetric bidders, dynamic auction models and model with risk averse bidders. Laffont and Vuong (1996) generalize the identification result in Guerre, Perrigne, and Vuong (2000) to symmetric APV models.[11]

The method in Guerre, Perrigne, and Vuong (2000) calls for a two step procedure. In the first step, a sample of pseudo private values is obtained while using nonparametric estimators for the distribution and density of observed bids. With this pseudo sample, the second step consists of estimating the density of bidders' private values nonparametrically. This estimator is shown to have desirable properties such us uniform consistency and the achievement of the optimal convergence rate by appropriate choice of vanishing rates for bandwidths.

In the same spirit as Guerre, Perrigne, and Vuong (2000) chapters 2 and 3 of this dissertation rely heavily on the use of nonparametric procedures. The following section discusses this kind of methods.

1.5 Nonparametric Estimators

For the last sixty years the statistical literature on nonparametric methods has developed considerably. This methodology proves to be especially useful in cases in which one has no precise information about the form and class of, e.g., the true density of a random variable. The histogram is one of the oldest nonparametric methods for density estimation; it has the disadvantage of being discontinuous and too "rough". There are several other methods available in this literature, such as kernels, splines, nearest–neighbor and local polynomials (see Härdle (1991) and Pagan and Ullah (1999) for a comprehensive discussion). By far, the most widely used is the kernel method.

Every method has some cost associated with it. The major problem cited in the literature faced by nonparametric procedures is the "curse of dimensionality". The precision of the estimator (exponentially) deteriorates when the number of variables increases. That is, large data sets are needed to get accurate estimators.

1.5.1 Kernel Density Estimators

The idea behind the estimation of a density, $f(\cdot)$, using kernel estimation is very similar to the histogram. However, in kernel estimation one averages over kernel functions instead of averaging over data points.

More formally, let $X_1, X_2 \dots$ be a sequence of random vectors in $I\!R^p$, where each X_i is distributed as $F(\cdot)$ with density $f(\cdot)$. A *kernel density estimator* of the density $f(\cdot)$

[11] Laffont and Vuong (1996) explicitly acknowledge that their identification result is a generalization of a result in a previous version of Guerre, Perrigne, and Vuong (2000).

at $x \in I\!R^p$ is defined as

$$\hat{f}(x) = \frac{1}{nh^p} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where $h > 0$ is a bandwidth and $K(\cdot)$ is a kernel (see Parzen (1962) and Rosenblatt (1956)). The bandwidth h is a smoothing parameter which regulates the degree of smoothness of the estimator.

The following assumptions are sufficient to obtain a pointwise consistent estimator,

1. The kernel $K(\cdot)$ is a bounded, even and integrable function from $I\!R^p$ to $I\!R$ with $\int K(x)dx = 1$. In particular, because $K(\cdot)$ is bounded, $\hat{f}(x)$ has finite moments of any order.

2. The random vectors $X_1, X_2 \ldots$ are independent and identically distributed.

3. $f(\cdot)$ is continuous at x.

4. $K(\cdot)$ is a Parzen–Rosenblatt kernel. More precisely, $K(\cdot)$ satisfies

$$\lim_{\|x\| \to \infty} \|x\|^p K(x) = 0,$$

   where $\| \cdot \|$ is the Euclidean norm on $I\!R^p$. For instance, a function $K(\cdot)$ with bounded support satisfies this condition.[12]

5. $h = h_n$, where $\{h_n\} \subset I\!R_+$ is a nonstochastic sequence satisfying

   (a) $h_n \to 0$ and

   (b) $nh^p \to +\infty$, as $n \to +\infty$.

The kernel estimator $\hat{f}(\cdot)$ has (further) desirable statistical properties. The asymptotic distribution of the estimator has been derived under additional regularity conditions. Also, uniform consistency has been established by strengthening some of the underlying assumptions. See e.g., Silverman (1986) or Bierens (1983) for a rigorous treatment and proofs of all these results mentioned above and the assumptions needed in each case.

As is well–known in the statistical literature, nonparametric estimators attain a lower convergence rate than parametric ones. For the kernel density estimator the best

[12] This assumption can be weakened if one is willing to make global assumptions on $f(\cdot)$ such as it is bounded over $I\!R^p$.

(uniform) rate of convergence has been established by Stone (1982) and it is given by $r^* = (n/\log n)^{R/(2R+p)}$, where R is the "degree" of smoothness of $f(\cdot)$.

Next, we provide expressions for the bias and variance of this estimator.

Bias of $\hat{f}(x)$:

$$\mathrm{E}[\hat{f}(x)] - f(x) = \frac{h^R}{R!}\left[\int M_x^{(R)}(0,u)K(u)du + o(1)\right]$$

where $\int M_x^{(R)}(0,u)K(u)du = (-1)^R \sum_{i_1=r}^{p} \cdots \sum_{i_r=1}^{p} \frac{\partial^r f(x)}{\partial x_{i_1} \ldots \partial x_{i_r}}$.

Variance of $\hat{f}(x)$:

$$\mathrm{Var}[\hat{f}(x)] = \frac{1}{nh^p}\left[f(x)\int K^2(u)du + o(1)\right].$$

1.5.2 Local Polynomial Fitting

In this section we describe the main characteristics of local polynomial estimators which are used in chapter 2 of this dissertation.[13] In line with chapter 2, we present the estimator for the case of one–dimensional explanatory variables $X_1, \ldots, X_n$.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent and identically distributed sample from a population (X, Y). The objective is to estimate the regression function $m(x_0) = \mathrm{E}(Y|X = x_0)$ and its derivatives $m'(x_0)$, $m''(x_0), \ldots, m^{(\rho)}(x_0)$. Assume that the $(\rho+1)$st derivative of $m(\cdot)$ exists at the point x_0. Using a Taylor expansion for x in a neighborhood of x_0, the regression function $m(\cdot)$ can be locally approximated by a polynomial of order ρ, that is

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{m^{(\rho)}(x_0)}{\rho!}(x - x_0)^\rho \quad (1.10)$$

In terms of a weighted least squares (LS) regression problem one can write

$$\min_\beta \sum_{i=0}^{n} \{Y_i - \sum_{j=0}^{\rho} \beta_j (X_i - x_0)^j\} \frac{1}{h} K_h(X_i - x_0) \quad (1.11)$$

where h is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$ with K a kernel function assigning weights to each observation.

[13] The discussion in this section follows closely Fan and Gijbels (1996). The reader is referred to this text for further details.

Let $\hat{\beta}_j$, $j = 1\ldots,\rho$ be the solution to the LS problem (1.11). By comparing the Taylor expansion in (1.10) to the LS problem (1.11) it is clear that $\hat{m}_\nu(x_0) = \nu!\hat{\beta}_\nu$ is an estimator for $m^{(\nu)}(\cdot)$, $\nu = 0, 1, \ldots, \rho$.

To derive an expression for $\hat{\beta}$ it is more convenient to work with matrix notation. Let X denote the matrix of regressors of problem (1.11) and organize the variables Ys and the estimators $\hat{\beta}_j$ in column vectors, i.e.

$$X = \begin{pmatrix} 1 & (X_1 - x_0) & \ldots & (X_1 - x_0)^\rho \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (X_n - x_0) & \ldots & (X_n - x_0)^\rho \end{pmatrix}, \; y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_\rho \end{pmatrix}.$$

Further, let W be the $n \times n$ diagonal matrix of weights: $W = \text{diag}\{K_h(X_i - x_0)\}$. Then the weighted LS problem above can be written as

$$\min_\beta (y - X\beta)^T W (y - X\beta),$$

with $\beta = (\beta_0, \ldots, \beta_\rho)^T$. From LS theory the solution to this problem is given by

$$\hat{\beta} = (X^T W X)^{-1} X^T W y$$

Like in the case of kernel estimators there are expressions for the (conditional) bias and variance of the LPE. The following theorem by Ruppert and Wand (1994) establishes the asymptotic expansions for the bias and variance of the estimator $\hat{m}_\nu(x_0) = \nu!\hat{\beta}_\nu$. First we need to introduce the following notation. Let $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$. Also let $S = (\mu_{j+l})_{0\leq j,l\leq\rho}$, $\tilde{S} = (\mu_{j+l+1})_{0\leq j,l\leq\rho}$, $S^* = (\nu_{j+l})_{0\leq j,l\leq\rho}$, $c_p = (\mu_{\rho+1}, \ldots, \mu_{2\rho+1})^T$, $\tilde{c}_\rho = (\mu_{\rho+2}, \ldots, \mu_{2\rho+2})^T$. Further, we consider the unit vector $e_{\nu+1} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, with 1 on the $(\nu + 1)$st position.

Theorem [Ruppert and Wand (1994)] *Assume that $f(x_0) > 0$ and that $f(\cdot)$, $m^{(\rho+1)}(\cdot)$ and $\sigma(\cdot)$ are continuous in a neighborhood of x_0. Further, assume that $h \to 0$ and $nh \to \infty$. Then the asymptotic conditional variance of $\hat{m}_\nu(x_0)$ is given by*

$$\text{Var}(\hat{m}_\nu(x_0)|X) \quad = \quad e_{\nu+1}^T S^{-1} S^* S^{-1} e_{\nu+1} \frac{\nu!^2 \sigma^2(x_0)}{f(x_0) n h^{1+2\nu}} + op\left(\frac{1}{nh^{1+2\nu}}\right)$$

The asymptotic conditional bias for $\rho - \nu$ odd is given by

$$\text{Bias}(\hat{m}_\nu(x_0)|X) = e_{\nu+1}^T S^{-1} c_\rho \frac{\nu!}{(\rho+1)!} m^{(\rho+1)}(x_0) h^{\rho+1-\nu} + op(h^{\rho+1-\nu})$$

Further, for $\rho - \nu$ even the asymptotic conditional bias is

$$\begin{aligned}\text{Bias}(\hat{m}_\nu(x_0)|X) &= e_{\nu+1}^T S^{-1} \tilde{c}_\rho \frac{\nu!}{(\rho+2)!} \{m^{(\rho+2)}(x_0) + (\rho+2) m^{(\rho+1)}(x_0) \frac{f'(x_0)}{f(x_0)}\} h^{\rho+2-\nu} \\ &\quad + op(h^{\rho+2-\nu})\end{aligned}$$

provided that $f'(\cdot)$ and $m^{(p+2)}(\cdot)$ are continuous in a neighborhood of x_0 and $nh^3 \to \infty$.

From the above theorem it is clear that there is a theoretical distinction between the cases $p - \nu$ odd and $p - \nu$ even. Moreover, polynomial fits with $p - \nu$ odd outperform those with $p - \nu$ even. For an exhaustive and detailed discussion see Fan and Gijbels (1996).

As with kernel estimators the crucial choice for LPEs is the bandwidth parameter. Of less importance (in practice) is the choice of the order of the polynomial, ρ and the kernel used for weighting. Optimal choices for bandwidths and kernels involved in LP estimation have been studied in the statistical literature. With respect to the choice of ρ, Fan and Gijbels (1996) emphasize that for many applications the "rule" $\rho = \nu + 1$ suffices.

LP fitting has a number of attractive features both from theoretical and practical point of view. The most relevant for this thesis is the behavior of these estimators at the boundaries. Put in other words, the bias at the boundary stays automatically of the same order as in the interior, without the use of specific boundary kernels. This is remarkably different from kernel estimation (and other nonparametric methods as well).[14] On top of the advantages at the boundaries, LPEs have nice minimax efficiency properties; the asymptotic minimax efficiency for commonly used orders is 100% among all linear estimators and only a small loss has to be tolerated beyond this class.[15]

[14] See Silverman (1986), Fan and Marron (1993).

[15] LPEs belong to the class of linear smoothers. This class includes the Nardaraya–Watson estimator, the Gasser–Müller estimator, orthogonal series estimators and spline smoothers.

Chapter 2

Semiparametric Estimation of First–Price Auction Models

2.1 Introduction

From a theoretical point of view, auctions are modeled as games of incomplete information in which asymmetric information among players (seller/buyer and bidders) is one of the key features. See for example Krishna (2002), McAfee and McMillan (1987) and Wilson (1992). From an applied perspective, since auctions are widely used mechanisms to allocate goods and services which are often public, many data sets are available for empirical research. By assuming that observed bids are the equilibrium outcomes of the underlying auction model under consideration, the structural approach to analyze auction data provides a framework in which the theoretical model and its empirical counterpart are closely related. The main objective of this approach is then to recover the structural elements of the auction model. This line of research has developed considerably over the last fifteen years. The difficulties in estimating auction models are many. First, auction models lead to highly nonlinear econometric models through the equilibrium strategies. Second, auction models may not lead to tractable solutions rendering even more difficult the derivation of an econometric model. Third, the estimation of auction models often requires the numerical computation of the equilibrium strategy and its inverse.

As documented by Perrigne and Vuong (1999, 2008) several contributions to the structural estimation of first–price auction models are available in the literature.[1] We

[1]See also Paarsch and Hong (2006) for an extensive survey on structural estimation of auction models

distinguish two kinds of methods for estimating structural auction models: direct methods and indirect methods. Direct methods were first developed in the literature relying on parametric econometric models. Starting from a specification of the underlying distribution of private values, the objective of direct methods is to estimate the parameter vector characterizing such a distribution. Within this class of methods, there are two major estimation procedures. The first methodology introduced by Paarsch (1992) and Donald and Paarsch (1993) is a fully parametric setup that uses Maximum Likelihood (ML)–based estimation procedures requiring the computation of the equilibrium strategy. This in turn could be highly computationally demanding, as recognized by Donald and Paarsch (1993), and thus only very simple distributions are considered in practice. In particular, because the upper bound of the bid distribution depends on the parameter(s) of the underlying distribution, the ML estimator has a nonstandard limiting distribution. In view of this, Donald and Paarsch (1993) develop a so–called piecewise pseudo ML estimator requiring the computation of the equilibrium strategy that can be obtained using specific parametric distribution(s). Laffont, Ossard, and Vuong (1995) introduced a second methodology, which is computationally more convenient. Relying on the revenue equivalence theorem, the authors propose a simulation–based method that avoids computation of the equilibrium strategy and therefore allows for more general parametric specifications for the private value distribution.

More recently Guerre, Perrigne, and Vuong (2000) (GPV (2000) hereafter) have developed a fully nonparametric indirect procedure introducing the use of indirect methods for the structural estimation of auction models. This alternative methodology relies on a simple but crucial observation, namely each private value can be expressed as a function of the corresponding bid, the distribution of observed bids and its density using the first–order condition of the bidder's optimization problem. Based on this equation, which defines the inverse of the equilibrium strategy, the authors show that the model is nonparametrically identified. Other papers by the same authors and others study other auction models in a similar fashion. Including the affiliated private value model, models with asymmetric bidders, dynamic auction models and models with risk averse bidders, etcetera. Therefore, in contrast to direct methods, indirect methods start from the distribution of observed bids in order to estimate the distribution of unobserved private values without computing the Bayesian–Nash equilibrium strategy or its inverse explicitly. This calls for a two step procedure. In the first step, a sample of pseudo private valuations is obtained while using (say) kernel estimators for the distribution

within the Independent Private Value (IPV) paradigm.

and density of observed bids. With this pseudo sample, the second step consists of nonparametrically estimating the density of bidders' private values. GPV (2000) also establish some asymptotic properties of their estimator, namely its uniform consistency and the achievement of the optimal convergence rate by appropriate vanishing rates for bandwidths, while the optimal rate is derived using the minimax theory as developed by Ibragimov and Has'minskii (1981).

Though a fully nonparametric estimator has some desirable properties such as flexibility and robustness to misspecification, it has a number of drawbacks such as a slow convergence rate and the problems associated with using a large number of covariates because of the so–called curse of dimensionality. In this chapter, we propose a semiparametric estimator within the class of indirect methods. In the same spirit as GPV (2000) our procedure consists of two steps. Unlike GPV (2000), the second step is fully parametric and therefore our resulting model falls within the category of semiparametric models. Specifically, we propose to model private valuations through a set of conditional moment restrictions.[2]

For sake of simplicity, we consider a symmetric first–price sealed–bid auction model within the independent private value (IPV) paradigm with a nonbinding reserve price. As shown below, our estimation procedure applies to a more general class of auction models, namely symmetric and asymmetric affiliated private value models. More generally, our method extends to models which have been estimated using a nonparametric indirect procedure. Let $V_{p\ell}$, $p = 1, \ldots, I_\ell$, $\ell = 1, \ldots, L$ denote the private value of the pth bidder for the ℓth auctioned object. Let $Z_\ell \equiv (X_\ell, I_\ell) \in I\!R^{d+1}$ denote the vector of exogenous variables, which includes the number of bidders I_ℓ and variables X_ℓ characterizing object heterogeneity across auctions.[3] We model private values by the following set of conditional moment restrictions

$$\mathrm{E}[M(V, Z; \theta_0)|Z] = 0, \tag{2.1}$$

for some known function $M(\cdot, \cdot; \theta) : I\!R^{d+2} \to I\!R^q$ and $\theta \in I\!R^p$. However, such moment

[2] A noticeable exception is found in Jofre-Bonet and Pesendorfer (2003) which use a fully parametric indirect procedure to entertain a large number of covariates. The authors do not provide, however, any asymptotic properties for their estimator.

[3] The dependence of the private value distribution on I_ℓ captures the idea that private values and the number of bidders can be dependent in general. For instance, objects of higher value may attract more bidders. The number of bidders may capture some unobserved heterogeneity. It may also result from endogenous participation. Note also that the reserve price is nonbinding so that the number of potential bidders I_ℓ in the ℓth auction is known. This assumption can be relaxed and our results can be straightforwardly extended in a similar way as in GPV (2000, section 4).

restrictions are infeasible in practice since private values are unobserved.

From the theoretical model we know that the equilibrium bid, B, can be expressed as $B = s(V, Z; \theta_0)$. Therefore the equilibrium strategy depends on the parameter vector θ_0 both directly since $B \sim G(\cdot|Z; \theta_0)$ (say) and indirectly through V since $V \sim F(\cdot|Z; \theta_0)$. We note that even in the simplest auction model the strategy $s(\cdot)$ depends on $F(\cdot)$ which is unknown. A natural way of expressing the above conditions would be

$$\mathrm{E}\big\{M[s^{-1}(B, Z; \theta_0), Z; \theta_0]\big|Z\big\} = 0.$$

The above condition requires the computation of the equilibrium strategy as well as its inverse. This could be highly computationally demanding for two different reasons. First, such computation has to be carried out for any trial value of the parameters. Second, in a more general class of models, the computation of the equilibrium strategy $s(\cdot, \cdot; \theta_0)$ and of its inverse is much more involved such as for affiliated private value models. Furthermore, in asymmetric models this computation becomes intractable.

The set of conditional moment restrictions given by (2.1) lead to the following (infeasible) unconditional moment restrictions

$$\mathrm{E}[m(V, Z; \theta_0)] = 0.$$

where $m(\cdot, \cdot; \theta) : I\!R^{d+2} \to I\!R^q$, $\theta \in I\!R^p$ and $q \geq p$.

We therefore propose to replace V in (2.1) by its nonparametric estimator $\hat{V} = \hat{\xi}(B, Z)$ to make the moment condition operational. Thus, as explained below, we minimize a quadratic form based on the following sample moment conditions

$$\hat{S}_L(\theta) = \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} m(\hat{V}_{p\ell}, Z_\ell; \theta).$$

It is worth noting that $\hat{V} = \hat{\xi}(B, Z)$ is obtained nonparametrically by means of local polynomial fitting. Unlike other nonparametric estimators, such us kernels, local polynomial fitting is an attractive method from a theoretical and practical point of view. As pointed out by Fan and Gijbels (1996), local polynomial estimators (LPE) have some advantages over other commonly used nonparametric estimators including the absence of boundary effects insofar as they are related to the convergence rate. As is well–known, many nonparametric estimators are ill–behaved close to the boundaries of the support. A common way to deal with this problem is to trim out observations in these problematic regions. In this chapter, we use LPEs and thus boundary effects do not affect the second step. Therefore we do not have to trim out observations. This is a remarkable advantage of our procedure. We notice that in this context the trimming would have to be performed on the endogenous variable of the model. This introduces an additional technical difficulty because this trimming on the bids, which implies an automatic trimming on private values, will affect the moments of the latter. In a standard econometric framework the trimming is usually applied on the exogenous variables. See for instance, Lavergne and Vuong (1996) and Robinson (1988).

In line with GPV (2000) our econometric model calls for a two step procedure. The first step is similar to the first step in GPV (2000) in which we recover a sample of pseudo private values while using an LPE. The second step departs from the fully nonparametric second step in GPV (2000) since we instead use a GMM procedure to obtain an estimate for θ_0. Thus our procedure is semiparametric.

We establish the consistency and asymptotic normality of our semiparametric estimator. Specifically, we show that our estimator converges at the parametric $\sqrt{L}$ rate. As is well–known, nonparametric estimators converge at a slower rate than $\sqrt{L}$ and their rates are usually negatively related to the dimension of the vector of exogenous variables. This makes these estimators less desirable in applications, especially when a limited number of observations is available and/or when the number of exogenous variables is relatively large.[4] The estimator we propose in this chapter does not share this drawback since it is not subject to the curse of dimensionality. In other words, the convergence rate of our estimator is independent of the dimension of the exogenous variables. A second major advantage of our estimation procedure is that it can be used to estimate more general auction models. Indirect methods in general do not require neither the computation of the equilibrium strategy nor of its inverse. Therefore these methods are especially convenient when there is no closed form solution to the differential equation(s) characterizing the equilibrium strategy as is the case in asymmetric auction models which lead to intractable expressions for the first–order conditions.

The rest of the chapter is organized as follows. In section 2.2, we introduce the theoretical model from which the structural econometric model is derived as well as our semiparametric two–step estimator. Section 2.3 establishes the asymptotic properties

[4]Examples of semiparametric estimators attaining $\sqrt{L}$ rate can be found in Newey and McFadden (1994) and Powell (1994). Some notable exceptions are the estimators proposed by Manski (1985), Horowitz (2000), Kryriazidou (1997) and Honoré and Kryriazidou (2000). An example of a semiparametric estimator converging at a slower than the parametric rate but not subject to the curse of dimensionality, i.e. its rate is independent of d, is given by Campo, Guerre, Perrigne, and Vuong (2006).

of our estimator. In section 2.4 we present some Monte Carlo experiments to illustrate the properties of our procedure in small samples and to assess its advantages relative to a nonparametric procedure. Section 2.5 discusses how to extend our procedure in a more general class of auction models. Section 2.6 contains the conclusions. Appendix A collects the proofs of our results.

2.2 The Model

2.2.1 The Symmetric IPV Model

We present the benchmark theoretical model underlying our structural econometric model, namely the symmetric IPV model with a nonbinding reserve price. Although this model is restrictive for applications, it allows us to develop our econometric procedure in a more transparent way. We postpone possible extensions until section 2.5. A single and indivisible object is auctioned to I_ℓ bidders who are assumed to be ex ante identical (i.e. the game is symmetric) and risk neutral. We allow for the possibility that the total number of bidders varies across auctions as well. Unobserved private values are denoted by V. More precisely, we assume that each valuation $V_{p\ell}$, $\ell = 1, \ldots, L$, $p = 1, \ldots, I_\ell$, is distributed according to $F(\cdot|Z_\ell;\theta_0)$, where $\theta_0 \in I\!R^p$ is the parameter of interest. The distribution function $F(\cdot)$ is "almost" arbitrary. That is, we do not assume a specific functional form for this function but $F(\cdot)$ is an arbitrary function of its arguments but varies as a parametric function of its conditional variables. The support of $F(\cdot|\cdot)$ is $[\underline{V}_\ell, \overline{V}_\ell]$, with $0 \leq \underline{V}_\ell = \underline{V}(Z_\ell) < \overline{V}_\ell = \overline{V}(Z_\ell) < \infty$. Among others, Riley and Samuelson (1981) have characterized the unique symmetric differentiable Bayesian Nash equilibrium. In particular, for every ℓ and $I_\ell \geq 2$ the equilibrium bid $B_{p\ell}$ in the ℓth auction is

$$B_{p\ell} = s_0(V_{p\ell}, Z_\ell) = V_{p\ell} - \frac{1}{F(V_{p\ell}|Z_\ell;\theta_0)^{I_\ell - 1}} \int_{\underline{v}_\ell}^{V_{p\ell}} F(v|Z_\ell;\theta_0)^{I_\ell - 1} dv, \tag{2.2}$$

for any $V_{p\ell}$ since $p_0 < \underline{v}(Z_\ell)$ subject to the boundary condition $s(\underline{V}_\ell) = \underline{V}_\ell$.

The distribution and density of observed bids in the ℓth auction are given by $G(\cdot|Z_\ell;\theta_0)$ $= G_0(\cdot|Z_\ell)$ and $g(\cdot|Z_\ell;\theta_0) = g_0(\cdot|Z_\ell)$, respectively. Observed bids are assumed to be the equilibrium outcome of the game. From GPV (2000), the observed bids and private values are related by the following equilibrium expression

$$V_{p\ell} = \xi_0(B_{p\ell}, Z_\ell) = B_{p\ell} + \frac{1}{I_\ell - 1} \frac{G_0(B_{p\ell}|Z_\ell)}{g_0(B_{p\ell}|Z_\ell)}, \tag{2.3}$$

for $\ell = 1, \ldots, L,\ p = 1, \ldots, I_\ell$. Equation (2.3) constitutes the basis for the identification result in GPV (2000), i.e. the authors show that the model is nonparametrically identified.

2.2.2 The Two Step Estimator

In line with GPV (2000), equation (2.3) is the basis for our econometric model. The difference with GPV (2000) is that we model private values as a set of moment conditions. Therefore knowledge of $G_0(\cdot|\cdot)$ and $g_0(\cdot|\cdot)$ would lead us to a GMM framework. However, these functions are unknown in practice but can be estimated from observed bids.

Ideally, one would like to specify the following set of conditional moment restrictions

$$\mathrm{E}[M(V, Z; \theta_0)|Z] = 0,$$

which would lead to the following unconditional moment restrictions

$$\mathrm{E}[m(V, Z; \theta_0)] = 0.$$

Therefore, the infeasible estimator $\tilde{\theta}$, (say), is the solution to

$$\tilde{\theta} = \arg\min_{\theta \in \Theta} S_L^T(\theta) \Omega S_L(\theta),$$

where $S_L(\theta) = 1/L \sum_{\ell=1}^{L} 1/I_\ell \sum_{p=1}^{I_\ell} m(V_{p\ell}, Z_\ell; \theta)$ and Ω is a positive definite matrix of order q.

As mentioned before we propose to estimate θ_0 by $\hat{\theta}$ as follows

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{S}_L^T(\theta) \Omega \hat{S}_L(\theta), \qquad (2.4)$$

where $\hat{S}_L(\theta) = 1/L \sum_{\ell=1}^{L} 1/I_\ell \sum_{p=1}^{I_\ell} m(\hat{V}_{p\ell}, Z_\ell; \theta)$.

As we show in Appendix A, the asymptotic distributions of $\hat{\theta}$ and $\tilde{\theta}$ are closely related though not the same.

Before presenting our two–step estimator, it is worth mentioning that some of our assumptions are similar or even identical to those in GPV (2000). This is not surprising since our methodology follows closely their methodology. In particular we follow GPV (2000) and indicate when some modifications are necessary. Our first two assumptions deal with the underlying data generating process and the smoothness of the latent joint distribution of $(V_{p\ell}, Z_\ell)$ for any $p = 1, \ldots, I_\ell$.

Assumption A1:

(i) $Z_\ell = (X_\ell, I_\ell) \in I\!R^{d+1}$, $\ell = 1, 2, \ldots$ *are independently and identically distributed with distribution function* $F_m(\cdot, \cdot)$ *and density* $f_m(\cdot, \cdot)$.[5]

(ii) For each ℓ, $V_{p\ell}$, $p = 1, \ldots, I_\ell$ *are independently and identically distributed conditionally upon* Z_ℓ *with distribution function* $F(\cdot|\cdot;\theta_0)$ *and density* $f(\cdot|\cdot;\theta_0)$

Let $\mathcal{I}$ be the set of possible values for I_ℓ. We denote by $S(*)$ the support of $*$ and by $S_i(*)$ the support when the number of bidders is equal to i.

Assumption A2: $\mathcal{I}$ *is a bounded subset of* $\{2, 3, \ldots\}$, *and:*

(i) for each $i \in \mathcal{I}$, $S_i(F) = \{(v, x) : x \in [\underline{x}, \overline{x}], v \in [\underline{v}(x), \overline{v}(x)]\}$, *with* $\underline{x} < \overline{x}$,

(ii) for $(v, x, i) \in S(F)$, $f(v|x, i; \theta_0) \geq c_f > 0$, *and for* $(x, i) \in S(F_m)$, $f_m(x, i) \geq c_f > 0$,

(iii) for each $i \in \mathcal{I}$, $F(\cdot|\cdot, i; \theta_0)$ *and* $f_m(\cdot, i)$ *admit up to* $R+1$ *continuous bounded partial derivatives on* $S_i(F)$ *and* $S_i(F_m)$, *with* $R > d + 1$.

These assumptions can be found in GPV (2000) as well, though A2-(iii) is stronger in our case. That is, we need to require R to be sufficiently large with respect to the dimension of X, i.e. $R > d + 1$. This kind of condition has been used elsewhere in the semiparametric literature, e.g. Powell, Stock, and Stoker (1989). Note that proposition 1 in GPV (2000) still holds under our assumption A2.

The next two assumptions are on kernels and bandwidths that we use in the first stage.

Assumption A3:

(i) The kernels $K_G(\cdot)$, $K_{1g}(\cdot)$ *and* $K_{2g}(\cdot)$ *are symmetric with bounded hypercube supports and twice continuous bounded derivatives with respect to their arguments,*

(ii) $\int K_G(x)dx = 1$, $\int K_{1g}(x)dx = 1$, $\int K_{2g}(b)db = 1$

(iii) $K_G(\cdot)$, $K_{1g}(\cdot)$ *and* $K_{2g}(\cdot)$ *are of order* $R - 1$. *Thus moments of order strictly smaller than* $R - 1$ *vanish.*

This is similar to assumption A3 in GPV (2000). It is a standard assumption in the nonparametric literature.

[5] In a slight abuse of terminology we use $f_m(\cdot)$ to denote the following: $\frac{\partial}{\partial x}\Pr(X \leq x, I = i)$ assuming that the variables in X are continuous. We also note that the vector X could contain discrete elements and our results still go through.

Assumption A4: *The bandwidths h_G, h_{1g} and h_{2g} satisfy*

(i) $h_G \to 0$ *and* $\dfrac{Lh_G^d}{\log L} \to \infty$, *as* $L \to \infty$,

(ii) $h_{1g} \to 0$, $h_{2g} \to 0$ *and* $\dfrac{Lh_{1g}^d h_{2g}}{\log L} \to \infty$, *as* $L \to \infty$.

As shown below, for consistency of our estimator it is possible to choose the optimal bandwidths in the first step, i.e. the bandwidths proposed in Stone (1982). Unlike GPV (2000) we do not need to specify a "boundary bandwidth" since the local polynomial method does not require knowledge of the location of the endpoints of the support. Therefore, it is not necessary to estimate the boundary of the support of the bid distribution.

For simplicity of presentation, in the rest of the chapter we only deal with the univariate X case, that is $d = 1$.

In order to describe our two–step estimator, we observe first that, analogously to the first step in GPV (2000), our objective is to estimate the ratio $\psi(\cdot|\cdot) = G_0(\cdot|\cdot)/g_0(\cdot|\cdot)$ by $\hat{\psi} = \hat{G}(\cdot|\cdot)/\hat{g}(\cdot|\cdot)$ (see (2.3) above). From proposition 1 in GPV (2000) we know that $G_0(\cdot|\cdot)$ is $R+1$ times continuously differentiable on its entire support and therefore $g_0(\cdot|\cdot)$ is R times continuously differentiable on its entire support as well.[6] Given the smoothness of each function we propose to use a LPE(R), i.e. a LPE of degree R, for $G_0(\cdot|\cdot)$ and a LPE($R-1$) for $g_0(\cdot|\cdot)$. We first introduce some notation.

Let $P_\rho(X;\beta)$ denote a polynomial of degree ρ in X with parameter β. Then for each each i we have

$$\hat{G}(b|x) = \arg\max_{\beta_G} \sum_{\{\ell:I_\ell=i\}}^{L} \sum_{p=1}^{i} \left\{ Y_{p\ell}^G - P_R(X_\ell - x;\beta_G) \right\}^2 \frac{1}{h_G} K_G\left(\frac{X_\ell - x}{h_G}\right)$$

where $Y_{p\ell}^G = \mathbb{1}(B_{pl} \leq b)$, and

$$\hat{g}(b|x) = \arg\max_{\beta_g} \sum_{\{\ell:I_\ell=i\}}^{L} \sum_{p=1}^{i} \left\{ Y_{p\ell}^g - P_{R-1}(X_\ell - x;\beta_g) \right\}^2 \frac{1}{h_{1g}} K_{1g}\left(\frac{X_\ell - x}{h_{1g}}\right)$$

where $Y_{p\ell}^g = \frac{1}{h_{2g}} K_{2g}\left(\frac{B_{p\ell}-b}{h_{2g}}\right)$.

[6] Observe that by proposition 1 in GPV (2000) we also know that the conditional density $g_0(\cdot|\cdot)$ is $R+1$ times continuously differentiable on a closed subset of the interior of the support and thus the degree of smoothness closed to the boundaries and at the boundaries of the support is not necessarily $R+1$.

More precisely we have,

$$\begin{aligned}\hat{G}(b|x,i) &= \frac{1}{h_G}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{p=1}^{i} e_1^T (X_{i,R+1}^T W_x^G X_{i,R+1})^{-1} \\ &\quad \times X_{R+1,\ell} K_G\left(\frac{X_\ell - x}{h_G}\right) 1\!\!I(B_{p\ell} \le b) \\ &= \frac{1}{Lh_G}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{p=1}^{i} e_1^T \left(\frac{X_{i,R+1}^T W_x^G X_{i,R+1}}{n_i}\right)^{-1} \\ &\quad \times X_{R+1,\ell} K_G\left(\frac{X_\ell - x}{h_G}\right) 1\!\!I(B_{p\ell} \le b) \qquad (2.5)\end{aligned}$$

and

$$\begin{aligned}\hat{g}(b|x,i) &= \frac{1}{h_{1g}h_{2g}}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{p=1}^{i} e_1^T (X_{i,R}^T W_x^g X_{i,R})^{-1} X_{R,\ell} K_{1g}\left(\frac{X_\ell - x}{h_{1g}}\right) \\ &\quad \times K_{2g}\left(\frac{B_{p\ell} - b}{h_{2g}}\right) \\ &= \frac{1}{Lh_{1g}h_{2g}}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{p=1}^{i} e_1^T \left(\frac{X_{i,R}^T W_x^g X_{i,R}}{n_i}\right)^{-1} X_{R,\ell} K_{1g}\left(\frac{X_\ell - x}{h_{1g}}\right) \\ &\quad \times K_{2g}\left(\frac{B_{p\ell} - b}{h_{2g}}\right) \qquad (2.6)\end{aligned}$$

where for $s = \{R, R+1\}$,

e_1 is the unit vector in $I\!R^s$ containing a 1 in its first entry,

$n_i = iL_i$,

$L_i = \#\{\ell : I_\ell = i\}$,

$X_{s,\ell} = [1 \quad (X_\ell - x)\ldots(X_\ell - x)^{s-1}]^T$, is a $s \times 1$ vector and,

$$X_{i,s} = \begin{pmatrix} 1 & (X_1 - x) & \ldots & (X_1 - x)^{s-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (X_{n_i} - x) & \ldots & (X_{n_i} - x)^{s-1} \end{pmatrix}$$

is the matrix of regressors of dimension $n_i \times s$ with the first i rows identical and similarly for the other rows,

$W_x^G = \text{diag}\left\{\frac{1}{h_G} K_G\left(\frac{X_\ell - x}{h_G}\right)\right\}$,

$W^g_x = \text{diag}\left\{\frac{1}{h_{1g}} K_{1g}\left(\frac{X_\ell - x}{h_{1g}}\right)\right\}$,

where $K_G(\cdot)$, $K_{1g}(\cdot)$ and $K_{2g}(\cdot)$ are kernels with bounded support and h_G, h_{1g},h_{2g} are bandwidths (see A3 and A4). Given (2.5) and (2.6), a natural way of recovering pseudo private values is

$$\hat{V}_{p\ell} = B_{p\ell} + \frac{1}{I_\ell - 1}\hat{\psi}(B_{p\ell}|Z_\ell).$$

Then, as outlined above, in a second step we implement a GMM procedure to obtain the estimator $\hat{\theta}$ of θ_0.

2.3 Asymptotic Properties

In this section we show that our two-step semiparametric estimator $\hat{\theta}$ of θ_0 is consistent and asymptotically normal distributed. Moreover, we establish that our estimator attains the parametric rate of convergence given an appropriate choice of the bandwidths used in the first step to estimate $G_0(\cdot|\cdot)$ and $g_0(\cdot|\cdot)$. As we will discuss below the optimal bandwidths given by Stone (1982), i.e. the one–step bandwidths, cannot be chosen instead our choice implies that in practice one needs to undersmooth.

We also discuss the assumptions under which our results hold.

2.3.1 Consistency

Our first result establishes that $\hat{\theta}$ is a (strongly) consistent estimator of θ_0. Moreover this is the case even if one uses the optimal bandwidths for estimating $G_0(\cdot|\cdot)$ and $g_0(\cdot|\cdot)$ in the first step, i.e. the bandwidths proposed by Stone (1982). To see this, we notice that the "optimal one–step" bandwidths satisfy our assumption A4 above (with $d = 1$) since they are of the form,

(i) $h_G = \lambda_G \left(\frac{\log L}{L}\right)^{1/(2R+3)}$

(ii) $h_{1g} = \lambda_{1g} \left(\frac{\log L}{L}\right)^{1/(2R+1)}$ *and* $h_{2g} = \lambda_{2g} \left(\frac{\log L}{L}\right)^{1/(2R+1)}$,

where λ_G, λ_{1g} *and* λ_{2g} *are strictly positive constants.*

As observed by GPV (2000), h_G, h_{1g} and h_{2g}, as given above are optimal bandwidth choices to estimate $G_0(\cdot|\cdot)$ and $g_0(\cdot|\cdot)$ given proposition 1 and A2-(iii) in that paper.[7]

[7] As pointed out before, A2-(iii) in our case is stronger than A2-(iii) in GPV (2000). Thus their proposition 1 also holds in our framework.

Thus, A4 implies that our consistency result can be established when using LPEs in the first stage that converge at the best possible rate. The remaining assumptions are standard in the literature.

Assumption A5:

(i) *The parameter space* $\Theta \subset I\!R^p$ *is compact and* θ_0 *is in the interior of* Θ*,*

(ii) *Identifying assumption:* $\mathrm{E}[m(V, Z; \theta)] = 0$ *if and only if* $\theta = \theta_0$*,*

(iii) $\sup_{\theta \in \Theta} \left| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \|m(V_{p\ell}, Z_\ell; \theta)\| - \mathrm{E}\,\|m(V, Z; \theta)\| \right| = o_{as}(1),$

(iv) $m(v, z; \theta)$ *is Lipschitz in* v*: there exists a measurable function* $K_1(\cdot)$ *such that*

$$\left\|m(v, z; \theta) - m(v', z; \theta)\right\| \leq K_1(z) \left|v - v'\right|,$$

for every v, $v' \in [\underline{V}, \overline{V}]$*, for every* z*, and every* $\theta \in \Theta$*. Moreover* $\mathrm{E}[K_1(\cdot)] < \infty$*.*

We now state our consistency result.

Proposition 1: *Let* $\hat{\theta}$ *be defined as in (2.4). Then, under A1-A5, we have*

$$\hat{\theta} \xrightarrow{a.s} \theta_0.$$

Proposition 1 is important since it establishes that our estimator possess one of the desirable asymptotic properties. This is the first step in order to be able to establish the asymptotic distribution of the estimator. Moreover, there is no need to undersmooth the distribution and density functions in the first step in order for $\hat{\theta}$ to be consistent.

2.3.2 Asymptotic Normality

Given that $\hat{\theta}$ is a (strongly) consistent estimator of θ_0, we can now establish its asymptotic distribution along with its convergence rate. This is the purpose of proposition 2 below. Before, we introduce some additional assumptions.

We need to modify our choice of bandwidths as mentioned earlier. Thus, as is made clear by A4.AN below, in order for $\hat{\theta}$ to achieve the parametric rate of convergence we need to specify bandwidths for our first step that rule out the optimal choice and moreover that imply undersmoothed estimates for $\hat{G}(\cdot|\cdot)$ and $\hat{g}(\cdot|\cdot)$.

Assumption A4.AN: *The bandwidths h_G, h_{1g} and h_{2g} satisfy*

(i) $\sqrt{L}h_G^{R+1} \to 0$ *and* $\dfrac{\log L}{\sqrt{L}h_G} \to 0$, *as* $L \to \infty$

(ii) $\sqrt{L}h_{1g}^R \to 0$, $\sqrt{L}h_{2g}^R \to 0$ *and* $\dfrac{\log L}{\sqrt{L}h_{1g}h_{2g}} \to 0$, *as* $L \to \infty$.

(iii) $h_{1g} = h_{2g}$

The assumption that h_{1g} and h_{2g} are the same, is to simplify the notation in the proof. In fact it is enough to choose any pair of bandwidths satisfying A4.AN-(ii).

We denote by $m_j(\cdot)$ the derivative of $m(\cdot)$ with respect to its jth argument. The next set of assumptions is standard.

Assumption A6:

(i) $m_3(v, z; \theta)$ *is Lipschitz in* v*: there exists a measurable function* $K_3(\cdot)$ *such that*

$$\left\|m_3(v, z; \theta) - m_3(v', z; \theta)\right\| \leq K_3(z)|v - v'|,$$

for every v, $v' \in [\underline{V}, \overline{V}]$, *for every* z *and* $\theta \in \Theta$. *Moreover* $\mathrm{E}[K_3(\cdot)] < \infty$,

(ii) $m_3(v, z; \theta)$ *is Lipschitz in* θ*: there exists a measurable function* $K_4(\cdot)$ *such that*

$$\left\|m_3(v, z; \theta) - m_3(v, z; \theta')\right\| \leq K_4(z) \left\|\theta - \theta'\right\|,$$

for every θ, $\theta' \in \Theta$, *for every* z, *and* $v \in [\underline{V}, \overline{V}]$. *Moreover* $\mathrm{E}[K_4(\cdot)] < \infty$,

(iii) $\sup_{\theta\in\Theta} \left\| \dfrac{1}{L}\sum_{\ell=1}^{L} \dfrac{1}{I_\ell}\sum_{p=1}^{I_\ell} m_3(V_{p\ell}, Z_\ell; \theta) - \mathrm{E}[m_3(V, Z; \theta)] \right\| = o_{as}(1)$ *and*

$\mathrm{E}[m_3^T(V, Z; \theta)]\Omega\mathrm{E}[m_3(V, Z; \theta)]$ *is non singular,*

(iv) $\sup_{\theta\in\Theta} \|m_3(V, Z; \theta)\| \leq K_5(V, Z)$ *with* $\mathrm{E}[K_5(\cdot, \cdot)] < \infty$,

(v) $m_1(v, z; \theta)$ *is Lipschitz in* v*: there exists a measurable function* $K_6(\cdot)$ *such that*

$$\left\|m_1(v, z; \theta) - m_1(v', z; \theta)\right\| \leq K_6(z)|v - v'|,$$

for every v, $v' \in [\underline{V}, \overline{V}]$, *for every* z, *and* $\theta \in \Theta$. *Moreover* $\mathrm{E}[K_6(\cdot)] < \infty$, *(vi)* $\sup_{\theta \in \Theta} \|m_1(V, Z; \theta)\| \leq K_7(V, Z)$ *with* $\mathrm{E}[K_7(\cdot, \cdot)^2] < \infty$.

Our last assumption concerns a technical condition we need for the proof of proposition 2. Our problem brings some technical difficulties that are not usually found in the semiparametric literature. As a consequence, we need technical devices to get around some of these problems. This is the purpose of Assumption A7.[8]

Assumption A7:

(i) Technical condition: $\mathrm{E}[m_1(V, Z; \theta_0)] = 0$.

We acknowledge that this assumption is restrictive. Nevertheless, we believe that Assumption A7 can be removed if one is willing to use Jackknife methods to deal with the bias term. We leave this for future work. An example in which this assumption is satisfied is the following. Suppose for simplicity that $Z = X \in I\!\!R$ and let $p = 1$. Let $\mathrm{E}(V|Z) = Z\theta_0$. Then, $\mathrm{E}(V - Z\theta_0|Z) = 0$. Thus, we can write the following unconditional moment restriction $\mathrm{E}(Z(V - Z\theta_0)) = 0$. Therefore if $\mathrm{E}(Z) = 0$, then Assumption A7 is satisfied.

Next, we establish our main result.

PROPOSITION 2: *Let $\hat{\theta}$ be defined as in (2.4). Then, under A1-A3, A4.AN and A5-A7, we have*

$$\sqrt{L}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma).$$

where, for each $i \in \mathcal{I}$

$$\Sigma = \mathrm{Var}(\psi_1),$$

$$\begin{aligned} \psi_1 \quad = \quad & -1/i \sum_{p=1}^{i} \Bigg\{ (C^T \Omega C)^{-1} C \Omega m(V_{p1}, X_1; \theta_0, i) + 2 \Bigg[\sum_i \frac{1}{i(i-1)} N(Y_{p1}, i) \\ & \times f_m^{-1}(X_1, i) g_0(Y_{p1}, i) - \mathrm{E}\Big[\sum_i \frac{1}{i(i-1)} N(Y_{p1}, i) f_m^{-1}(X_1, i) g_0(Y_{p1}, i) \Big] \Bigg] \Bigg\}, \end{aligned}$$

$$C \quad = \quad E[m_3(V, X; \theta_0, i)],$$

[8] Assumptions A7 basically implies that we can apply a Taylor expansion around h at $h = 0$ to a bias term we need to deal with.

and

$$N(Y_{p1}, i) = [m_1(V_{p1}, X_1; \theta_0, i)/g_0(B_{p1}|X_1, i)^2]G_0(B_{p1}|X_1, i).$$

Proposition 2 is important for several reasons. First it establishes that our semiparametric estimator has a standard limiting distribution. Asymptotic normality is fundamental since most econometric tests rely on it. Second, although slow estimators are used in the first step of our estimation procedure to recover pseudo private values, the estimator of the parameter of interest converges at the best possible rate, thereby avoiding the curse of dimensionality. Finally, proposition 2 can be used to conduct inference on θ_0.

2.4 Monte Carlo Experiments

In this section we present the results of a set of Monte Carlo simulations in order to assess the performance of our semiparametric estimator relative to the nonparametric estimator proposed in GPV (2000). In line with the theoretical framework analyzed in the previous section we consider a setup with observed object heterogeneity ($d = 1$). The setup here is fully parametric. We use $L = 200$ with $I = 5$ bidders, which gives a total of 1,000 observed bids. The choice $L = 200$ corresponds to a realistic size of an auction data set and makes the comparison with GPV (2000) easier since these authors also use this sample size although they consider homogeneous auctions. In order to account for object heterogeneity, we generate X from a log–normal distribution with mean 0 and variance 1 truncated at 0.055 and 30 to satisfy A2-(i). The true distribution $F(\cdot|Z; \theta_0)$ of private values is also log–normal. Moreover we have that conditional on X, private values are log–normally distributed with mean $1+X$ and variance 1. We use $\theta_0 = (1, 1)^T$. To satisfy assumption A2-(ii) we truncate this distribution at 0.055 and 30. We consider that this function has four continuous bounded partial derivatives, so that $R = 3$. In line with assumption A3, we choose the triweight kernel $(35/32)(1 - u^2)^3 1\!\!I(|u| \leq 1)$ for the three kernels involved in our first step estimators.

We choose the bandwidths according to A4.AN. In particular, we consider $h_G = 1.06\hat{\sigma}_x(IL)^{-1/6.5}$, $h_{1g} = 1.06\hat{\sigma}_x(IL)^{-1/4.5}$, $h_{2g} = 1.06\hat{\sigma}_b(IL)^{-1/4.5}$, where $\hat{\sigma}_b$ and $\hat{\sigma}_x$ are the estimated standard deviations of observed bids and object heterogeneity, respectively. The factor 1.06 follows from the so–called rule of thumb (see Härdle (1991)). The use of I arises because we have I bidders per auction.[9]

[9]To replicate the GPV (2000) estimator we choose the bandwidths according to the optimal rates.

We use 1,000 replications. More precisely, for each replication we generate first private values using the truncated log–normal distribution. Next we compute the corresponding bids $B_{p\ell}$ using (2.2). With these observed bids we can now apply our estimation procedure for each replication. First, we estimate the distribution and density functions of observed bids using (2.5) and (2.6).

Next, we compute pseudo private values $\hat{V}_{p\ell}$ corresponding to $B_{p\ell}$ as

$$\hat{V}_{p\ell} = B_{p\ell} + \frac{1}{I-1}\hat{G}(B_{p\ell}|Z_\ell)/\hat{g}(B_{p\ell}|Z_\ell)$$

for $p = 1, \ldots, 5$ and $\ell = 1, \ldots, L$.

In a second step we formulate the following (infeasible) moment condition

$$\mathrm{E}\left[1\!\!I(\underline{V} \le V \le \overline{V}) \frac{\partial \ln f(V|I, X; \theta_0)}{\partial \theta} \right] = 0.$$

Thus our sample moment condition can be expressed as follows

$$\frac{1}{IL}\sum_{\ell=1}^{L}\sum_{p=1}^{I} 1\!\!I(0.055 \le \hat{V}_{p\ell} \le 30)\frac{\partial \ln f(\hat{V}_{p\ell}|I, X; \hat{\theta})}{\partial \theta} = 0.$$

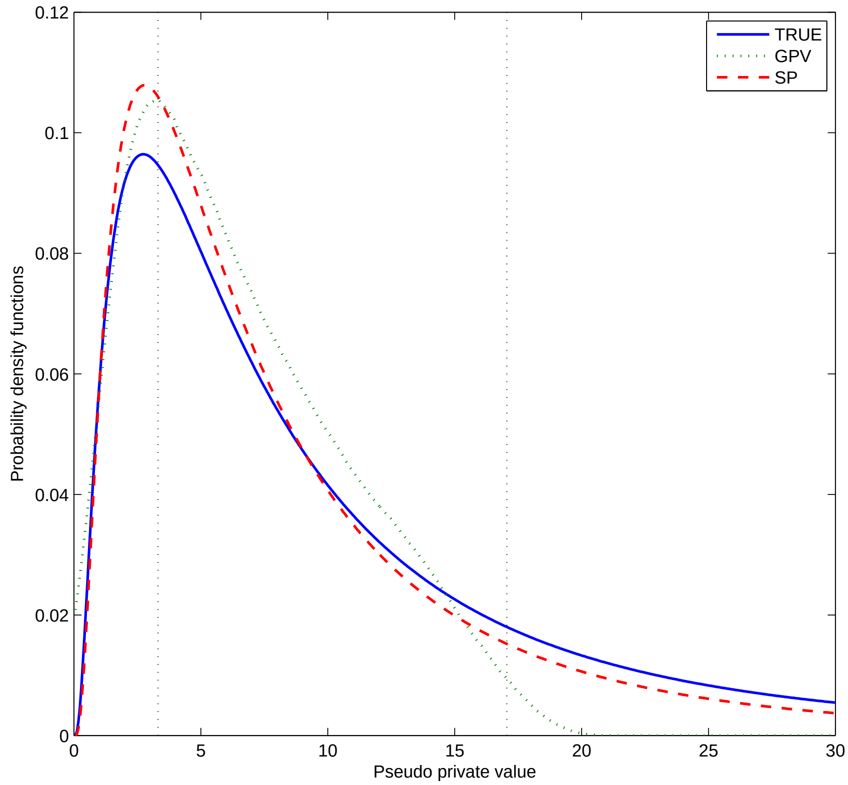
Using the above condition we implement an efficient GMM procedure to obtain $\hat{\theta}$.

We represent our results for a fix value of X. In particular, we set X equal to its median. Figure 2.1 below shows the true density of private values against the two estimators we are comparing. The semiparametric estimator developed in this chapter (dashed line) does a good job in matching the true density. Moreover, when comparing with the nonparametric GPV (2000) estimator, we can see that our semiparametric estimator is not subject to boundary effects.[10]

Thus, the order of the bandwidths is $L^{-1/9}$ for h_G and the second step bandwidth h_x and $L^{-1/10}$ for h_{gb} and h_{gx} and the second step bandwidths h_{fv} and h_{fx}. Specifically we use $h_G = 1.06\hat{\sigma}_x(IL)^{-1/9}$, $h_{gx} = 1.06\hat{\sigma}_x(IL)^{-1/10}$, $h_{gb} = 1.06\hat{\sigma}_b(IL)^{-1/10}$ where $\hat{\sigma}_b$ and $\hat{\sigma}_x$ are as defined above. The second step bandwidths are $h_{fv} = 1.06\hat{\sigma}_{\hat{v}}(n_t)^{-1/10}$, $h_{fx} = 1.06\hat{\sigma}_x(n_t)^{-1/10}$ and $h_x = 1.06\hat{\sigma}_x(L)^{-1/9}$, where n_t is the number of observations remaining after trimming.

[10]The vertical lines in the figure correspond to the trimming we have conducted for the GPV (2000) estimator to which one h_f is added (and subtracted) to eliminate remaining boundary effects.

Figure 2.1. Densities of bidders' private values



2.5 A More General Class of Models

In this section we indicate how to extend our procedure to a more general class of auction models. To keep the notation as simple as possible we consider models without observed object heterogeneity. This is not restrictive since the distribution and density functions can be replaced by with their conditional counterparts.

Binding Reserve Price

The first natural extension of the model considered in Section 2.2 is the symmetric IPV first–price auction model with a binding reserve price, announced or random.

Announced Reserve Price

An announced binding reserve price ($p_0 > \underline{V}$) constitutes a screening device for participating in the auction. As pointed out by GPV (2000) the Bayesian–Nash equilibrium strategy is still given by (2.2) in this set up, but the number I of potential bidders becomes unobserved and typically different from the observed number, I^*, of actual bidders who have submitted a bid ($\geq p_0$). Hence the model has a new structural element, namely I, in addition to the latent distribution of bidders' private values. As shown in GPV (2000), the differential equation defining the equilibrium strategy can be rewritten as

$$V_i = \xi_0(B_i, G_0^*, F(p_0), I) = B_i + \frac{1}{I-1}\left(\frac{G_0^*(B_i)}{g_0^*(B_i)} + \frac{F(p_0)}{1-F(p_0)}\frac{1}{g_0^*(B_i)}\right),$$

for $i = 1, \ldots, I^*$ and where $G_0^*(\cdot)$ is the truncated distribution of an observed bid conditional upon the fact that the corresponding private value is grater than or equal to p_0. Provided one can estimate I and $F(p_0)$ this equation is the basis for a two step procedure analogous to that of section 2.2.[11]

Random Reserve Price

In some cases, as in timber and wine auctions, the seller may decide not to announce the reserve price at the time the auction takes place. Hence, the reserve price is said to be secret or random. Since bidders do not know it when submitting their bids, this fact brings into the model a new kind of uncertainty that has to be taken into account. To present the basic equation underlying our two–step procedure in this model we need first to introduce additional notation. Let V_0 be the private value of the risk–neutral seller for the auctioned object. Moreover, we assume that V_0 is distributed according to $H(\cdot)$ defined on the same support as $F(\cdot)$ and that $H(\cdot)$ is common knowledge. Elyakime, Laffont, Loisel, and Vuong (1994) have shown that in a first–price sealed bid auction $p_0 = V_0$. In addition the bidders' equilibrium strategy is the solution of a differential equation which in general cannot be solved explicitly (see Li and Perrigne (2003)). However, this differential equation can be rewritten as follows

$$V_i = \xi_0(B_i, H, G_0, I) = B_i + \frac{1}{(I-1)\left(\frac{g_0(B_i)}{G_0(B_i)} + \frac{h(B_i)}{H(B_i)}\right)},$$

for $i = 1, \ldots, I$. As mentioned by Perrigne and Vuong (1999) since the reserve price is kept secret, all potential bidders submit a bid. Hence I is typically observed. The above equation can be used as the basis of a two–step procedure similar to the one described in section 2.2. In a first step, observed bids and reserve prices can be used to estimate the distribution $G_0(\cdot)$, its density $g_0(\cdot)$ as well as the distribution $H(\cdot)$ and its density $h(\cdot)$ nonparametrically. Next, pseudo private values can be recovered using the equation

[11] Introduction of heterogeneity across auctioned objects can be easily implemented if $p_{0\ell}$ is an unknown deterministic function of exogenous variables. See GPV (2000), section 4 for further details.

above in order to define a set of moment conditions for estimating the parameter of interest θ_0 in a second step.

The Symmetric Affiliated Private Value (APV) Model

To assume independence across private values can be restrictive since one can expect some degree of affiliation or positive correlation among private values. Thus, a second natural extension of our framework is to consider the more general class of model encompassed by symmetric APV models. Affiliation means that if one bidder draws a high valuation for the auctioned object, then others bidders are likely to draw higher valuations too. Laffont and Vuong (1996) study the problem of identification and theoretical restrictions in a general framework, namely in Affiliated Value (AV) models. In particular they show that any symmetric AV model is observationally equivalent to some symmetric APV model because the utility function is not identified from observed bids only.[12] Therefore when only data on observed bids are available the result in Laffont and Vuong (1996) implies that APV models can be considered without loss of generality, provided that we have identification.

We briefly indicate here how to adapt our estimation procedure to this kind of models. We assume that all bids are observed and that the reserve price is nonbinding. Let $Y_i = \max_{j \neq i} V_j$. The differential equation defining the equilibrium strategy in the APV model can be written as follows

$$V_i = \xi_0(B_i, G_0) \equiv B_i + \frac{G_{0,b_1|B_1}(B_i|B_i)}{g_{0,b_1|B_1}(B_i|B_i)},$$

for all V_i subject to the boundary condition $s(\underline{V}) = \underline{V}$, $G_{0,b_1|B_1}(x_1|X_1) = F_{Y_1|V_1}(s^{-1}(x_1))$ and $B_1 = s(Y_1)$. This equation is again the basis for the identification result and estimation procedure.

The theoretical restrictions as shown by Li, Perrigne, and Vuong (1999) indicate that the joint distribution of bids $G_0(\cdot)$ can be rationalized by a symmetric APV model if and only if (i) $G_0(\cdot)$ is symmetric and affiliated and (ii) the function $\xi_0(\cdot, G_0)$ is strictly increasing on its support. Moreover, if these two conditions are satisfied, then the joint distribution $F(\cdot)$ of private values is identified.

Regarding estimation, the equation above suggests a two–step procedure analogous to the one described in section 2.2. In the first step the ratio $G_{b_1|B_1}(\cdot|\cdot)/g_{b_1|B_1}(\cdot|\cdot)$ can be estimated nonparametrically and then pseudo private values can be recovered. In the

[12] Two auction models are said to be observationally equivalent given observed bids if they lead to the same equilibrium bid distribution.

second step a GMM procedure can be implemented to estimate the parameter of interest of the underlying distribution of private values.

Asymmetric Models

Assuming that bidders are ex ante identical may constitute a limitation. Therefore, in some cases one needs to use models relaxing this assumption. However, a common feature shared by asymmetric auction models is that they lead to systems of differential equations without a closed form solution. Hence, the direct approach becomes extremely difficult to implement. Nevertheless, using our indirect two–step procedure, asymmetric models can be structurally estimated while avoiding solving for the equilibrium strategy as well as of its inverse.

The Asymmetric IPV Model

Following the exposition in Perrigne and Vuong (2008) we assume that asymmetry is ex ante known to all bidders. Let $F_1(\cdot), \ldots, F_I(\cdot)$ be the private value distributions of the I bidders whose identities are observed and let $G_1(\cdot), \ldots, G_I(\cdot)$ be the corresponding bid distributions.[13] We can express the intractable system of differential equations as follows

$$V_i = B_i + \frac{1}{\sum_{j \neq i} \frac{g_j(B_i)}{G_j(B_i)}}$$

for $i = 1, \ldots, I$.

The above system of equations leads naturally to a two–step procedure similar to the one proposed in section 2.2

The Asymmetric APV Model

For simplicity we consider only two types of bidders. That is, the model assumes that the I-dimensional vector $(V_{11}, \ldots, V_{1I_1}, V_{01}, \ldots, V_{0I_0})$ has distribution function $F(\cdot)$ which is exchangeable in its first I_1 and last I_0 arguments. We can interpret this structure as follows. There is symmetry within each subgroup as bidders of the same type are assumed to be ex ante identical. Furhter, since $F(\cdot)$ is affiliated, there is general positive dependence among private values. From Campo, Perrigne, and Vuong (1998) the system of differential equation defining equilibrium strategies is

$$V_1 = \xi_1(B_1, G) \equiv B_1 + \frac{G_{B_1^*, B_0|B_1}(B_1, B_1|B_1)}{\partial G_{B_1^*, B_0|B_1}(B_1, B_1|B_1)/\partial T},$$

[13] In the context of procurements, Flambard and Perrigne (2006) use an asymmetric model to analyze snow removal contracts in Canada.

$$V_0 = \xi_0(B_0, G) \equiv B_0 + \frac{G_{B_1,B_0^*|B_0}(B_0, B_0|B_0)}{\partial G_{B_1,B_0^*|B_0}(B_0, B_0|B_0)/\partial T},$$

where $B_j^* = \max_{i\neq 1, i\in G_j} B_{ji}$, $B_j = \max_{i\in G_j} B_{ji}$, for $j = 1, 0$ and the partial derivatives with respect to T indicate the total derivative with respect to the first two arguments. Campo, Perrigne, and Vuong (1998) establish that $F(\cdot, \ldots, \cdot)$ is identified. Moreover they use a nonparametric two–step procedure following GPV (2000) to estimate the model. We propose instead to use our two–step semiparametric procedure using the above system of equations to recover pseudo private values after obtaining nonparametric estimates for $G_{B_1^*,B_0|B_1}(\cdot, \cdot|\cdot)$ and $G_{B_1,B_0^*|B_0}(\cdot, \cdot|\cdot)$. Next in a second step a model for the private values is specified through a set of moment conditions.

2.6 Conclusions

In this chapter we develop an indirect procedure to estimate first–price sealed–bid auction models, contributing in this way to the structural analysis of auction data that has been developed over the last fifteen years. Following GPV (2000) our procedure is in two steps. The difference with GPV (2000) is that our second step is implemented using a GMM procedure so that our resulting model is semiparametric. We show that our semiparametric estimator converges at the parametric $\sqrt{L}$ rate while the nonparametric estimator in GPV (2000) was shown to converge at the best possible rate according to the minimax theory which is slower than the parametric rate. Moreover, our procedure is not subject to the curse of dimensionality or in other words the convergence rate is independent of the dimension of the exogenous variables. We establish consistency and asymptotic normality of our estimator.

Given the nature of our procedure it is not necessary to solve for the equilibrium strategy or its inverse explicitly. This is a valuable advantage with respect to direct methods especially when estimating models that lead to intractable first–order conditions, such as asymmetric auction models. More generally, our method extends to models which have been estimated using a nonparametric indirect procedure. In this respect, we briefly outline how this can be done in models with a binding reserve price (announced or random), affiliated private value models and asymmetric models.

Finally, we conducted a set of Monte Carlo simulations. The main purpose of this was to assess the performance of our estimator in finite samples relative to the nonparametric estimator proposed by GPV (2000). Our semiparametric estimator does a good job in matching the true density. Moreover, when comparing with the nonparametric GPV

(2000) estimator, we can see that the estimator developed in this chapter is not subject to boundary effects.

Chapter 3

Detecting Collusion on Highway Procurement

3.1 Introduction

Despite the vulnerability to bidder collusion, auctions and procurements are widely used mechanisms for allocating goods and services. Most government acquisitions are competitively procured (see Kelman (1990)). As Marshall and Meurer (2001) point out, construction and highway projects are typically procured by the government, assets of bankrupt businesses are usually liquidated by means of an auction, the federal government is the biggest auctioneer in the U.S. and offshore oil leases as well as timber from national forests are sold by means of auctions.

Marshall and Meurer (2001) argue that criminal and civil enforcement of the antitrust laws has deterred price–fixing in some market settings, but not bidder collusion. In particular they mention that many cases in the 1980s and more recent high–profile cases serve as a reminder that the success of anti–collusive policies is limited in auction and procurement markets. The construction industry has experienced many instances of bid–rigging (see Bajari (2001)).

There are several reasons why a good understanding of cooperative behavior in auctions or procurements is desired. The argument made most frequently in the literature is that collusion creates inefficiencies. In an auction context, the traditional view is that bidder collusion depresses seller revenue. In particular, in markets involving the government as a buyer or seller it is argued that collusion leads to increased government expenditures at procurements and decreased revenues at auctions. An additional problem in the case of the government is that raising governments funds through distortionary taxes creates inefficiencies. Thus, the increased revenue spent in procurements because of collusion is not simply a wealth transfer.

Identifying the characteristics of competitive behavior is a necessary first step towards collusion detection. Using the structural approach to analyze auction data, the main objective of this chapter is to develop a methodology to detect collusive behavior. The idea is to compare two alternative models. Both models share the feature that bidders are allowed to be ex ante asymmetric (across types). As argued by Bajari (1997) and Bajari and Ye (2003), realistic models of bidding for procurement contracts should consider asymmetries among bidders. There are many sources that can create asymmetries among which the most important ones are location and capacity constraints. Other reasons often cited in the literature are different managerial skills, different information about the project, and the presence of a bidding ring, to name a few.

In this chapter we identify two different sets of firms as potential ring members. Relying on an exogenous number of bidders and the assumption that within each type bidders are symmetric, we find evidence supporting the collusive scheme for the two mentioned sets of firms.

This chapter is organized as follows. Section 3.2 outlines the theoretical model that leads to the econometric model. We first discuss the maintained assumptions throughout the chapter. Then we present the theoretical framework that encompasses both, the competitive model and the collusive model. In this section we also provide some arguments for distinguishing the two competing models. Section 3.3 contains the econometric methodology followed in this chapter. A description of the data and the market for construction projects in California is given in section 3.4. Section 3.5 shows how we classify different firms into different types of firms. In section 3.6 we present the main results of this chapter and section 3.7 concludes. Finally, Appendix B collects some practical issues; in particular the choices of kernels and bandwidths used to estimate the models are discussed.

3.2 Structural Analysis

In this section we describe the environment for a procurement model with private information in which firms compete for a construction project. Specifically we consider a first–price sealed–bid auction within the IPV paradigm with asymmetric bidders and an exogenous number of bidders. First we discuss the assumptions maintained throughout the chapter in the following section. Then, we introduce the case in which firms bid competitively and after that we adapt the model to allow for collusion.

3.2.1 Assumptions

A single and indivisible project is procured to $N^* \geq 2$ risk neutral potential bidders. We assume that bidders of type $j = 0, 1, 2$ draw their private costs independently from a distribution $F_j(\cdot)$. We further assume that cost distributions depend on the number of bidders only via X which is a vector of exogenous characteristics. In other words the distributions of private costs do not depend directly on the number of bidders. We will call this the exogeneity assumption. Therefore, the number of bidders is exogenous. That is, firms do not make entry decisions on the basis of perceived profitability. Thus, the number of potential bidders, N^*, and the number of actual bidders, N, is the same. This implies that the reserve price is nonbinding. An announced binding reserve price, p_0, or an entry fee, e, are screening devices for participating in the auction. As pointed out by Perrigne and Vuong (1999) this complicates the nonparametric identification and estimation of the model.

There are few theoretical models in the literature which address endogenous entry decisions by means of a two–stage game (see e.g. Levin and Smith (1994)). However, in this kind of models the participation decision and the bidding decision of each firm are independent. This implies that in the second stage the bidding behavior is basically the same as the one described in this chapter. To the best of our knowledge, there is no model in the literature considering a two–stage game in which both decisions are correlated. Thus, this is outside the scope of this chapter.

Let N_0, N_1 and N_2 denote the number of participants for type 0, 1, and 2, respectively, which are observed by all firms. In other words, firm i knows its actual competitors.

The distribution of private costs is given by $F(\cdot, \ldots, \cdot) = F_0^{N_0}(\cdot)F_1^{N_1}(\cdot)F_2^{N_2}(\cdot)$. These three distributions are common knowledge with common support $[\underline{c}, \overline{c}]$. Let $f_j(\cdot)$ denote the corresponding densities which are assumed to be continuously differentiable and bounded away from zero on their support.

3.2.2 Model for Competitive Bidding (Model A)

In the competitive framework, group 1 characterizes large firms that bid simultaneously (on a pairwise basis) more than a handful of times. Group 2 contains the remaining large firms and group 0 the other (small) bidders.

Each firm i of type j submits a bid, b_{ij}, which depends on its own project cost c_{ij}. Firm i maximizes its expected profit.

The expected profit of type $j = 0, 2$ bidders is ,

$$\begin{aligned}\pi_{ij} &= (b_{ij} - c_{ij})\Pr\left(b_{ij} < \min_{k \neq i} B_{kj}\right)\Pr\left(b_{ij} < \min_{k=1,\dots,n_1} B_{k1},\right)\Pr\left(b_{ij} < \min_{k=1,\dots,n_{(2-j)}} B_{k(2-j)}\right)\\ &= (b_{ij} - c_{ij})\left(1 - F_j[s_j^{-1}(b_{ij})]\right)^{n_j - 1}\left(1 - F_1[s_1^{-1}(b_{ij})]\right)^{n_1}\left(1 - F_{(2-j)}[s_{(2-j)}^{-1}(b_{ij})]\right)^{n_{(2-j)}}.\end{aligned}$$

For type 1 bidders it is

$$\begin{aligned}\pi_{i1} &= (b_{i1} - c_{i1})\Pr\left(b_{i1} < \min_{k=1,\dots,n_0} B_{k0}\right)\Pr\left(b_{i1} < \min_{k \neq i} B_{k1}\right)\Pr\left(b_{i1} < \min_{k=1,\dots,n_2} B_{k2}\right)\\ &= (b_{i1} - c_{i1})\left(1 - F_0[s_0^{-1}(b_{i1})]\right)^{n_0}\left(1 - F_1[s_1^{-1}(b_{i1})]\right)^{n_1 - 1}\left(1 - F_2[s_2^{-1}(b_{i1})]\right)^{n_2}\end{aligned}$$

where $s_j(\cdot)$ denotes type $j's$ equilibrium strategy.

Lebrun (1996, 1999) and Maskin and Riley (2000a,b, 2003) among others, have studied the existence and uniqueness of the Bayesian–Nash equilibrium in asymmetric first–price, sealed–bid auctions.[1] It is known from this literature that the equilibrium strategies $s_0(\cdot)$, $s_1(\cdot)$ and $s_2(\cdot)$ satisfy the following system of differential equations.

$$\begin{aligned}s_j'(c_j, b_j, n) &= (b_j - c_j)\left[(n_j - 1)\frac{f_j(c_j)}{1 - F_j(c_j)} + n_1\frac{f_1(s_1^{-1}(b_j))}{1 - F_1(s_1^{-1}(b_j))}\frac{s_0'(c_0)}{s_1'(s_1^{-1}(b_j))}\right.\\ &\quad\left. + n_{(2-j)}\frac{f_{(2-j)}(s_{(2-j)}^{-1}(b_j))}{1 - F_{(2-j)}(s_{(2-j)}^{-1}(b_j))}\frac{s_j'(c_j)}{s_{(2-j)}'(s_{(2-j)}^{-1}(b_0))}\right] \quad \text{for } j = 0, 2\end{aligned}$$

$$\begin{aligned}s_1'(c_1, b_1, n) &= (b_1 - c_1)\left[n_0\frac{f_0(s_0^{-1}(b_1))}{1 - F_0(s_0^{-1}(b_1))}\frac{s_1'(c_1)}{s_0'(s_0^{-1}(b_1))} + (n_1 - 1)\frac{f_1(c_1)}{1 - F_1(c_1)}\right.\\ &\quad\left. + n_2\frac{f_2(s_2^{-1}(b_1))}{1 - F_2(s_2^{-1}(b_1))}\frac{s_1'(c_1)}{s_2'(s_2^{-1}(b_1))}\right], \end{aligned} \tag{3.1}$$

subject to the boundary conditions $s_0(\underline{c}) = s_1(\underline{c}) = s_2(\underline{c})$, and $s_0(\overline{c}) = s_1(\overline{c}) = s_2(\overline{c}) = \overline{c}$. The above system of equations does not have a closed form solution, introducing a major difficulty for estimating the model. In this way, the application of direct estimation procedures to field data becomes cumbersome and only some numerical methods could be use but they require the numerical determination of the equilibrium strategies for any

[1] The assumptions about c_{ij} and f_j described above guarantee that the type–specific equilibrium of this game exists and is unique.

trial parameter value (see Perrigne and Vuong (1999, 2008) for further details).

To complete the specification of the econometric model that follows from the theoretical model, let $G_j(\cdot)$ be the distribution of bids corresponding to bidders of type $j = 0, 1, 2$, and let $g_j(\cdot)$ denote the corresponding density. Following a similar argument to achieve identification as the one in Guerre, Perrigne, and Vuong (2000), system (3.1) above can be expressed as follows

$$c_j = b_j - \frac{1}{(n_j - 1)\frac{g_j(b_j)}{1-G_j(b_j)} + n_1 \frac{g_1(b_j)}{1-G_1(b_j)} + n_{(2-j)} \frac{g_{(2-j)}(b_j)}{1-G_{(2-j)}(b_j)}} \quad \text{for } j = 0, 2$$

$$c_{1A} = b_1 - \frac{1}{n_0 \frac{g_0(b_1)}{1-G_0(b_1)} + (n_1 - 1)\frac{g_1(b_1)}{1-G_1(b_1)} + n_2 \frac{g_2(b_1)}{1-G_2(b_1)}}. \tag{3.2}$$

This set of conditions establishes that *unobserved* private costs are identified from *observed* bids and bidders' identities.

3.2.3 Model for Efficient Collusion (Model B)

The environment in which the collusive game takes place is similar to the one for the competitive model. That is, we assume that firms are engaged in a first–price sealed–bid auction where they compete for construction projects. The difference with the previous model arises in the way type 1 firms decide their bidding strategies.

In order to adapt the model developed above, we assume that the cartel behaves efficiently. This assumption can be justified if one thinks that there are side payments among ring members, a practice that has been used in cases in which collusion has been detected. We assume that side payments are cleared among all cartel firms before the target auction takes place. The cartel operates as follows; all ring members submit bids according to the strategy given by the first order conditions (see below). Moreover, the designated winner is the firm with the lowest cost. Hence, cartel members communicate before an auction is conducted to compare their cost estimates. Therefore the model in section 3.2.2 can be adapted to this case. Thus, this model is a special case of the asymmetric IPV framework described above.

Under efficient collusion, both competitive firms and cartel firms participate in an auction. As before, there are 3 types of bidders. We label cartel firms as type 1 bidders. Large competitive firms are named as type 2 firms and small competitive (fringe firms) will be type 0 bidders. From the perspective of a type 1 bidder, there is only one such firm participating (seriously) in an auction. Hence, $N_1 = 1$ for this group of firms.[2] As before, there are N_0 and N_2 bidders of type 0 and type 2, respectively. We maintain the assumption that bidders of type j draw their private costs independently from a distribution $F_j(\cdot)$, $j = 0, 1, 2$.

We present now the maximization problem for each type of bidder and derive the first order conditions that private costs satisfy under a Bayes–Nash equilibrium. The expected profits for this model are as follows.

For type $j = 0, 2$ bidders

$$\begin{aligned}\pi_{ij} &= (b_{ij} - c_{ij})\Pr\left(b_{ij} < \min_{k \neq i} B_{kj}\right)\Pr\left(b_{ij} < \min_{k=1,\ldots,n_1} B_{k1},\right)\Pr\left(b_{ij} < \min_{k=1,\ldots,n_{(2-j)}} B_{k(2-j)}\right) \\ &= (b_{ij} - c_{ij})\left(1 - F_j[s_j^{-1}(b_{ij})]\right)^{n_j - 1}\left(1 - F_1[s_1^{-1}(b_{ij})]\right)^{n_1}\left(1 - F_{(2-j)}[s_{(2-j)}^{-1}(b_{ij})]\right)^{n_{(2-j)}}.\end{aligned}$$

For type 1 bidders,

$$\begin{aligned}\pi_{i1} &= (b_{i1} - c_{i1})\Pr\left(b_{i1} < \min_{k=1,\ldots,n_0} B_{k0}\right)\Pr\left(b_{i1} < \min_{k=1,\ldots,n_2} B_{k2}\right) \\ &= (b_{i1} - c_{i1})\left(1 - F_0[s_0^{-1}(b_{i1})]\right)^{n_0}\left(1 - F_2[s_2^{-1}(b_{i1})]\right)^{n_2}.\end{aligned}$$

Side payments are not included in the expected profits for type 1 bidders given that they are paid before the target auction whether or not the bidder wins (see above).

The first order conditions derived from the above maximization problem, after using a similar argument for identification as the one in Guerre, Perrigne, and Vuong (2000), can be expressed as follows

$$c_j = b_j - \frac{1}{(n_j - 1)\frac{g_j(b_j)}{1 - G_j(b_j)} + n_1\frac{g_1(b_j)}{1 - G_1(b_j)} + n_{(2-j)}\frac{g_{(2-j)}(b_j)}{1 - G_{(2-j)}(b_j)}} \quad \text{for } j = 0, 2$$

and

$$c_{1B} = b_1 - \frac{1}{n_0\frac{g_0(b_1)}{1 - G_0(b_1)} + n_2\frac{g_2(b_1)}{1 - G_2(b_1)}}, \tag{3.3}$$

Notice that under efficient collusion, expected profits, π_{i1}, are different only for type 1 bidders. Hence, the expression for private costs is also different for this type of bidders as compared to the competitive case.

[2]What the efficient cartel actually does is to limit the level of competition among its members.

3.2.4 Comparing the Two Alternative Models

As shown in sections 3.2.2 and 3.2.3, for each competing model there is an expression for private costs for each type of bidder (see equations (3.2) and (3.3)).

The idea now is to test whether the data are best explained by an Asymmetric IPV model in which there is no collusion (Model A), or by an Asymmetric IPV model in which type 1 bidders collude efficiently (Model B). To control for possible heterogeneity across auctions we consider the variable $X_\ell \in I\!R^p$, $\ell = 1, \ldots, L$, which denotes relevant characteristics of the ℓth project.[3] Therefore all the distributions are conditional distributions namely $F_j(\cdot|x_\ell)$ and $G_j(\cdot|x_\ell)$. In particular, bid distributions depend on the number of bidders. We further assume that the vectors $(X_\ell, N_{0\ell}, N_{1\ell}, N_{2\ell})$, are independent and identically distributed across ℓ.[4]

In order to be able to distinguish which of the two models described above best explains the behavior of bidders during the sample period considered, we rely on the following observation. The underlying distributions of private costs in the "right" model should not change with the number of bidders. This is a direct consequence of our exogeneity assumption. In other words, if it were the case that bidders are just asymmetric and therefore Model A is the appropriate one to use, then the underlying distributions of private costs for each type of bidder should be the same regardless of how many bidders participate in an auction when the competitive model is estimated. At the same time, under this scenario we expect to see more variation in cost distributions for the collusive model across N. On the other hand, if type 1 bidders indeed act as an efficient cartel so that Model B is the relevant one to use, it must also be the case that the distributions of private costs do not change with the number of bidders and again some variation is expected for the competitive framework in this case.

It is important to emphasize that failure to see unchanged cost distributions as the number of bidders varies could also be due to asymmetries within types. The model allows for asymmetric bidders across types, but symmetry is assumed for bidders of the same type. It is also possible that not all ring members are included in the group defining type 1 bidders. Moreover, other (non-identified) cartels could be operating during the sample period as well. If this were the case, then these firms would be (wrongly) labeled as type 0 or type 2 bidders thus yielding misleading results.

Finally, the comparison of the models boils down to distinguishing between the cost

[3] In the empirical application $p = 1$ and X_ℓ is the Engineers' Estimate for the ℓth auction.

[4] Notice that this assumptions allows the number of bidders to depend upon the characteristics of the project since it does not require independence between X_ℓ and the number of bidders.

distributions for type 1 bidders in each competing model.

3.3 The Econometric Approach

In this section we outline the econometric strategy implemented to obtain the distribution of private costs for each subgroup. In the same spirit as Guerre, Perrigne, and Vuong (2000) we use a two step nonparametric procedure. In the first step we apply kernel methods to estimate the distribution and density of observed bids for each group of bidders. The second step then uses these estimated functions to recover pseudo private costs which are used to obtain the corresponding estimated densities.

We first discuss some practical issues. The skewness of the bid distribution is a typical problem encountered with auction data. In addition, the use of kernel estimators is subject to the so–called boundary effect so that some kind of trimming is often used.[5] As a consequence it is common practice among empirical researchers to use a logarithmic transformation in order to keep a substantial number of observations after trimming (see for example Li and Perrigne (2003)). For notational simplicity we suppress the dependance of the distributions on (X, N). Later, when presenting the estimators we include these variables explicitly. Applying the log transformation to system (3.2) yields

$$c_j = \xi_j(d_j, n) = 10^{d_j} - \frac{10^{d_j}}{(n_j - 1)\frac{g_{jd}(d_j)}{1-G_{jd}(d_j)} + n_1\frac{g_{1d}(d_j)}{1-G_{1d}(d_j)} + n_{(2-j)}\frac{g_{(2-j)d}(d_j)}{1-G_{(2-j)d}(d_j)}}$$

for j=0,2 and

$$c_{1A} = \xi_1(d_1, n) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + (n_1 - 1)\frac{g_{1d}(d_1)}{1-G_{1d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}}. \tag{3.4}$$

For system (3.3) the transformed system of equations differs from system (3.4) above only in the expression for $c_1(\cdot)$, namely

$$c_{1B} = \xi_1(d_1) = 10^{d_1} - \frac{10^{d_1}}{n_0\frac{g_{0d}(d_1)}{1-G_{0d}(d_1)} + n_2\frac{g_{2d}(d_1)}{1-G_{2d}(d_1)}} \tag{3.5}$$

where $d_j = \log(b_j)$, $G_{jd}(\cdot)$ is the cdf of $\log(b_j)$ and $g_{jd}(\cdot)$ is its corresponding density, for $j = 0, 1, 2$.

[5]To avoid trimming we could have used LPEs instead of kernels in the first step. However, here it does not matter because we are mainly interested in assessing the center of the distributions of private costs.

As noted earlier, some kind of trimming is often needed due to the bad behavior of kernel estimators close to the boundaries of the support of bids. In line with Guerre, Perrigne, and Vuong (2000) we adopt the following:

$$\hat{c}_{ji\ell} = \begin{cases} \xi_j(d_{i\ell}) & \text{if } d_{\min} + \varrho \max\{h_g, h_G\}/2 \leq d_{i\ell} \leq d_{\max} - \rho \max\{h_g, h_G\}/2; \\ +\infty & \text{otherwise.} \end{cases}$$

for $j = 0, 1, 2,\ i = 1, \ldots, n_j$ and $\ell = 1, \ldots, L$, where d_{min} and d_{max} are the minimum and maximum of log bids respectively, h_g, h_G are bandwidths and ϱ is the length of the support of the kernel.[6]

Let $S_{dj}(d|x,n) = \Pr(D \geq d|x,n)$. Then, the *hazard rate* functions involved in the expressions for private costs given by the system of equations (3.4) and (3.5) can be written as follows,

$$\frac{g_{dj}(d|x,n)}{1 - G_{dj}(d|x,n)} = \frac{g_{dj}(d|x,n)}{S_{dj}(d|x,n)} = \frac{g_{dj}(d,x,n)}{S_{dj}(d,x,n)}$$

for $j = 0, 1, 2$. Let T_j denote the total number of observations for bidders of type j. I consider L auctions in which different types of bidders participate. Thus bidder i, $i = 1, \ldots, N_j$ of type j participates in auction $\ell = 1, \ldots, L$. Relabeling bidders such that $k = (i, \ell)$, i.e. the ith bidder in auction ℓ, the sample consists of observation (d_k, x_k, n_k).[7,8] Thus, the estimators involved in the first step are

$$\hat{g}_j(d,x,n) \quad = \quad \frac{1}{T_j h_g^{p+1}} \sum_{k=1}^{T_j} K_g\left(\frac{d - D_k}{h_g}, \frac{x - X_k}{h_g}, \frac{n - N_k}{h_{gn}}\right)$$

and

$$\hat{S}_j(d,x,n) = \frac{1}{T_j h_{Gx}^p} \sum_{k=1}^{T_j} \mathbb{1}(d_k \geq d) K_G\left(\frac{x - X_k}{h_G}, \frac{n - N_k}{h_{Gn}}\right).$$

With the sample of pseudo private costs $\hat{C}$ we estimate the cost densities in a second

[6]Without loss of generality we set $d_{\min} = 0$.

[7]To keep the notation simple, we just include n_k in the formulas above. However, for the computation of the estimator we have used n_{0k}, n_{1k} and n_{2k} separately.

[8]Recall that X characterizes auction heterogeneity, thus it only varies across auctions. In terms of the notation used this means that $X_k = X_\ell$. In other words, for each auction ℓ the value x is the same for all bidders participating in that auction. A similar argument applies to the number of bidders, N_ℓ.

step as follows,

$$\hat{f}_j(c|x,n) = \frac{\hat{f}_j(c,x,n)}{\hat{q}_m(x,n)},$$

where

$$\hat{f}_j(c,x,n) = \frac{1}{T_j h_f^{p+1}} \sum_{k=1}^{T_j} K_f \left(\frac{c - \hat{C}_k}{h_f}, \frac{x - X_k}{h_f}, \frac{n - N_k}{h_{fn}} \right)$$

and

$$\hat{q}_m(x,n) = \frac{1}{T_j h_q^p} \sum_{k=1}^{T_j} K_q \left(\frac{x - X_k}{h_q}, \frac{n - N_k}{h_{qn}} \right).$$

The functions $K_g(\cdot)$, $K_G(\cdot)$, K_f and $K_q(\cdot)$ are kernels. The bandwidths for the continuous variables are denoted h_G, h_g, h_q and h_f. The bandwidths for the discrete variables are h_{Gn}, h_{gn}, h_{qn} and h_{fn}. Appendix B discusses the choices of kernels and bandwidths.

3.4 Data Description and Awarding Process

The California Department of Transportation (Caltrans) allocates construction projects using a First Price Sealed–Bid mechanism. The awarding process used by Caltrans is subject to Federal Acquisition Regulations and is therefore similar to other states' procedures. This process is conducted in three steps: First, the Caltrans Headquarters Office Engineer announces a project that is going to be let and invites firms to submit bids. This corresponds to the *Advertising Period*, which lasts between 4 and 10 weeks depending on the size and complexity of the job. Second, potential bidders may submit sealed bids based on bid proposals that explain the project's characteristics. Third, on the letting day, the bids received are opened and ranked. The project is awarded to the lowest bidder, provided that the firm fulfills certain responsibility criteria. After each letting, the information about all bids and their ranking is made public. The winning firm is awarded the job no more than 30 days after the letting date.

This section discusses the observable variables and provides descriptive statistics of the data. The sample consists of a subset of the Caltrans database on procurements of highway and road construction projects between January 2002 and January 2008.[9]

[9] We obtain the data from Caltrans web site: `http://www.dot.ca.gov/hq/esc/oe/awards/bidsum/`

During the sample period, Caltrans awarded 2,152 contracts for a total of $7,645 million. The information available on every project awarded consists of the *Bid Opening Date*, *Contract Number*, *Location*, *Number of Bidders*, *Number of Working Days*, *Engineers' Estimate*, *Amount of the Bid* and the *Rank of the Bid* for each of the bidding firms. Also there is information on the identity of each bidder and the address of the firm.

In line with the theoretical model we only consider auctions in which at least 2 bidders participate and the winning bidder is the one with the lowest bid. There are 1,907 such projects, with a total of $6,989 million. A total of 823 firms submit bids on at least one of these 1,907 projects. The subset containing big projects, which we define as those for which the engineers' estimate is at least $1 million, has 438 bidders in 780 contracts awarded for a total of $6,502 million, 85% of the total. Given that the main purpose of this chapter is to develop a methodology to detect potential collusive behavior, we restrict our attention further to big projects with engineers' estimates ranging between $1 and $20 million. In this subset there are 724 projects and 413 bidders participating with 202 winning at least once. The total value of the winning bids is $2,408 million, which represents 31% of the total. We make a first classification of bidders on the basis of their revenue share in the sample. Thus, there are 25 firms with at least 1% revenue share. We call these firms *Main Firms*. In the theoretical model there are three types of firms. Thus, it is reasonable to think that potential cartel candidates are among the main firms. Table 3.1 summarizes the bidding activity of the 25 main firms.

The first column in Table 3.1 gives the number of bids of each main firm. These bids represent 34% of all bids in the sample. The second and third column show the number of times each main firm has won a contract and the "expected number" of wins, respectively. For example, firm A bids on a total of 50 projects against a varying number of firms, n_ℓ for $\ell = 1, \ldots, 50$, then expected number of wins is defined to be $\sum_{\ell=1}^{50} 1/n_\ell$. By comparing these two columns it can be seen that with the exception of five firms, main firms tend to win more contracts than expected. In other words, this is suggesting that some firms win too often. The fourth column reports the average bid of each main firm in the sample and the fifth column the revenue share computed as the total value of the firm's winning bid as a fraction of the total value of winning bids for all contracts. The last column in Table 3.1 contains the participation rate (i.e. the bid frequency rate). There is variation in this rate across firms with a remarkable 44% for firm D.

Table 3.2 provides summary statistics for a subset of variables in the sample. The mean number of bidders per project is above four with most of the contracts receiving between two and five bids. On average the winning bid is $3.33 million. This number

Table 3.1. Revenue Shares and Participation of Main Firms

Firm ID	Number of Bids	Number of wins	Exp. Number of wins	Average bid (Mill. $)	Revenue Share	Participation rate
A	50	9	10.34	4.83	0.020	0.07
B	34	13	10.51	3.21	0.012	0.05
C	43	9	10.46	5.32	0.013	0.06
D	319	97	87.32	3.61	0.145	0.44
E	46	11	10.15	4.49	0.015	0.06
F	42	15	10.70	3.63	0.016	0.06
G	25	12	5.84	4.09	0.027	0.03
H	26	6	5.16	5.03	0.011	0.04
I	21	7	4.27	4.54	0.012	0.03
J	20	9	4.69	3.84	0.015	0.03
K	34	4	6.90	8.44	0.019	0.05
L	35	16	7.95	4.32	0.020	0.05
M	29	13	6.94	3.69	0.016	0.04
N	9	3	1.55	6.33	0.012	0.01
O	31	5	6.82	6.37	0.011	0.04
P	50	16	12.95	4.03	0.027	0.07
Q	33	9	6.31	3.35	0.017	0.05
R	28	10	8.10	3.48	0.012	0.04
S	47	12	8.82	4.37	0.021	0.06
T	25	13	5.99	3.75	0.021	0.03
U	68	16	15.22	4.77	0.026	0.09
V	26	7	4.78	5.75	0.025	0.04
W	41	11	7.18	2.92	0.019	0.06
X	41	7	10.27	4.50	0.021	0.06
Y	11	4	1.89	6.04	0.012	0.02
Total	1148	351	282		0.57	

Only firms with revenue shares $\geq 1\%$ are reported.

is smaller than the average engineers' estimate which is $3.77 million. The difference between the winning bid and the second lowest bid, "Money on the table", reveals the existence of imperfect information among bidders. In the sample this difference is on average $300,000. The engineers' estimate is highly positively correlated with bids (the correlation coefficient is 0.95). Despite this high correlation, it seems that the engineers' estimate is not binding as a screening device since in 30% of the cases the winning bid is above the engineers' estimate.

With the information in the database it is possible to construct measures of distance and backlog for each firm in each project. Distance is expressed in miles and refers to the distance between the location of each firm and that of the county where the project takes place. One would expect that closer firms have a cost advantage which should be reflected in bidding strategies. Even though there is a positive correlation between distance and bids in the sample, the magnitude of this correlation is small (0.012) suggesting that the location of the project does not influence bidding decisions much. The way the variable distance was constructed takes into account the longitudinal and latitudinal coordinates of the county where the project takes place and the coordinates corresponding to the zip code the firms have reported as their location. This variable is subject to measurement problems which could result in a low distance–bid correlation coefficient.

The variable backlog is defined as the sum of the dollar values of Caltrans contracts won but not yet completed by a particular firm. Firms are assumed to work at a constant pace during the working–days period of each project. To account for differences across firms, we have constructed a measure of capacity defined as the maximum backlog carried by a firm during the sample period. On average, firms' capacity is about 60% of the average bid, however there is considerable variation in this variable. The last variable constructed from the information in the sample is the utilization rate which is meant to measure a firm's backlog at a given point in time; it is defined as the ratio of backlog to capacity.[10]

Table 3.2. Summary Statistics

	No. observations	Mean	SD
No. Bidders	724	4.62	2.37
Winning bid	724	3.33	3.11
Money on the table	724	0.30	0.46
Engineers' Estimate	724	3.77	3.49
All Bids	3347	3.79	3.51
Backlog	3347	4.30	9.76
Distance (miles)	3347	123.98	162.93
Capacity (across firms)	413	2.30	5.69
Utilization rate	3347	0.20	0.32

All dollar figures are expressed in millions.

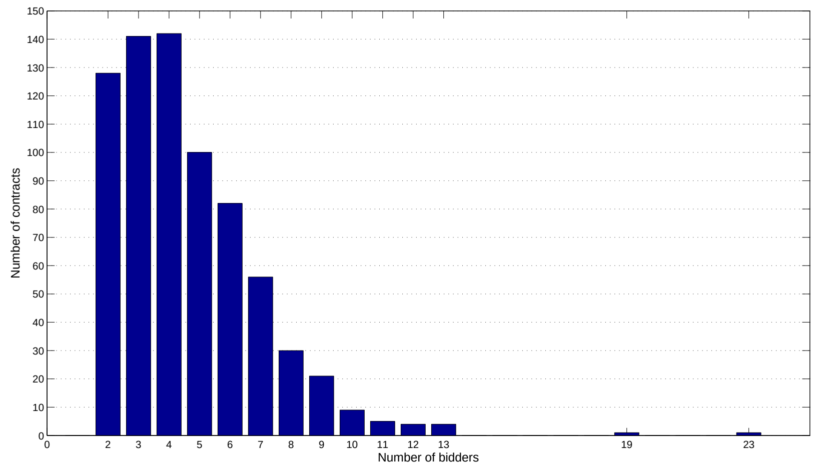
It is reasonable to expect that bidding rings would prefer to operate in markets with limited competition. Figure 3.1 below displays the distribution of the number of bidders per contract. The chart shows that most of the contracts have between two and five bidders with a peak at four. The fact the we see many few bidders is in line with the idea that competition is low in big projects.

3.5 Classifying Bidders

Recall that according to the theoretical model there are three types of bidders. Here we explain how we classify participating firms as type 0, type 1 or type 2 bidders. The key point is to determine which firms are considered type 1 firms, since then the remaining

[10] $\text{Util}_{\text{it}} = \text{Backlog}_{\text{it}}/\text{Cap}_{\text{i}}$ (if Cap=0, then Util=0 for all t).

Figure 3.1. Bidder Concentration



main firms will be considered type 2 bidders and all other (small) firms will be treated as type 0 bidders. The natural candidates for type 1 bidders are the 25 main firms. We start by looking at the number of simultaneous bids among these main firms on a pairwise basis, we select those pairs with at least fifteen simultaneous bids as potential type 1 bidders. The result is fifteen pairs of firms involving fifteen main firms. Table 3.3 below shows the pairs selected. In the first column the total number of simultaneous bids submitted by each pair are reported. The second column gives the "expected" number of wins in those projects computed according to the level of competition in each one. These two columns together reveal that main firms participate (simultaneously with another candidate) more than expected. The next two columns contain the actual number of times the first member of the pair (third column) as well as the second (fourth column) win a contract, respectively. Comparing the numbers in each of these columns to their expected counterparts in column two suggests that at least one member of the pair wins often which is in line with previous findings (see Table 3.1).

A couple of interesting features arising from the comparison between Table 3.1 and Table 3.3 are worth mentioning. First, firm A bids almost exclusively against firm D. Second, firm E bids remarkably frequently with both firm A and firm D. This triplet of firms could be in principle one candidate. Also from Table 3.3 it can be seen that firms D and P (along with (A,D)) have the largest number of simultaneous bids. Due to availability of data for each pair, we concentrate especially on those with a large number of simultaneous bids; the pair (D,P) constitutes a candidate in this respect. To further investigate the behavior of these pairs of firms we follow Bajari and Ye (2003). These

Table 3.3. Simultaneous Bids

Firm Pair	Simultaneous Bids	Expected Wins	First Bidder Wins	Second Bidder Wins
(A,D)	44	9.03	9	5
(A,E)	20	4.05	3	6
(B,D)	29	9.51	12	10
(C,D)	17	5.65	5	9
(D,E)	41	8.67	8	9
(D,F)	26	7.46	5	9
(D,H)	19	3.92	7	3
(D,I)	18	3.68	1	7
(D,O)	25	5.16	7	5
(D,P)	44	11.08	13	14
(D,R)	27	7.96	10	10
(D,V)	22	4.20	5	6
(D,W)	19	2.97	2	3
(M,X)	22	4.91	11	2
(W,X)	15	2.81	5	2

authors develop two conditions that must hold in equilibrium when bidding is competitive. The first condition states that conditional on observables, bids are independently distributed. The second condition refers to exchangeability of the bid distribution. As Bajari and Ye (2003) point out, these conditions may fail when bidding is collusive. In order to assess which pair of firms may be labeled as type 1 bidders we test for conditional independence and exchangeability.[11]

To test for independence we use a regression–based approach and consider the fifteen pairs of firms bidding frequently described above.[12] The model used is the following

$$\frac{BID_{i\ell}}{EE_\ell} = \beta_{0i} + \beta_{1i}LDIST_{i\ell} + \beta_{2i}CAP_{i\ell} + \beta_{3i}UTIL_{i\ell} + \beta_{4i}LMDIST_{i\ell} + u_{i\ell} \qquad (3.6)$$

$$\frac{BID_{i\ell}}{EE_\ell} = \alpha_0 + \alpha_1 LDIST_{i\ell} + \alpha_2 CAP_{i\ell} + \alpha_3 UTIL_{i\ell} + \alpha_4 LMDIST_{i\ell} + \varsigma_{i\ell} \qquad (3.7)$$

where the regressors have been already discussed above. $\mathrm{LDIST}_{i\ell}$ refers to the logarithm of distance and $\mathrm{LMDIST}_{i\ell}$ refers to the logarithm of the minimum of distances of all firms on project ℓ, excluding i.

Thus, if firm i is among the fifteen firms in Table 3.3, i.e. a main firm that frequently bids against another main firm, we use equation (3.6) with firm–varying coefficients. If

[11] This set of conditions are necessary for competitive bidding. However rejection does not imply that bidding is collusive.

[12] The main reason for conducting pairwise tests is basically driven by the amount of data. There are relatively few observations for the triplet (A,D,E) in the sample.

firm i is not one of the largest fifteen firms we use equation (3.7). For the estimation both equations are pooled and we include project fixed effects.

Let ρ_{ij} be the correlation between the residual to firm i's bid function and firm j's bid function, $\hat{u}_{i\ell}$ and $\hat{u}_{j\ell}$, respectively. We use Pearson's correlation test. Among all pairs, the null hypothesis of independence is rejected for all but one pair using a 5% two sided test.

Next, we test for exchangeability and, as before, we follow Bajari and Ye (2003) to construct two kinds of tests: Exchangeability at the Market Level by pooling the fifteen firms in one group and Exchangeability on a Pairwise basis. The null hypothesis of the test is: H_0: $\beta_{ik} = \beta_{jk}$ for all $i, j, i \neq j$ and for all $k = 1, \dots, 4$.

Let $T = 3,347$ be the number of observations, m the number of regressors and r the number of constraint implied by H_0. We consider the following statistic

$$F = \frac{(SSR_C - SSR_U)/r}{SSR_U/(T - m)}$$

which is asymptotically distributed as F with parameters $(r, T - m)$ under the null hypothesis.

At the market level, the restricted model imposes that the effect of the four explanatory variables is the same for potential cartel members and the remaining firms (i.e this is the exchangeability hypothesis). The null hypothesis of exchangeability is rejected when comparing the group of potential cartel members against the remaining bidders. Next, we conduct pairwise tests by pooling firms accordingly and find that the hypothesis of exchangeability is rejected at conventional levels for 13 out of 15 pairs including the pair (D,P) as well as (A,D) and (D,E).

Based on the previous analysis all pairs of firms considered do not pass at least one of the tests for competitive bidding. However, as mentioned above, taking into account the number of simultaneous bids, firms D and P bid simultaneously more than a handful of times. Also, the triplet (A,D,E) is chosen as a potential cartel candidate. Therefore for the subsequent analysis we concentrate on two groups of candidates, namely the pair (D,P) and the triplet (A,D,E) as type 1 bidders.

3.5.1 Summary Statistics for Type 1 Bidders

Firms D and P bid, on average, in projects of smaller size than the remaining thirteen large firms (i.e type 2 bidders in the model) and roughly of the same size as the small firms (type 0 bidders). At least one of the firms participates in 325 projects winning 113 out of

724 contracts with and average winning bid of $3.67 million. On average the engineers' estimate in these projects is above the winning bid. The average number of bidders participating in the 325 contracts is 4.65. Generally speaking, the data reveal that this pair tends to participate more often in small size projects with less competition. The other main firms tend to bid on larger projects and participate in 312 lettings. Type 0 bidders (i.e the remaining smaller firms in the sample) participate in almost all auctions (666 out of 724). Table 3.4 below contains summary statistics per type when type 1 bidders are the pair (D,P).

Table 3.4. Summary Statistics per Type

	Type 0		Type 1=(D,P)		Type 2	
	Number of observations	Mean SD	Number of observations	Mean SD	Number of observations	Mean SD
No. Bidders	666	4.81	325	4.65	312	5.17
		2.36		2.46		2.77
Winning bid	488	3.07	113	3.67	123	4.01
		2.93		3.08		3.65
Money on the table	488	0.28	113	0.29	123	0.36
		0.46		0.34		0.53
Engineers' Estimate	666	3.64	325	3.74	312	4.32
		3.38		3.27		3.72
All Bids	2520	3.69	369	3.66	458	4.41
		3.49		3.18		3.81
Backlog	2520	1.37	369	24.60	458	4.05
		3.40		16.44		6.00
Distance (miles)	2520	116.98	369	194.29	458	105.85
		168.91		98.51		157.12
Capacity (across firms)	398	1.67	2	39.12	13	15.73
		4.09		32.07		6.09
Utilization rate	2520	0.16	369	0.42	458	0.25
		0.32		0.26		0.32

All dollar figures are expressed in millions.

Firms in the triplet (A,D,E) tend to bid also in smaller size projects relative to type 2 bidders. At least one of the firms participate in 329 projects winning 117 times. The average winning bid for this group is $3.70 million which is below the average of the engineers' estimate. There are about five bidders participating in the projects where the triplet bids. The next table shows some summary statistics.

Table 3.5. Summary Statistics per Type

	Type 0		Type 1=(A,D,E)		Type 2	
	Number of observations	Mean SD	Number of observations	Mean SD	Number of observations	Mean SD
No. Bidders	666	4.81	329	4.66	306	5.08
		2.36		2.45		2.76
Winning bid	488	3.07	117	3.70	119	3.99
		2.93		3.12		3.63
Money on the table	488	0.28	117	0.30	119	0.36
		0.46		0.34		0.54
Engineers' Estimate	666	3.64	329	3.76	306	4.35
		3.38		3.34		3.77
All Bids	2520	3.69	415	3.85	412	4.30
		3.49		3.34		3.75
Backlog	2520	1.37	415	22.75	412	3.62
		3.40		16.64		5.39
Distance (miles)	2520	116.98	415	146.87	412	143.74
		168.91		100.69		172.66
Capacity (across firms)	398	1.67	3	31.72	12	15.63
		4.09		26.84		5.72
Utilization rate	2520	0.16	415	0.42	412	0.23
		0.32		0.28		0.30

All dollar figures are expressed in millions.

3.6 Empirical Results

In this section we present the empirical evidence obtained from the structural analysis. As explained above, there are two (alternative) groups of firms that have been selected as type 1 bidders, namely the triplet (A,D,E) and the pair (D,P). As a consequence the set of firms labeled as type 2 bidders is different for each case. Recall that both type 1 and type 2 bidders are large firms winning often. On the other hand, type 0 bidders (i.e. the small, fringe firms) include the same set of firms in each case (see Tables 3.4 and 3.5). Thus, there are two sets of results: one corresponding to the situation in which the firms (A,D,E) are type 1 bidders and the other to the case in which the firms (D,P) are type 1 bidders. We will refer to the former as the "triplet–case" and to the latter as the "pair–case".

There are both continuous and discrete variables involved in the estimation procedure. The set of continuous variables is given by $d_{i\ell}$, the log of the bid for the ith bidder in project ℓ and x_ℓ, the log of the engineers' estimate in project ℓ.[13] For the discrete variables we include $n_{0\ell}$, $n_{1\ell}$ and $n_{2\ell}$, namely the number of bidders of type 0, 1 and 2 in each project.

[13] All logarithms are base 10. Since the only continuous variable included as an explanatory variable is the engineers' estimate, $p = 1$ (see the definitions of the estimators in section 3.3).

As discussed above, the main purpose is to determine which model best describes the bidding behavior of the different types of bidders considered in this analysis. Moreover, both the competitive model and the collusive model differ only in the underlying distribution of private costs for type 1 bidders. It is then natural to attempt to find differences across the models by looking at these two distributions. Inspection of the expressions for the costs for each type of bidder in each model (see equations (3.4) and (3.5)) reveals that one should expect to find the greatest differences when: both n_0 and n_2 are small and n_1 is large. A number of combinations among n_0, n_1 and n_2 satisfy these conditions. We first discuss the results obtained from changing the number of bidders for types 0 and 2, respectively. The idea is then to assess the effect on type 1 distributions across models. We present results for two values of the (log) engineers' estimate, namely 6.1 and 6.5. The first corresponds to fairly small projects (around $1.3 million). The second value is roughly the log of the average value in the sample.

Figures 3.2 and 3.3 below contain the estimated densities of private values for type 0 bidders in the triplet–case and the pair–case, respectively. The distribution of private costs for type 0 bidders exhibits some variation with respect to n_0 for both the triplet–case and the pair–case. The variability observed could be reflecting the randomness in the data. However, at least three other explanations are possible. First, the exogeneity assumption could be inappropriate for these firms. Second, it could be that there are asymmetries across type 0 bidders, which are assumed away in the theoretical model. The case of endogenous entry would require one to explicitly include in the model a description of how firms decide whether or not to participate in an auction, which as discussed before is outside the scope of this chapter. Finally, it could be that not all cartel members are captured in the group of type 1 bidders. It is most likely that the results found for type 0 bidders are a combination of the premises outlined above. Nevertheless, type 0 bidders are fringe firms which hardly ever win a contract.

Figure 3.2. Type 0 densities for various values of n_0 - triplet–case

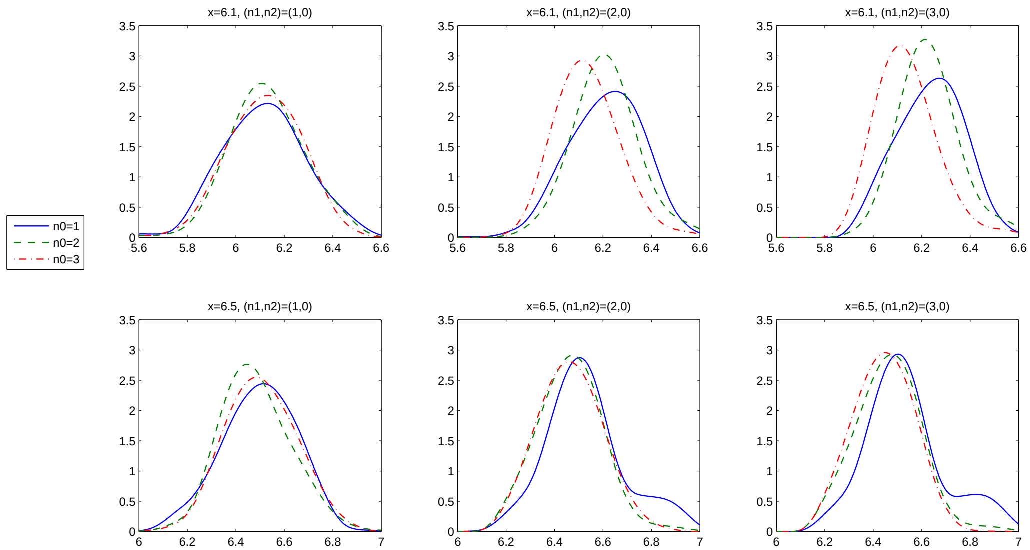
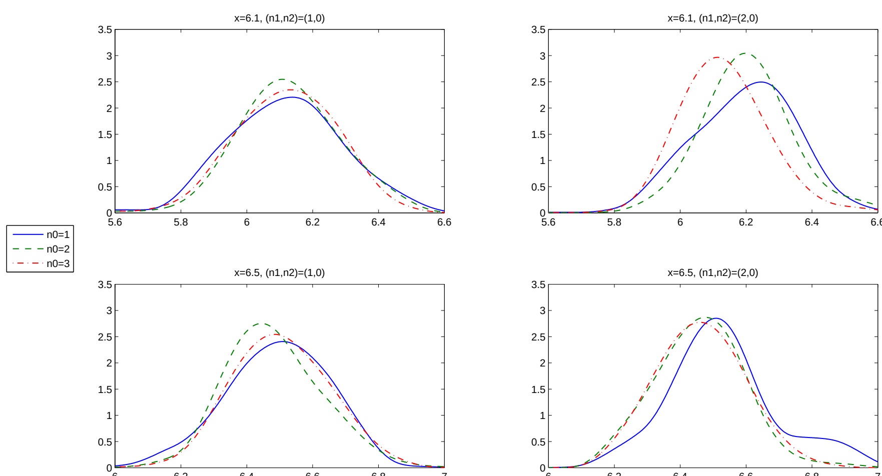


Figure 3.3. Type 0 densities for various values of n_0 - pair–case



Type 2 bidders' distributions do not show great variation for different values of n_2. In Figures 3.4 and 3.5 we present the results for the triplet–case and in Figure 3.6 for the pair–case. Unlike the case of type 0 bidders, these results are more in line with what one would expect if bidders are symmetric (within types) and the number of bidders is exogenous, as assumed in this chapter. Moreover, even when there are type 0 bidders participating (see Figure 3.5 and second row of Figure 3.6), the distributions for type 2 bidders are remarkably similar.

Figure 3.4. Type 2 densities for various values of n_2 and $n_0 = 0$ - triplet–case

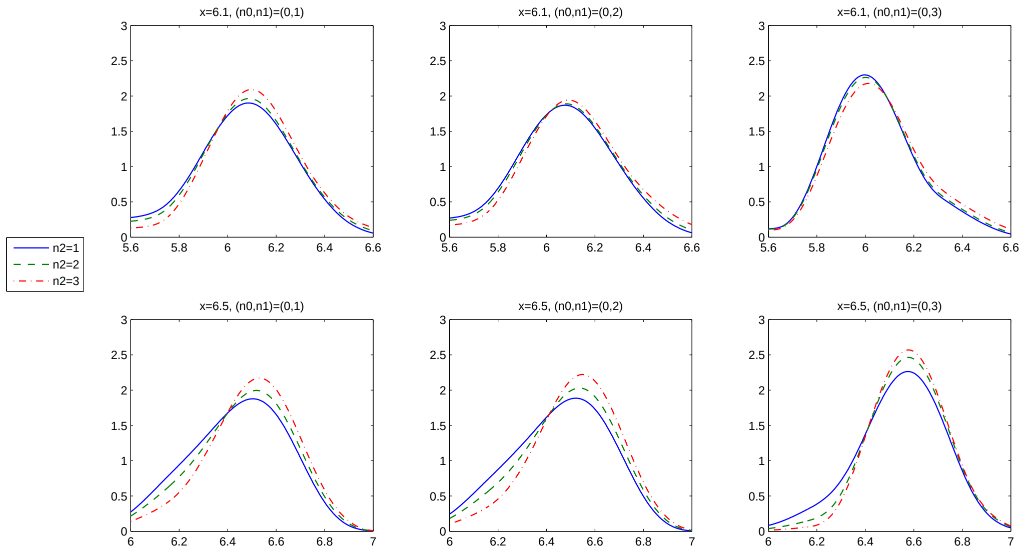


Figure 3.5. Type 2 densities for various values of n_2 and $n_0 > 0$ - triplet–case

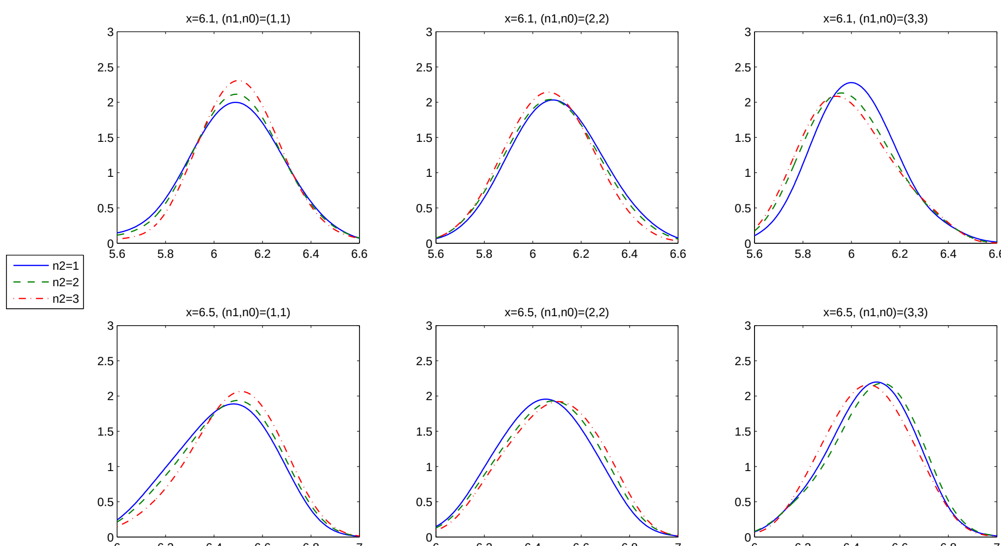
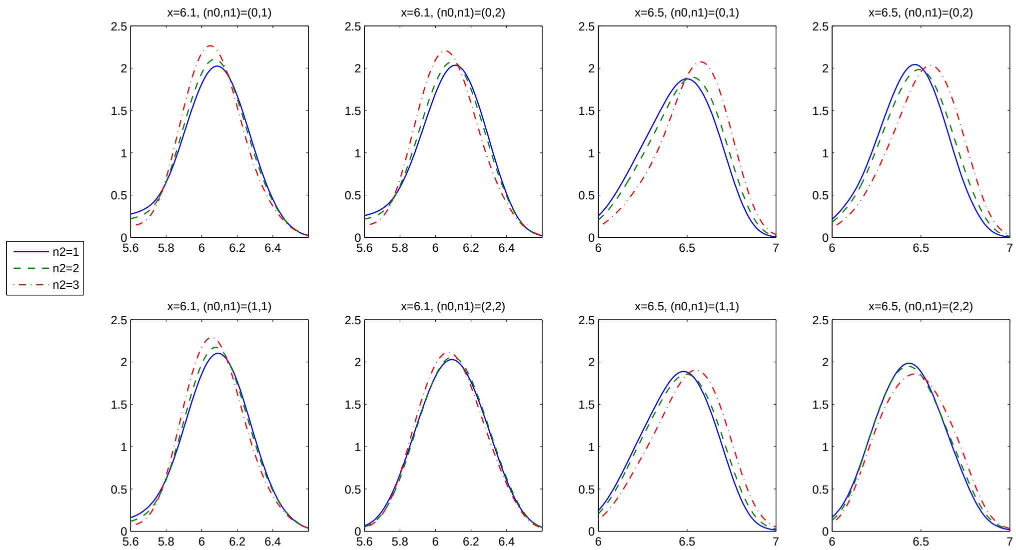


Figure 3.6. Type 2 densities for various values of n_2 - pair–case



In order to control for (to the greatest extent possible) other sources of variation, we decide to analyze how the distributions for type 1 bidders change as n_2 changes for various values of n_1. This is mainly driven by the above considerations regarding how the distributions of type 0 and type 2 bidders behave when the number of bidders changes.

For the triplet–case, Figure 3.7 shows the effect on the distributions of type 1 bidders in the competitive model (Model A, see the first row) and in the collusive model (Model B, see the second row) when $n_0 = 0$. The results for the case $n_0 > 0$ (not reported) are similar. The distribution of type 1 bidders shows less variation in the collusive setup. That is, under our exogeneity assumption and the assumption of symmetry within types, this piece of evidence suggests that firms (A,D,E) could be engaged in a collusive agreement. With respect to the pair–case (see Figures 3.8 and 3.9), the results are along the same lines as those for the triplet–case. The distributions of private costs in Model B exhibit less variation than in Model A, thus, providing additional evidence supporting the collusive setup. This is so for the cases in which $n_0 = 0$ and also when $n_0 = 1$. Recall that firm D is a type 1 bidder in both the triplet–case and the pair–case. Moreover, this firm participates in 44% of the projects in the sample. Thus, the similarity in the results for the triplet–case and the pair–case could be driven by the fact that firm D is a type 1 bidder in both cases.

Figure 3.7. Effect on type 1 distributions of changing n_2 - triplet–case

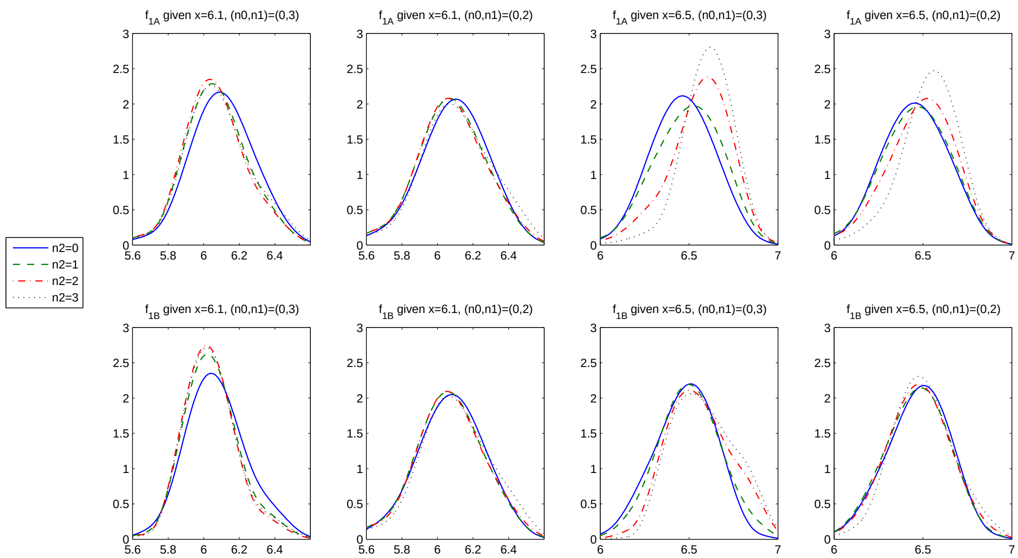
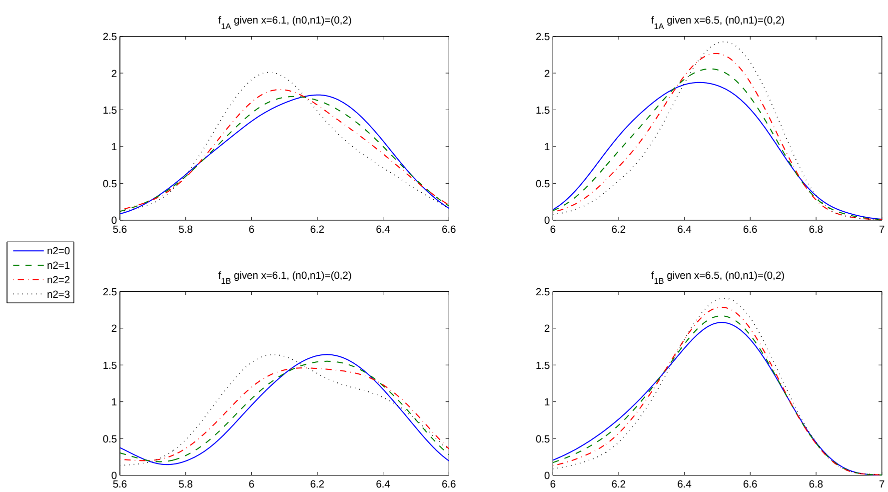
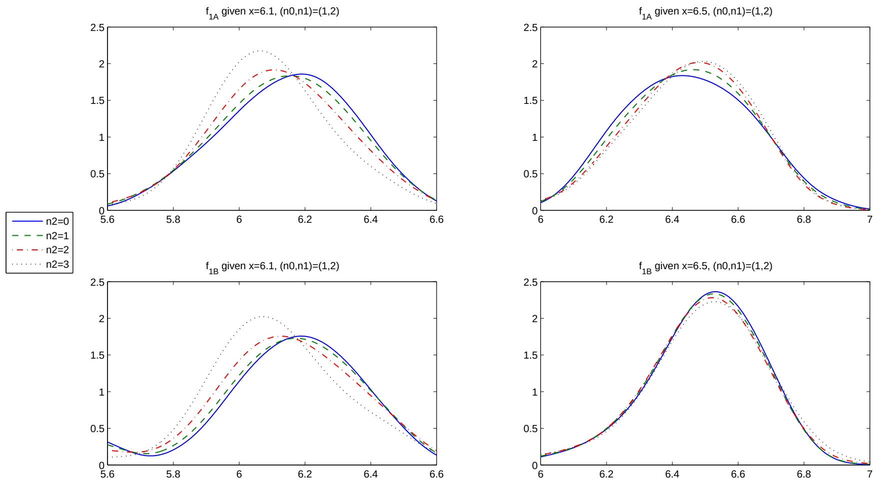


Figure 3.8. Effect on type 1 distributions of changing n_2 when $n_0 = 0$ - pair–case



Overall, the evidence in the sample tends to favor the collusive model over the competitive model.

Figure 3.9. Effect on type 1 distributions of changing n_2 when $n_0 = 1$ - pair–case



It is worth noting that we do not use a formal test, such as the Kolmogorov–Smirnov (KS) test, to distinguish between the two models. The main reason is that conducting this kind of test is not straightforward in our case. Recall that private costs c_{ij} are unobserved and we recover pseudo private costs, $\hat{c}_{ij}$, nonparametrically. Therefore a formal test should take into account the nuisance parameters introduced by the fact that $\hat{c}$ is used instead of c to estimate the distribution of private costs. This is not a trivial issue. Moreover, the distribution of private costs obtained are conditional on the engineers' estimate which is a continuous variable. This further complicates the formal comparison of the distributions of type 1 bidders across models. For the outlined reasons the KS test merits a separate paper.

3.7 Conclusions

This chapter proposes a methodology to detect cartels acting in procurement–auctions. Two competing models within the asymmetric IPV paradigm are used to investigate the behavior of firms competing for construction projects. In the first model (Model A) firms are engaged in a competitive game. On the other hand, in the second model type 1 bidders behave cooperatively. The method is applied to field data on highway construction projects in California. Relying on the assumptions of an exogenous number of bidders and symmetry among firms of the same type (but not across types) we find evidence suggesting collusive behavior during the sample period analyzed. We acknowledge that some of our assumptions are strong. However even under this restricted framework we find evidence supporting the operation of cartels.

Relatively few empirical papers analyze the presence of bid–rings in auction markets within the structural approach. This chapter contributes to this literature. Our exogeneity assumption is restrictive. For instance it precludes endogenous entry. However, a model in which entry is endogenous leads to a number of challenges. It is therefore desirable to develop a model with endogenous entry decisions affecting bidding decisions of firms to better understand the possibility of operating cartels. From a more applied perspective, the econometric model that follows from the theoretical model becomes also more involved under endogenous entry. We leave these issues for future research.

Appendix A

Proofs of Asymptotic Properties

This Appendix gives the proofs of our asymptotic results (propositions 1 and 2 in chapter 2). We state first two important results.

Results: Under A4 we have,

(i) $\sup_{(b,x,i)} |\hat{g}(b|x,i) - g_0(b|x,i)| = O_{as}\left(h_{1g}^R + h_{2g}^R + \sqrt{\frac{\log L}{Lh_{1g}h_{2g}}}\right)$

(ii) $\sup_{(b,x,i)} \left|\hat{G}(b|x,i) - G_0(b|x,i)\right| = O_{as}\left(h_G^{R+1} + \sqrt{\frac{\log L}{Lh_G}}\right)$

For a proof of the above results we refer the reader to Korostelev and Tsybakov (1993). We observe that the above results imply that $\sup_{p\ell}|\hat{V}_{p\ell} - V_{p\ell}| = o_{as}(1)$.

Proof of proposition 1: It suffices to show that $\sup_{\theta\in\Theta} \parallel S_L(\theta) - \hat{S}_L(\theta) \parallel = o_{as}(1)$. From the triangle inequality, A5-(iv) it follows that

$$
\begin{aligned}
\sup_{\theta\in\Theta}\|S_L(\theta) - \hat{S}_L(\theta)\| &= \sup_{\theta\in\Theta}\left\|\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}m(V_{p\ell}, Z_\ell;\theta) - \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}m(\hat{V}_{p\ell}, Z_\ell;\theta)\right\| \\
&= \sup_{\theta\in\Theta}\left\|\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}[m(V_{p\ell}, Z_\ell;\theta) - m(\hat{V}_{p\ell}, Z_\ell;\theta)]\right\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\sup_{\theta\in\Theta}\left\|m(V_{p\ell}, Z_\ell;\theta) - m(\hat{V}_{p\ell}, Z_\ell;\theta)\right\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_1(Z_\ell)|\hat{V}_{p\ell} - V_{p\ell}| \\
&= \{\mathrm{E}[K_1(Z)] + o_{as}(1)\}\sup_{p\ell}|\hat{V}_{p\ell} - V_{p\ell}| = o_{as}(1) \qquad \text{(A.1)}
\end{aligned}
$$

Where we use the fact that $\hat{V}_{p\ell}$ is a consistent estimator of $V_{p\ell}$, i.e. we make use of the 2 results stated at the beginning of this Appendix. Therefore, the desired result follows. *Q.E.D.*

Proof of proposition 2

From the FOCs that characterize $\tilde{\theta}$ and $\hat{\theta}$ respectively, we have

$$\frac{1}{2}\frac{\partial Q_L}{\partial \theta}(\tilde{\theta}) = \frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega S_L(\tilde{\theta}) = 0 \quad \text{(A.2)}$$

$$\frac{1}{2}\frac{\partial \hat{Q}_L}{\partial \theta}(\hat{\theta}) = \frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega \hat{S}_L(\hat{\theta}) = 0. \quad \text{(A.3)}$$

We can use a Taylor expansion around θ_0 to obtain

$$S_L(\tilde{\theta}) = S_L(\theta_0) + \frac{\partial S_L}{\partial \theta^T}(\overline{\theta})(\tilde{\theta} - \theta_0) \quad \text{(A.4)}$$

$$\hat{S}_L(\hat{\theta}) = \hat{S}_L(\theta_0) + \frac{\partial \hat{S}_L}{\partial \theta^T}(\overline{\theta}^*)(\hat{\theta} - \theta_0), \quad \text{(A.5)}$$

where $\overline{\theta}$ and $\overline{\theta}^*$ are vectors between $\tilde{\theta}$ and θ_0, and $\hat{\theta}$ and θ_0, respectively.

Thus using (A.4) in (A.2) we get

$$\begin{aligned} &\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega\left[S_L(\theta_0) + \frac{\partial S_L}{\partial \theta^T}(\overline{\theta})(\tilde{\theta} - \theta_0)\right] \\ = \;& \frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega S_L(\theta_0) + \frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega \frac{\partial S_L}{\partial \theta^T}(\overline{\theta})(\tilde{\theta} - \theta_0) = 0. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \sqrt{L}(\tilde{\theta} - \theta_0) &= -\left[\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega \frac{\partial S_L}{\partial \theta^T}(\overline{\theta})\right]^{-1} \frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega \sqrt{L} S_L(\theta_0) \\ &= -\tilde{A}^{-1}\tilde{B}\sqrt{L}S_L(\theta_0). \end{aligned}$$

Similarly using (A.5) in (A.3) yields

$$\begin{aligned} \sqrt{L}(\hat{\theta} - \theta_0) &= -\left[\frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega \frac{\partial \hat{S}_L}{\partial \theta^T}(\overline{\theta}^*)\right]^{-1} \frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega \sqrt{L} \hat{S}_L(\theta_0) \\ &= -\hat{A}^{-1}\hat{B}\sqrt{L}\hat{S}_L(\theta_0). \end{aligned}$$

we need to show: (i) $\tilde{B} - \hat{B} = o_{as}(1)$, (ii) $\tilde{A} - \hat{A} = o_{as}(1)$ since this and A6-(iii) imply that the difference of the inverses is also $o_{as}(1)$ and (iii) $\sqrt{L}[S_L(\theta_0) - \hat{S}_L(\theta_0)] = O_p(1)$.

The proof consists of three steps. Step 1: We prove $\tilde{B} - \hat{B} = o_{as}(1)$. The term $\tilde{B} - \hat{B}$ can be written as

$$
\begin{aligned}
\tilde{B} - \hat{B} &= \frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega - \frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega \\
&= \left(\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta}) - \frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta}) \right) \Omega \\
&= \left(\frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \Big(m_3^T(V_{p\ell}, Z_\ell, \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta}) \Big) \right) \Omega.
\end{aligned}
$$

It suffices to show that the norm of the term between brackets is $o_{as}(1)$ since Ω is a positive definite matrix. Namely

$$
\begin{aligned}
&\left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} m_3^T(V_{p\ell}, Z_\ell, \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta})) \right\| \\
&= \left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \Big[(m_3^T(V_{p\ell}, Z_\ell, \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta})) + m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta})) \Big] \right\| \\
&\leq \left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \Big[(m_3^T(V_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta})) \Big] \right\| \\
&\quad + \left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta})) \right\| \\
&= C + D, \qquad \text{(A.6)}
\end{aligned}
$$

where the last line follows from the triangle inequality. The term C is

$$
\begin{aligned}
C &= \left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \Big[(m_3^T(V_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta})) \Big] \right\| \\
&\leq \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \left\| \Big[m_3^T(V_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta}) \Big] \right\| \\
&\leq \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} K_3(Z_\ell) |V_{p\ell} - \hat{V}_{p\ell}| \\
&\leq \{ \mathrm{E}[K_3(Z)] + o_{as}(1) \} \sup_{p\ell} |\hat{V}_{p\ell} - V_{p\ell}| \\
&= o_{as}(1)
\end{aligned}
$$

where we use A6-(i) and the fact that $\hat{V}_{p\ell}$ is uniformly consistent, i.e., we use the two results stated at the beginning of this Appendix.

We consider now the term D

$$
D = \left\| \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{I_\ell} \sum_{p=1}^{I_\ell} \Big[m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta}) \Big] \right\|
$$

$$
\begin{aligned}
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\left\|m_3^T(\hat{V}_{p\ell}, Z_\ell; \tilde{\theta}) - m_3^T(\hat{V}_{p\ell}, Z_\ell; \hat{\theta})\right\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_4(Z_\ell)\|\tilde{\theta} - \hat{\theta}\| \\
&= \{\mathrm{E}[K_4(Z)] + o_{as}(1)\}o_{as}(1)
\end{aligned}
$$

where we have used A6-(ii) and the fact that $\tilde{\theta}$ and $\hat{\theta}$ are consistent estimators for θ_0.

STEP 2: We prove $\tilde{A} - \hat{A} = o_{as}(1)$. The term $\tilde{A} - \hat{A}$ is

$$
\begin{aligned}
\tilde{A} - \hat{A} &= \left(\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega\frac{\partial S_L}{\partial \theta^T}(\overline{\theta})\right) - \left(\frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega\frac{\partial \hat{S}_L}{\partial \theta^T}(\overline{\theta}^*)\right) \\
&= \left[\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega\left(\frac{\partial S_L}{\partial \theta^T}(\tilde{\theta}) + o_{as}(1)\right)\right] - \left[\frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega\left(\frac{\partial \hat{S}_L}{\partial \theta^T}(\hat{\theta}) + o_{as}(1)\right)\right] \\
&= \left(\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta})\Omega\frac{\partial S_L}{\partial \theta^T}(\tilde{\theta})\right) - \left(\frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\Omega\frac{\partial \hat{S}_L}{\partial \theta^T}(\hat{\theta})\right) + o_{as}(1) \\
&= \left[\left(\frac{\partial S_L^T}{\partial \theta}(\tilde{\theta}) - \frac{\partial \hat{S}_L^T}{\partial \theta}(\hat{\theta})\right)\Omega\right]\left(\frac{\partial S_L}{\partial \theta^T}(\tilde{\theta}) + \frac{\partial \hat{S}_L}{\partial \theta^T}(\hat{\theta})\right). \qquad (A.7)
\end{aligned}
$$

where the second equality comes from the following

$$
\begin{aligned}
\left\|\frac{\partial S_L}{\partial \theta^T}(\overline{\theta}) - \frac{\partial S_L}{\partial \theta^T}(\tilde{\theta})\right\| &\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(V_{p\ell}, Z_\ell; \overline{\theta}) - m_3(V_{p\ell}, Z_\ell; \tilde{\theta})\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_4(Z_\ell)\|\overline{\theta} - \tilde{\theta}\| \\
&= \{\mathrm{E}[K_4(Z)] + o_{as}(1)\}o_{as}(1) \\
&= o_{as}(1)
\end{aligned}
$$

where we use A6-(ii), the fact that $\tilde{\theta} \leq \overline{\theta} \leq \theta_0$ and that $\tilde{\theta} \xrightarrow{a.s} \theta_0$. A similar argument can be used to show that

$$
\frac{\partial \hat{S}_L}{\partial \theta^T}(\overline{\theta}^*) = \frac{\partial \hat{S}_L}{\partial \theta^T}(\hat{\theta}) + o_{as}(1),
$$

since $\hat{\theta} \leq \overline{\theta}^* \leq \theta_0$ and $\hat{\theta} \xrightarrow{a.s} \theta_0$.

Now, for the last line in (A.7) we observe that by Step 1, the first factor in (A.7) is $o_{as}(1)$ and the second factor can be expressed as follows

$$
\left\|\left(\frac{\partial S_L}{\partial \theta^T}(\tilde{\theta}) + \frac{\partial \hat{S}_L}{\partial \theta^T}(\hat{\theta})\right)\right\| = \left\|\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}[m_3(V_{p\ell}, Z_\ell; \tilde{\theta}) + m_3(\hat{V}_{p\ell}, Z_\ell; \hat{\theta})]\right\|
$$

$$
\begin{aligned}
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|[m_3(V_{p\ell},Z_\ell;\tilde{\theta})+m_3(\hat{V}_{p\ell},Z_\ell;\hat{\theta})]\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(V_{p\ell},Z_\ell;\tilde{\theta})\|+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(\hat{V}_{p\ell},Z_\ell;\hat{\theta})\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\sup_{\theta\in\Theta}\|m_3(V_{p\ell},Z_\ell;\theta)\|+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(\hat{V}_{p\ell},Z_\ell;\hat{\theta})-m_3(\hat{V}_{p\ell},Z_\ell;\theta_0) \\
&\quad +m_3(\hat{V}_{p\ell},Z_\ell;\theta_0)\| \\
&\leq \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_5(V_{p\ell},Z_\ell)+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(\hat{V}_{p\ell},Z_\ell;\hat{\theta})-m_3(\hat{V}_{p\ell},Z_\ell;\theta_0)\| \\
&\quad +\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(\hat{V}_{p\ell},Z_\ell;\theta_0)-m_3(V_{p\ell},Z_\ell;\theta_0)+m_3(V_{p\ell},Z_\ell;\theta_0)\| \\
&\leq \{\mathrm{E}[K_5(V,Z)]+o_{as}(1)\}+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_4(Z_\ell)\|\hat{\theta}-\theta_0\| \\
&\quad +\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(\hat{V}_{p\ell},Z_\ell;\theta_0)-m_3(V_{p\ell},Z_\ell;\theta_0)\|+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_3(V_{p\ell},Z_\ell;\theta_0)\| \\
&\leq \{\mathrm{E}[K_5(V,Z)]+o_{as}(1)\}+\{\mathrm{E}[K_4(Z)]+o_{as}(1)\}o_{as}(1)+\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_3(Z_\ell)|\hat{V}_{p\ell}-V_{p\ell}| \\
&\quad +\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\sup_{\theta\in\Theta}\|m_3(V_{p\ell},Z_\ell;\theta)\| \\
&\leq \{\mathrm{E}[K_5(V,Z)]+o_{as}(1)\}+\{\mathrm{E}[K_4(Z)]+o_{as}(1)\}o_{as}(1)+\{\mathrm{E}[K_3(Z)]+o_{as}(1)\}\sup_{p\ell}|\hat{V}_{p\ell}-V_{p\ell}| \\
&\quad +\{\mathrm{E}[K_5(V,Z)]+o_{as}(1)\} \\
&= 2\{\mathrm{E}[K_5(V,Z)]+o_{as}(1)\}<\infty
\end{aligned}
$$

where we use A6-(ii),(iv),(v) and the two results stated at the beginning of this Appendix.

Therefore the second factor in the last line of (A.7) converges to a finite limit and since the first factor is $o_{as}(1)$ the desired result follows.

Step 3: We prove $\sqrt{L}(S_L(\theta_0)-\hat{S}_L(\theta_0))=O_p(1)$.

The term $\sqrt{L}(S_L(\theta_0)-\hat{S}_L(\theta_0))$ is

$$
\begin{aligned}
B=\sqrt{L}(S_L(\theta_0)-\hat{S}_L(\theta_0)) &= \sqrt{L}\left(\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}m(V_{p\ell},Z_\ell;\theta_0)-\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}m(\hat{V}_{p\ell},Z_\ell;\theta_0)\right) \\
&= \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\Big[m(V_{p\ell},Z_\ell;\theta_0)-m(\hat{V}_{p\ell},Z_\ell;\theta_0)\Big]
\end{aligned}
$$

We prove

$$\begin{aligned} B &= \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\Big[m(V_{p\ell},Z_\ell;\theta_0)-m(\hat{V}_{p\ell},Z_\ell;\theta_0)\Big] \\ &= O_p(1)+o_{as}(1). \end{aligned}$$

The above expression can be rewritten as

$$\begin{aligned} B &= -\sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\Big[m_1(V_{p\ell},Z_\ell;\theta_0)(\hat{V}_{p\ell}-V_{p\ell})\Big] \\ &\quad +\sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\left[m_1(V_{p\ell},Z_\ell;\theta_0)-m_1(V^*_{p\ell},Z_\ell;\theta_0)\right](\hat{V}_{p\ell}-V_{p\ell}) \\ &= B_1+B_2, \end{aligned} \tag{A.8}$$

where the second equality comes from a Taylor expansion of order one and the following

$$\begin{aligned} & m(V_{p\ell},Z_\ell;\theta_0)-m(\hat{V}_{p\ell},Z_\ell;\theta_0) \\ &= m_1(V^*_{p\ell},Z_\ell;\theta_0)(V_{p\ell}-\hat{V}_{p\ell}) \\ &= m_1(V_{p\ell},Z_\ell;\theta_0)(\hat{V}_{p\ell}-V_{p\ell})+m_1(V^*_{p\ell},Z_\ell;\theta_0)(V_{p\ell}-\hat{V}_{p\ell})-m_1(V_{p\ell},Z_\ell;\theta_0)(\hat{V}_{p\ell}-V_{p\ell}) \\ &= -m_1(V_{p\ell},Z_\ell;\theta_0)(\hat{V}_{p\ell}-V_{p\ell})+[m_1(V_{p\ell},Z_\ell;\theta_0)-m_1(V^*_{p\ell},Z_\ell;\theta_0)](\hat{V}_{p\ell}-V_{p\ell}) \end{aligned}$$

Step 3.1: We consider B_1 in (A.8) and moreover we observe that for each i we can write

$$\begin{aligned} \|B_1\| &= \left\|\sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)(\hat{V}_{p\ell}-V_{p\ell})\right\| \\ &= \left\|\sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)\frac{1}{i-1}\left[\frac{\hat{G}(B_{p\ell}|X_\ell,i)}{\hat{g}(B_{p\ell}|X_\ell,i)}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\right]\right\| \\ &= \Bigg\|\sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)\Bigg\{\frac{\hat{G}(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0^2(B_{p\ell|X_\ell,i})} \\ &\quad \times\hat{g}(B_{p\ell}|X_\ell,i)+\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\frac{1}{\hat{g}(B_{p\ell}|X_\ell,i)g_0(B_{p\ell}|X_\ell,i)}\Big[\hat{g}(B_{p\ell}|X_\ell,i)-g_0(B_{p\ell}|X_\ell,i)\Big]^2 \\ &\quad -\frac{1}{\hat{g}(B_{p\ell}|X_\ell,i)g_0(B_{p\ell}|X_\ell,i)}\Big[\hat{G}(B_{p\ell}|X_\ell,i)-G_0(B_{p\ell}|X_\ell,i)\Big] \\ &\quad \times\left[\hat{g}(B_{p\ell}|X_\ell,i)-g_0(B_{p\ell}|X_\ell,i)\right]\Bigg\}\Bigg\| \\ &\leq \left\|\sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)\left[\frac{\hat{G}(B_{p\ell}|X_\ell,i)}{\hat{g}(B_{p\ell}|X_\ell,i)}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0^2(B_{p\ell|X_\ell,i})}\right.\right. \end{aligned}$$

$$
\begin{aligned}
&\times\hat{g}(B_{p\ell}|X_\ell,i)\Bigg]\Bigg\| + \Bigg\|\sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)\\
&\times\Bigg(\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\frac{1}{\hat{g}(B_{p\ell}|X_\ell,i)g_0(B_{p\ell}|X_\ell,i)}\left[\hat{g}(B_{p\ell}|X_\ell,i)-g_0(B_{p\ell}|X_\ell,i)\right]^2\\
&-\frac{1}{\hat{g}(B_{p\ell}|X_\ell,i)g_0(B_{p\ell}|X_\ell,i)}\left[\hat{G}(B_{p\ell}|X_\ell,i)-G_0(B_{p\ell}|X_\ell,i)\right]\\
&\times\left[\hat{g}(B_{p\ell}|X_\ell,i)-g_0(B_{p\ell}|X_\ell,i)\right]\Bigg)\Bigg\|\\
= \;&\|B_{11}\|+\|B_{12}\| \qquad \text{(A.9)}
\end{aligned}
$$

where the third line uses the following identity

$$
\frac{\tilde{a}}{\tilde{b}}-\frac{a}{b}=\frac{\tilde{a}-\frac{a}{b}\tilde{b}}{b}+\frac{a}{b}\frac{1}{\tilde{b}b}[\tilde{b}-b]^2-\frac{1}{\tilde{b}b}[\tilde{a}-a][\tilde{b}-b].
$$

The term B_{11} can be written as

$$
\begin{aligned}
B_{11} &= \sqrt{L}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}m_1(V_{p\ell},X_\ell,i;\theta_0)\left[\frac{\hat{G}(B_{p\ell}|X_\ell,i)-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\hat{g}(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\right]\\
&= \sqrt{L}(R_L+\frac{L(L-1)}{L^2}U_L)\\
&= B_{111}+B_{112}, \qquad \text{(A.10)}
\end{aligned}
$$

where

$$
\begin{aligned}
R_L &= \frac{1}{L^2}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(0)1\!\!1(B_{p\ell}\le B_{p\ell})\\
&-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(0)K_{2g,h_g}(0)\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
U_L &= \frac{1}{L(L-1)}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{\{j:I_j=i,j\neq\ell\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\sum_{q=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\\
&\times\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(X_j-X_\ell)1\!\!1(B_{qj}\le B_{p\ell})\\
&-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(X_j-X_\ell)K_{2g,h_g}(B_{qj}-B_{p\ell})\Bigg].
\end{aligned}
$$

To see how to obtain the last line in (A.10), we observe that the term within brackets in the first line of (A.10) can be expressed as

$$
\begin{aligned}
&\frac{\hat{G}(B_{p\ell}|X_\ell,i)-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\hat{g}(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} = \\
= \; & \frac{1}{g_0(B_{p\ell}|X_\ell,i)}\left[\hat{G}(B_{p\ell}|X_\ell,i)-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\hat{g}(B_{p\ell}|X_\ell,i)\right] \\
= \; & \frac{1}{g_0(B_{p\ell}|X_\ell,i)} \\
& \times\left[\frac{1}{Lh_G}\frac{L}{n_i}\sum_{\{j:I_j=i\}}^{L}\sum_{q=1}^{i}e_1^T\left(\frac{X_{i,R+1}^T W_x^G X_{i,R+1}}{n_i}\right)^{-1}X_{R+1,j}K_G\left(\frac{X_j-X_\ell}{h_G}\right)\mathbb{1}(B_{qj}\le B_{p\ell})\right. \\
& -\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} \\
& \left.\times\frac{1}{Lh_g^2}\frac{L}{n_i}\sum_{\{j:I_j=i\}}^{L}\sum_{q=1}^{i}e_1^T\left(\frac{X_{i,R}^T W_x^g X_{i,R}}{n_i}\right)^{-1}X_{R,j}K_{1g}\left(\frac{X_j-X_\ell}{h_g}\right)K_{2g}\left(\frac{B_{qj}-B_{p\ell}}{h_g}\right)\right] \\
= \; & \frac{1}{g_0(B_{p\ell}|X_\ell,i)}\left[\frac{1}{L}\frac{L}{n_i}\sum_{\{j:I_j=i\}}^{L}\sum_{q=1}^{i}\omega_{i,R+1,j}^G K_{G,h_G}(X_j-X_\ell)\mathbb{1}(B_{qj}\le B_{p\ell})-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\right. \\
& \left.\times\frac{1}{L}\frac{L}{n_i}\sum_{\{j:I_j=i\}}^{L}\sum_{q=1}^{i}\omega_{i,R,j}^g K_{1g,h_g}(X_j-X_\ell)K_{2g,h_g}(B_{qj}-B_{p\ell})\right] \qquad \text{(A.11)}
\end{aligned}
$$

where we have used the following notations

$$K_{G,h_G}(X_j-X_\ell)=\frac{1}{h_G}K_G\left(\frac{X_j-X_\ell}{h_G}\right), \qquad \text{(A.12)}$$

$$K_{1g,h_g}(X_j-X_\ell)=\frac{1}{h_g}K_{1g}\left(\frac{X_j-X_\ell}{h_g}\right), \qquad \text{(A.13)}$$

$$K_{2g,h_g}(B_{qj}-B_{p\ell})=\frac{1}{h_g}K_{2g}\left(\frac{B_{qj}-B_{p\ell}}{h_g}\right), \qquad \text{(A.14)}$$

$$\omega_{i,R+1,j}^G=e_1^T\left(\frac{X_{i,R+1}^T W_x^G X_{i,R+1}}{n_i}\right)^{-1}X_{R+1,j}, \qquad \text{(A.15)}$$

$$\omega_{i,R,j}^g=e_1^T\left(\frac{X_{i,R}^T W_x^g X_{i,R}}{n_i}\right)^{-1}X_{R,j}, \qquad \text{(A.16)}$$

Now using (A.11) in the first line of (A.10), we get

$$\begin{aligned} B_{11} &= \sqrt{L}\Bigg(\frac{1}{L^2}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{\{j:I_j=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\sum_{q=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)} \\ &\quad\times\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(X_j-X_\ell)1\!\!1(B_{qj}\le B_{p\ell}) \\ &\quad-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(X_j-X_\ell)K_{2g,h_g}(B_{qj}-B_{p\ell})\Bigg]\Bigg). \qquad (A.17) \end{aligned}$$

The term between parenthesis in (A.17) can be decomposed as follows. Namely,
1) Diagonal terms ($\ell = j, p = q$)

$$\begin{aligned} R_L &= \frac{1}{L^2}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(0)1\!\!1(B_{p\ell}\le B_{p\ell}) \\ &\quad-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(0)K_{2g,h_g}(0)\Bigg], \qquad (A.18) \end{aligned}$$

2) Off-diagonal terms ($\ell \neq j$)

$$\begin{aligned} \frac{L(L-1)}{L^2}U_L &= \frac{1}{L^2}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\sum_{\{j:I_j=i,j\neq\ell\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\sum_{q=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)} \\ &\quad\times\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(X_j-X_\ell)1\!\!1(B_{qj}\le B_{p\ell}) \\ &\quad-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(X_j-X_\ell)K_{2g,h_g}(B_{qj}-B_{p\ell})\Bigg]. \qquad (A.19) \end{aligned}$$

From (A.18) and (A.19) we have the expression in the last line of (A.10). It remains to show that $B_{11} = B_{111} + B_{112} = o_{as}(1)$. We consider first $B_{111} = \sqrt{L}R_L$ in (A.10). Specifically,

$$\begin{aligned} B_{111} &= \sqrt{L}\|R_L\| = \sqrt{L}\Bigg\|\Bigg(\frac{1}{L^2}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\Bigg[\omega^G_{i,R+1,j}K_{G,h_G}(0) \\ &\quad-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g,h_g}(0)K_{2g,h_g}(0)\Bigg]\Bigg)\Bigg\| \\ &= \sqrt{L}\Bigg\|\frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\Bigg[\omega^G_{i,R+1,j}\frac{K_{G,h_G}(0)}{L}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} \\ &\quad\times\omega^g_{i,R,j}\frac{K_{1g,h_g}(0)K_{2g,h_g}(0)}{L}\Bigg]\Bigg\| \end{aligned}$$

$$
\begin{aligned}
&\leq \left(\frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\|m_1(V_{p\ell},X_\ell,i;\theta_0)\|^2\right)^{\frac{1}{2}} \\
&\quad\times\sqrt{L}\Bigg(\frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{1}{g_0(B_{p\ell}|X_\ell,i)^2}\Bigg[\omega^G_{i,R+1,j}\frac{K_{G,h_G}(0)}{L}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} \\
&\quad\times\omega^g_{i,R,j}\frac{K_{1g,h_g}(0)K_{2g,h_g}(0)}{L}\Bigg]^2\Bigg)^{\frac{1}{2}} \\
&= CD,
\end{aligned}
$$

where the inequality comes from Cauchy-Schwartz. First we show that $C^2 < \infty$. Using A6-(vi), $0 < (1/(i-1)) < 1$ for each $i \in \mathcal{I}$ and $L/n_i = L/(iL_i) < \infty$ we get

$$
\begin{aligned}
C^2 &= \frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\|m_1(V_{p\ell},X_\ell,i;\theta_0)\|^2 \\
&\leq \frac{L}{n_i}\frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\frac{1}{i(i-1)}\sup_{\theta\in\Theta}\|m_1(V_{p\ell},X_\ell,i;\theta)\|^2 \\
&< \frac{1}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}K_7(V_{p\ell},X_\ell,i)^2 \\
&= \mathrm{E}[K_7(V,X,i)^2] + o_{as}(1) < \infty.
\end{aligned}
$$

It remains to consider the D term above. Namely,

$$
\begin{aligned}
D &\leq \sqrt{L}\Bigg(\frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{1}{g_0(B_{p\ell}|X_\ell,i)^2}\Bigg[\omega^G_{i,R+1,j}\frac{K_G(0)}{Lh_G}-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} \\
&\quad\times\omega^g_{i,R,j}\frac{K_{1g,h_g}(0)K_{2g,h_g}(0)}{Lh_g^2}\Bigg]^2\Bigg)^{\frac{1}{2}} \\
&= \sqrt{L}\Bigg(\frac{1}{L}\frac{L}{n_i}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i(i-1)}\sum_{p=1}^{i}\frac{1}{g_0(B_{p\ell}|X_\ell,i)^2}\Bigg[O_p(1)O_p\left(\frac{1}{Lh_G}\right)-\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)} \\
&\quad\times O_p(1)O_p\left(\frac{1}{Lh_g^2}\right)\Bigg]^2\Bigg)^{\frac{1}{2}} \\
&< \sqrt{L}\kappa_1\left[O_p\left(\frac{1}{Lh_G}\right)-\kappa_2 O_p\left(\frac{1}{Lh_g^2}\right)\right] \\
&= \kappa_1\left[O_p\left(\frac{1}{\sqrt{L}h_G}\right)-\kappa_2 O_p\left(\frac{1}{\sqrt{L}h_g^2}\right)\right] \\
&= \kappa_1[o_p(1)-\kappa_2 o_p(1)] \\
&= o_p(1),
\end{aligned}
$$

where after the first equality we use (A.12)- (A.16). The second line follows from observing that

$$\begin{aligned}
\omega^G_{i,R+1,j} &= e_1^T \left(\frac{X^T_{i,R+1} W^G_x X_{i,R+1}}{n_i} \right)^{-1} X_{R+1,j} \\
&= e_1^T \left[\frac{1}{n_i h_G} \sum_{s=1}^{n_i} \mathbf{x}_s^T \mathbf{x}_s K_G \left(\frac{x_s - x_j}{h_g} \right) \right]^{-1} e_1 \\
&= O_p(1)
\end{aligned}$$

and similarly for $\omega^g_{i,R,j}$.

The third line uses the fact that densities are bounded away from zero and $0 < (1/i(i-1)) < 1$ for all i. The last line follows from Assumption A4.AN. Thus, $B_{111} = CD = o(1)$ as desired.

Let $Y_{p\ell} = (B_{p\ell}, X_\ell)$ and for each i define ,

$r_L(Y_{p\ell}, i) = E[p_L((Y_{p\ell}, i), (Y_{qj}, i))|(Y_{p\ell}, i)]$, where $p_L(\cdot, \cdot)$ is a symmetric function,

$\theta_L = E[r_L(Y_{p\ell}, i)] = E[p_L((Y_{p\ell}, i), (Y_{qj}, i))]$,

$\hat{U}_L = \theta_L + \frac{2}{L} \sum_{\{\ell:I_\ell=i\}}^{L} \frac{1}{i} \sum_{p=1}^{i} [r_L(Y_{p\ell}, i) - \theta_L]$.

Next, we consider B_{112} in (A.10)

$$\begin{aligned}
B_{112} &= \frac{L(L-1)}{L^2} \sqrt{L} U_L \\
&= \frac{L(L-1)}{L^2} \sqrt{L} (U_L - \hat{U}_L) + \frac{L(L-1)}{L^2} \sqrt{L} \hat{U}_L \\
&= B_{1121} + B_{1122},
\end{aligned} \tag{A.20}$$

where U_L can be written as a U-statistic. Namely,

$$\begin{aligned}
U_L &= \frac{1}{L(L-1)} \frac{L}{n_i} \sum_{\{\ell:I_\ell=i\}}^{L} \sum_{\{j:I_j=i, j\neq\ell\}}^{L} \frac{1}{i(i-1)} \sum_{p=1}^{i} \sum_{q=1}^{i} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \\
&\quad \times \Bigg[\omega^G_{i,R+1,j} K_{G,h_G}(X_j - X_\ell) 1\!\!1(B_{qj} \leq B_{p\ell}) \\
&\quad - \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \omega^g_{i,R,j} K_{1g,h_g}(X_j - X_\ell) K_{2g,h_g}(B_{qj} - B_{p\ell}) \Bigg] \\
&= \frac{1}{L(L-1)} \sum_{\{\ell:I_\ell=i\}}^{L-1} \sum_{\{j:I_j=i, j=\ell+1\}}^{L} \frac{1}{i} \sum_{p=1}^{i} \sum_{q=1}^{i} \Bigg\{ \frac{L}{n_i(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \\
&\quad \times \Bigg[\omega^G_{i,R+1,j} K_{G,h_G}(X_j - X_\ell) 1\!\!1(B_{qj} \leq B_{p\ell}) \\
&\quad - \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \omega^g_{i,R,j} K_{1g,h_g}(X_j - X_\ell) K_{2g,h_g}(B_{qj} - B_{p\ell}) \Bigg] \Bigg\}
\end{aligned}$$

$$= \frac{2}{L(L-1)} \sum_{\{\ell: I_\ell = i\}}^{L-1} \sum_{\{j: I_j = i, j = \ell+1\}}^{L} \frac{1}{i} \sum_{p=1}^{i} \sum_{q=1}^{i} \left[\frac{m_1(V_{p\ell}, X_\ell, i; \theta_0) K^{**}(B_{p\ell}, B_{qj}, X_\ell, X_j, i)}{2} \right.$$

$$\left. + \frac{m_1(V_{qj}, X_j, i; \theta_0) K^{**}(B_{qj}, B_{p\ell}, X_j, X_\ell, i)}{2} \right]$$

$$= \binom{L}{2}^{-1} \sum_{\{\ell: I_\ell = i\}}^{L-1} \sum_{\{j: I_j = i, j = \ell+1\}}^{L} \frac{1}{i} \sum_{p=1}^{i} \sum_{q=1}^{i} p_L\left((B_{p\ell}, X_\ell, i), (B_{qj}, X_j, i)\right).$$

We prove $B_{1121} = \sqrt{L}(U_L - \hat{U}_L) = o_p(1)$. By Lemma 3.1 in Powell, Stock, and Stoker (1989) it suffices to show that $\mathrm{E}[\|p_L((Y_{p\ell}, i), (Y_{qj}, i))\|^2] = o(L)$. We will show that $\mathrm{E}[\|p_L((Y_{p\ell}, i), (Y_{qj}, i))\|^2 | i] = o(L)$ since it implies the aforementioned condition.

$$\mathrm{E}[\|p_L((Y_{p\ell}, i), (Y_{qj}, i))\|^2 | i] = \int \|p_L((Y_{p\ell}, i), (Y_{qj}, i))\|^2 g_0(Y_{p\ell}|i) g_0(Y_{qj}|i) dY_{p\ell} dY_{qj}$$

$$= \frac{1}{4} \int \left\| \frac{L}{n_i(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \left[\frac{1}{h_G} \omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!1(B_{qj} \le B_{p\ell}) \right.\right.$$

$$\left. - \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \frac{1}{h_g^2} \omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) \right]$$

$$+ \frac{L}{n_i(i-1)} \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \left[\frac{1}{h_G} \omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!1(B_{p\ell} \le B_{qj}) \right.$$

$$\left.\left. - \frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)} \frac{1}{h_g^2} \omega^g_{i,R,j} K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right) \right] \right\|^2$$

$$\times g_0(Y_{p\ell}|i) g_0(Y_{qj}|i) dY_{p\ell} dY_{qj}$$

$$= \frac{1}{4} \int \left\| \left[\frac{L}{n_i(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{1}{h_G} \omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!1(B_{qj} \le B_{p\ell}) \right.\right.$$

$$\left. + \frac{L}{n_i(i-1)} \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \frac{1}{h_G} \omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!1(B_{p\ell} \le B_{qj}) \right]$$

$$+ \left[\frac{L}{n_i(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \frac{1}{h_g^2} \omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) \right.$$

$$\times K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) + \frac{L}{n_i(i-1)} \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)} \frac{1}{h_g^2} \omega^g_{i,R,j} K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right)$$

$$\left.\left. \times K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right) \right] \right\|^2 g_0(Y_{p\ell}|i) g_0(Y_{qj}|i) dY_{p\ell} dY_{qj}$$

$$= \frac{1}{4} \int \left\| \frac{1}{h_G} \frac{L}{n_i} \left[\frac{1}{(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!1(B_{qj} \le B_{p\ell}) \right.\right.$$

$$\left. + \frac{1}{(i-1)} \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!1(B_{p\ell} \le B_{qj}) \right]$$

$$+\frac{1}{h_g^2}\frac{L}{n_i}\left[\frac{1}{(i-1)}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_j-X_\ell}{h_g}\right)K_{2g}\left(\frac{B_{qj}-B_{p\ell}}{h_g}\right)\right.$$

$$\left.\left.+\frac{1}{i-1}\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\frac{G_0(B_{qj}|X_j,i)}{g_0(B_{qj}|X_j,i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_\ell-X_j}{h_g}\right)K_{2g}\left(\frac{B_{p\ell}-B_{qj}}{h_g}\right)\right]\right\|^2$$

$$\times g_0(Y_{p\ell}|i)g_0(Y_{qj}|i)dY_{p\ell}dY_{qj}$$

$$\leq \frac{1}{2}\int\left\{\left\|\frac{L}{n_i}\frac{1}{h_G}\left[\frac{1}{(i-1)}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_j-X_\ell}{h_G}\right)1\!\!1(B_{qj}\leq B_{p\ell})\right.\right.\right.$$

$$\left.\left.+\frac{1}{(i-1)}\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_\ell-X_j}{h_G}\right)1\!\!1(B_{p\ell}\leq B_{qj})\right]\right\|^2$$

$$+\left\|\frac{L}{n_i}\frac{1}{h_g^2}\left[\frac{1}{(i-1)}\frac{m_1(V_{p\ell},X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\frac{G_0(B_{p\ell}|X_\ell,i)}{g_0(B_{p\ell}|X_\ell,i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_j-X_\ell}{h_g}\right)K_{2g}\left(\frac{B_{qj}-B_{p\ell}}{h_g}\right)\right.\right.$$

$$\left.\left.\left.+\frac{1}{(i-1)}\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\frac{G_0(B_{qj}|X_j,i)}{g_0(B_{qj}|X_j,i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_\ell-X_j}{h_g}\right)K_{2g}\left(\frac{B_{p\ell}-B_{qj}}{h_g}\right)\right]\right\|^2\right\}$$

$$\times g_0(Y_{p\ell}|i)g_0(Y_{qj}|i)dY_{p\ell}dY_{qj}$$

$$= C+D \tag{A.21}$$

where the inequality comes from using $(a+b)^2 \leq 2(a^2+b^2)$. Therefore we need to show that both C and D are $o(L)$.

We consider first the C term in (A.21), and note that we can write $V_{p\ell} = \xi(B_{p\ell}, X_\ell, i)$. It gives

$$C = \frac{1}{2h_G^2}\int\left(\frac{L}{n_i}\frac{1}{(i-1)}\right)^2\left\|\frac{m_1(\xi(B_{p\ell},X_\ell,i),X_\ell,i;\theta_0)}{g_0(B_{p\ell}|X_\ell,i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_j-X_\ell}{h_G}\right)1\!\!1(B_{qj}\leq B_{p\ell})\right.$$

$$\left.+\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_\ell-X_j}{h_G}\right)1\!\!1(B_{p\ell}\leq B_{qj})\right\|^2 g_0(Y_{p\ell}|i)g_0(Y_{qj}|i)dY_{p\ell}dY_{qj}$$

$$= \frac{1}{2}\int\left(\frac{L}{n_i}\frac{1}{(i-1)}\right)^2\left\|\frac{m_1(\xi(uh_G+Y_{qj},i),u_2h_G+X_j,i;\theta_0)}{g_0(u_1h_G+B_{qj}|u_2h_G+X_j,i)}\omega^G_{i,R+1,j}K_G(-u_2)\right.$$

$$\left.\times 1\!\!1(B_{qj}\leq u_1h_G+B_{qj})+\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\omega^G_{i,R+1,j}K_G(u_2)1\!\!1(u_1h_G+B_{qj}\leq B_{qj})\right\|^2$$

$$\times g_0(uh_G+Y_{qj}|i)g_0(Y_{qj}|i)dudY_{qj}$$

$$\leq \int\left[\left\|\frac{m_1(\xi(uh_G+Y_{qj},i),u_2h_G+X_j,i;\theta_0)}{g_0(u_1h_G+B_{qj}|u_2h_G+X_j,i)}\omega^G_{i,R+1,j}K_G(-u_2)1\!\!1(B_{qj}\leq u_1h_G+B_{qj})\right\|^2\right.$$

$$\left.+\left\|\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\omega^G_{i,R+1,j}K_G(u_2)1\!\!1(u_1h_G+B_{qj}\leq B_{qj})\right\|^2\right]$$

$$\times g_0(uh_G+Y_{qj}|i)g_0(Y_{qj}|i)dudY_{qj}$$

$$= C_1+C_2,$$

where we have used the change of variable

$$u = \frac{Y_{p\ell} - Y_{qj}}{h_G} = \left(\frac{B_{p\ell} - B_{qj}}{h_G}, \frac{X_\ell - X_j}{h_G}\right) = (u_1, u_2),$$

and the inequality comes from using $(a+b)^2 \leq 2(a^2+b^2)$ and $\left(\frac{L}{n_i}\frac{1}{(i-1)}\right)^2 < \infty$.

Next we consider C_1

$$\begin{aligned}
C_1 &= \int \left\| \frac{m_1(\xi(uh_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0)}{g_0(u_1 h_G + B_{qj} | u_2 h_G + X_j, i)} \omega^G_{i,R+1,j} K_G(-u_2) 1\!\!1(B_{qj} \leq u_1 h_G + B_{qj}) \right\|^2 \\
&\quad \times g_0(uh_G + Y_{qj}|i) g_0(Y_{qj}|i) du dY_{qj} \\
&= \int \left\| \frac{m_1(\xi(uh_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0)}{g_0(u_1 h_G + B_{qj} | u_2 h_G + X_j, i)} \right\|^2 \left\| \omega^G_{i,R+1,j} K_G(-u_2) 1\!\!1(B_{qj} \leq u_1 h_G + B_{qj}) \right\|^2 \\
&\quad \times g_0(uh_G + Y_{qj}|i) g_0(Y_{qj}|i) du dY_{qj} \\
&= \int \| m_1(\xi(uh_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0) \|^2 \left\| \omega^G_{i,R+1,j} K_G(-u_2) 1\!\!1(B_{qj} \leq u_1 h_G + B_{qj}) \right\|^2 \\
&\quad \times \frac{g_0(uh_G + Y_{qj}|i) g(Y_{qj}|i)}{g_0(u_1 h_G + B_{qj} | u_2 h_G + X_j, i)^2} du dY_{qj} \\
&\leq \int \| m_1(\xi(uh_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0) \|^2 \left\| \omega^G_{i,R+1,j} K_G(-u_2) 1\!\!1(B_{qj} \leq u_1 h_G + B_{qj}) \right\|^2 \\
&\quad \times g_0(Y_{qj}|i) du dY_{qj},
\end{aligned}$$

where the last inequality comes from the assumption that densities are bounded.

By the Lebesgue Dominated Convergence (LDC) Theorem and A6-(vi), the above integral converges to

$$\int \| O_p(1) K_G(-u_2) \|^2 \, du \int \| m_1(V_{qj}, X_j, i; \theta_0) \|^2 g_0(Y_{qj}|i) dY_{qj} < \infty.$$

Hence, $C_1 = o(L)$ as $L \to \infty$

A similar argument can be used to show that $C_2 = o(L)$ as $L \to \infty$. Therefore, $C = C_1 + C_2 = o(L)$.

Next we consider the D term in (A.21). Namely

$$\begin{aligned}
D &= \frac{1}{2} \int \Bigg\| \frac{L}{n_i} \frac{1}{h_g^2} \Bigg[\frac{1}{(i-1)} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) \\
&\quad \times K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) + \frac{1}{(i-1)} \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)} \omega^g_{i,R,j} \\
&\quad \times K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right) \Bigg] \Bigg\|^2 g_0(Y_{p\ell}|i) g_0(Y_{qj}|i) dY_{p\ell} dY_{qj} \\
&= \frac{1}{2h_g^4} \left(\frac{L}{n_i}\frac{1}{(i-1)}\right)^2 \int \Bigg\| \frac{m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \omega^g_{i,R,j} \\
&\quad \times K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\frac{G_0(B_{qj}|X_j,i)}{g_0(B_{qj}|X_j,i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_\ell-X_j}{h_g}\right)K_{2g}\left(\frac{B_{p\ell}-B_{qj}}{h_g}\right)\Bigg\|^2\\
&\times g_0(Y_{p\ell}|i)g_0(Y_{qj}|i)dY_{p\ell}dY_{qj}\\
\leq\ & \frac{1}{2h_g^2}\int\Bigg\|\frac{m_1(\xi(uh_G+Y_{qj},i),u_2h_G+X_j,i;\theta_0)}{g_0(u_1h_G+B_{qj}|u_2h_G+X_j,i)}\frac{G_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)}{g_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)}\\
&\times\omega^g_{i,R,j}K_{1g}(-u_2)K_{2g}(-u_1)\\
&+\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\frac{G_0(B_{qj}|X_j,i)}{g_0(B_{qj}|X_j,i)}\omega^g_{i,R,j}K_{1g}(u_2)K_{2g}(u_1)\Bigg\|^2 g_0(uh_g+Y_{qj}|i)g_0(Y_{qj}|i)dudY_{qj}\\
\leq\ & \frac{1}{h_g^2}\int\Bigg[\Bigg\|\frac{m_1(\xi(uh_G+Y_{qj},i),u_2h_G+X_j,i;\theta_0)}{g_0(u_1h_G+B_{qj}|u_2h_G+X_j,i)}\frac{G_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)}{g_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)}\omega^g_{i,R,j}\\
&\times K_{1g}(-u_2)K_{2g}(-u_1)\Bigg\|^2+\Bigg\|\frac{m_1(V_{qj},X_j,i;\theta_0)}{g_0(B_{qj}|X_j,i)}\frac{G_0(B_{qj}|X_j,i)}{g_0(B_{qj}|X_j,i)}\omega^g_{i,R,j}K_{1g}(u_2)K_{2g}(u_1)\Bigg\|^2\Bigg]\\
&\times g_0(uh_g+Y_{qj}|i)g_0(Y_{qj}|i)dudY_{qj}\\
=\ & D_1+D_2,
\end{aligned}
$$

where we have used the change of variable
$$u=\frac{Y_{p\ell}-Y_{qj}}{h_g}=\left(\frac{B_{p\ell}-B_{qj}}{h_g},\frac{X_\ell-X_j}{h_g}\right)=(u_1,u_2),$$
the first inequality comes from the fact that $\left(\frac{L}{n_i}\frac{1}{(i-1)}\right)^2<\infty$, and the second inequality comes from using $(a+b)^2\leq 2(a^2+b^2)$.

We consider first D_1. Specifically,

$$
\begin{aligned}
D_1 &= \frac{1}{h_g^2}\int\left\|m_1(\xi(uh_g+Y_{qj},i),u_2h_g+X_j,i;\theta_0)G_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)\right\|^2\\
&\times\left\|\omega^g_{i,R,j}K_{1g}(-u_2)K_{2g}(-u_1)\right\|^2\frac{g_0(uh_g+Y_{qj}|i)g_0(Y_{qj}|i)}{g_0(u_1h_g+B_{qj}|u_2h_g+X_j,i)^4}dudY_{qj}\\
&\leq \frac{1}{h_g^2}\int\|m_1(\xi(uh_g+Y_{qj},i),u_2h_g+X_j,i;\theta_0)\|^2\left\|\omega^g_{i,R,j}K_{1g}(-u_2)K_{2g}(-u_1)\right\|^2\\
&\times g_0(Y_{qj}|i)dudY_{qj}
\end{aligned}
$$

where the inequality uses the fact that $G(\cdot|\cdot,i)$ is bounded and that densities are bounded from above.

By the LDC Theorem and A6-(vi) the above integral converges to

$$\int\|O_p(1)K_{1g}(-u_2)K_{2g}(-u_1)\|^2du\int\|m_1(V_{qj},X_j,i;\theta_0)\|^2g_0(Y_{qj}|i)dY_{qj}<\infty.$$

Hence, $D_1=o(L)$ if and only if $Lh_g^2\to\infty$ as implied by A4.AN-(ii) since

$$Lh_g^2 = \sqrt{L}\sqrt{L}h_g^2\to\infty.$$

A similar argument can be used to show that $D_2 = o(L)$. That is, $D_2 = o(L)$ if and only if $Lh_g^2 \to \infty$, as implied by A4.AN-(ii).

Therefore, $C + D = C_1 + C_2 + D_1 + D_2 = o(L)$ and the desired result follows, i.e by Lemma 3.1 in Powell, Stock, and Stoker (1989) $\sqrt{L}(U_L - \hat{U}_L) = o_p(1)$.

Next we consider the second term in (A.20)

$$\begin{aligned}
B_{1122} &= \frac{L(L-1)}{L^2}\sqrt{L}\hat{U}_L \\
&= \frac{L(L-1)}{L^2}\sqrt{L}\left\{\theta_L + \frac{2}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}[r_L(Y_{p\ell}, i) - \theta_L]\right\} \\
&= \frac{L(L-1)}{L^2}\sqrt{L}E[p_L((Y_{p\ell}, i), (Y_{qj}, i))] + \frac{L(L-1)}{L^2}\sqrt{L}\frac{2}{L}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}[r_L(Y_{p\ell}, i) - \theta_L].
\end{aligned}$$

By the Central Limit Theorem (CLT), the second term above is $O_p(1)$ (See Lemma A1 below). Therefore, it remains to show that the first term is $o_{as}(1)$. We consider the expectation in the first term above. Namely

$$\begin{aligned}
&\mathrm{E}[p_L((Y_{p\ell}, i), (Y_{qj}, i))] \\
&= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int\left\{\frac{m_1(V_{p\ell}, X_\ell, i;\theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\left[\omega^G_{i,R+1,j}K_{G,h_G}(X_j - X_\ell)1\!\!1(B_{qj} \le B_{p\ell})\right.\right. \\
&\quad \left. -\frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)}\omega^g_{i,R,j}K_{1g,h_g}(X_j - X_\ell)K_{2g,h_g}(B_{qj} - B_{p\ell})\right] \\
&\quad +\frac{m_1(V_{qj}, X_j, i;\theta_0)}{g_0(B_{qj}|X_j, i)}\left[\omega^G_{i,R+1,j}K_{G,h_G}(X_\ell - X_j)1\!\!1(B_{p\ell} \le B_{qj})\right. \\
&\quad \left.\left. -\frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)}\omega^g_{i,R,j}K_{1g,h_g}(X_\ell - X_j)K_{2g,h_g}(B_{p\ell} - B_{qj})\right]\right\} \\
&\quad \times g_0(B_{p\ell}, X_\ell, i)g_0(B_{qj}, X_j, i)dY_{p\ell}dY_{qj} \\
&= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int\frac{1}{h_G}\left[\frac{m_1(V_{p\ell}, X_\ell, i;\theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_j - X_\ell}{h_G}\right)1\!\!1(B_{qj} \le B_{p\ell})\right. \\
&\quad \left. +\frac{m_1(V_{qj}, X_j, i;\theta_0)}{g_0(B_{qj}|X_j, i)}\omega^G_{i,R+1,j}K_G\left(\frac{X_\ell - X_j}{h_G}\right)1\!\!1(B_{p\ell} \le B_{qj})\right] \\
&\quad \times g_0(B_{p\ell}, X_\ell, i)g_0(B_{qj}, X_j, i)dY_{p\ell}dY_{qj} \\
&\quad -\frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int\frac{1}{h_g^2}\left[\frac{m_1(V_{p\ell}, X_\ell, i;\theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right)\right. \\
&\quad \times K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) + \frac{m_1(V_{qj}, X_j, i;\theta_0)}{g_0(B_{qj}|X_j, i)}\frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)}\omega^g_{i,R,j}K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) \\
&\quad \left. \times K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right)\right]g_0(B_{p\ell}, X_\ell, i)g_0(B_{qj}, X_j, i)dY_{p\ell}dY_{qj}
\end{aligned}$$

$$= \quad A_1 - A_2. \tag{A.22}$$

We consider first A_1.

$$\begin{aligned}
\|A_1\| &= \left\| \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)} \int \frac{1}{h_G}\left[\frac{m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right)\right.\right. \\
&\quad \left. \times 1\!\!I(B_{qj} \le B_{p\ell}) + \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)}\omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!I(B_{p\ell} \le B_{qj})\right] \\
&\quad \left. \times g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \right\| \\
&\le \|A_{11}\| + \|A_{12}\|
\end{aligned}$$

It is enough to show that $\|A_{11}\| = o_{as}(1/\sqrt{L})$ since the same argument can be used to show that $\|A_{12}\| = o_{as}(1/\sqrt{L})$. We observe the following

$$\begin{aligned}
\|A_{11}\| &= \left\| \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)} \int \frac{1}{h_G}\left[\frac{m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right)\right.\right. \\
&\quad \left.\left. \times 1\!\!I(B_{qj} \le B_{p\ell})\right] g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \right\| \\
&\le \left\| \sum_i \int \frac{1}{h_G} \frac{m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!I(B_{qj} \le B_{p\ell})\right. \\
&\quad \left. \times g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \right\| \\
&\le \left\| \sum_i \int \frac{1}{h_G} m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i; \theta_0)\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!I(B_{qj} \le B_{p\ell})\right. \\
&\quad \left. \times g_0(B_{p\ell}, X_\ell, i) dY_{p\ell} dY_{qj} \right\| \\
&= h_G \left\| \sum_i \int m_1(\xi(u h_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0)\omega^G_{i,R+1,j} K_G(-u_2)\right. \\
&\quad \left. \times 1\!\!I(B_{qj} \le u_1 h_G + B_{qj}) g_0(u h_G + Y_{qj}, i) du dY_{qj} \right\| \\
&\le h_G \left\| \sum_i \int m_1(\xi(u h_G + Y_{qj}, i), u_2 h_G + X_j, i; \theta_0) K_G(-u_2)\right. \\
&\quad \left. \times 1\!\!I(B_{qj} \le u_1 h_G + B_{qj}) g_0(u h_G + Y_{qj}, i) du dY_{qj} \right\|
\end{aligned} \tag{A.23}$$

where we have used the fact that densities are bounded and the change of variable $u = \dfrac{Y_{p\ell} - Y_{qj}}{h_G} = \left(\dfrac{B_{p\ell} - B_{qj}}{h_G}, \dfrac{X_\ell - X_j}{h_G}\right) = (u_1, u_2)$. The last inequality comes from observing that $\omega^G_{i,R+1,j} = O_p(1)$.

We consider now the expectation inside the norm of the term above evaluated at $h_G = 0$, namely

$$\begin{aligned}
&\sum_i \int m_1(\xi(Y_{qj}, i), X_j, i; \theta_0) g_0(Y_{qj}, i) dY_{qj} \\
&= \sum_i \int \left[\int m_1(\xi(Y_{qj}, i), X_j, i; \theta_0) g_0(B_{qj}|X_j, i) dB_{qj} \right] f_m(X_j, i) dX_j \\
&= \mathrm{E}[m_1(V, X, i; \theta_0)]
\end{aligned}$$

where we use A3-(ii) and the Law of Iterated Expectations. The last line in the expression above follows from observing that the integral inside can be solved by using twice integration by parts as follows

$$\begin{aligned}
&\int_{\underline{B}(X_j)}^{\overline{B}(X_j,i)} m_1(\xi(B_{qj}, X_j, i), X_j, i; \theta_0) g_0(B_{qj}|X_j, i) dB_{qj} \\
&= m_1(\xi(\overline{B}(X_j, i), X_j, i), X_j, i; \theta_0) G_0(\overline{B}(X_j, i)|X_j, i) - m_1(\xi(\underline{B}(X_j), X_j, i), X_j, i; \theta_0) \\
&\quad \times G_0(\underline{B}(X_j)|X_j, i) - \int_{\underline{B}(X_j)}^{\overline{B}(X_j,i)} m_{11}(\xi(B_{qj}, X_j, i), X_j, i; \theta_0) G_0(B_{qj}|X_j, i) dB_{qj} \\
&= m_1(\overline{V}, X_j, i; \theta_0) - \int_{\underline{B}(X_j)}^{\overline{B}(X_j,i)} m_{11}(\xi(B_{qj}, X_j, i), X_j, i; \theta_0) G_0(B_{qj}|X_j, i) dB_{qj} \\
&= m_1(\overline{V}, X_j, i; \theta_0) - m_1(\xi(\overline{B}(X_j, i), X_j, i), X_j, i; \theta_0) G_0(\overline{B}(X_j, i)|X_j, i) \\
&\quad + m_1(\xi(\underline{B}(X_j), X_j, i), Z_j; \theta_0) G_0(\underline{B}(X_j)|X_j, i) \\
&\quad + \int_{\underline{B}(X_j)}^{\overline{B}(X_j,i)} m_1(\xi(B_{qj}, X_j, i), X_j, i; \theta_0) g_0(B_{qj}|X_j, i) dB_{qj} \\
&= m_1(\overline{V}, X_j, i; \theta_0) - m_1(\overline{V}, X_j, i; \theta_0) \\
&\quad + \int_{\underline{B}(X_j)}^{\overline{B}(X_j,i)} m_1(\xi(B_{qj}, X_j, i), X_j, i; \theta_0) g_0(B_{qj}|X_j, i) dB_{qj} \\
&= \int_{\underline{V}(X_j,i)}^{\overline{V}(X_j,i)} m_1(V_{qj}, X_j, i; \theta_0) \frac{g_0(B_{qj}|X_j, i)}{\xi_1(B_{qj}, X_j, i)} dV_{qj} \\
&= \mathrm{E}[m_1(V, X, i; \theta_0)|X, i]
\end{aligned}$$

where the fifth equality uses $G_0(B_{qj}|X_j, i) = F(\xi(B_{qj}, X_j, i)|X_j, i)$, so that $g_0(B_{qj}|X_j, i) = f(V_{qj}|X_j, i)\xi_1(B_{qj}, X_j, i)$.

Therefore at $h_G = 0$ the integral inside the norm in (A.23) vanishes by A7-(i). Thus, we can apply a Taylor expansion of order $R + 1$ in the RHS of (A.23) around h_G to obtain

$$\|A_{11}\| \leq h_G \sum_i \left\| d_1 h_G + d_2 \frac{h_G^2}{2} + \ldots + d_R \frac{h_G^R}{R!} + O(h_G^{R+1}) \right\|$$

$$= \sum_i \left\| d_1 h_G^2 + d_2 \frac{h_G^3}{2} + \ldots + d_R \frac{h_G^{R+1}}{R!} + O(h_G^{R+2}) \right\|$$

We note that the remainder term vanishes, i.e. $\sqrt{L} h_G^{R+2} = o(1)$, and also that $\sqrt{L} h_G^{R+1} = o(1)$, by A4.AN-(i). The remaining $R-1$ terms also vanish by A3-(iii), i.e, since the kernels are of order $R-1$. To see this observe that the kth coordinate of d_ρ, $\rho = 1, \ldots, R-1$ is

$$\begin{aligned} d_{k_\rho} &= \frac{\partial^\rho}{\partial h_G^\rho} \int [H_k(uh_G + \overline{Y}) - H_k(uh_G + \underline{Y})] K_G(-u_2) du|_{h_G=0} \\ &= \sum_{k_1,\ldots,k_\rho=1}^{2} \int (u_{k_1} \ldots u_{k_\rho}) K_G(-u_2) \frac{\partial^\rho}{\partial Y_{k_1} \ldots \partial Y_{k_\rho}} H_k(\overline{Y}) du \\ &\quad - \sum_{k_1,\ldots,k_\rho=1}^{2} \int (u_{k_1} \ldots u_{k_\rho}) K_G(-u_2) \frac{\partial^\rho}{\partial Y_{k_1} \ldots \partial Y_{k_\rho}} H_k(\underline{Y}) du \\ &= 0, \end{aligned}$$

where $dH_k/dY(y) = m_{1,k}(\xi(y,i), x, i; \theta_0) g_0(y,i)$. The third equality uses A3-(iii), that is since $K_G(\cdot)$ is a higher order kernel, all moments of order strictly smaller than $R-1$ vanish.

It remains to consider now A_2 in (A.22). Namely

$$\begin{aligned} \|A_2\| &= \Bigg\| \frac{1}{2} \frac{L}{n_i} \sum_i \frac{1}{(i-1)} \int \frac{1}{h_g^2} \Bigg[\frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \\ &\quad \times \omega_{i,R,j}^g K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) \\ &\quad + \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)} \frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)} \omega_{i,R,j}^g K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right) \Bigg] \\ &\quad \times g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \Bigg\| \\ &\leq \|A_{21}\| + \|A_{21}\| \end{aligned}$$

We show only that $A_{21} = o_{as}(1/\sqrt{L})$ since a similar argument can be used to show that $A_{22} = o_{as}(1/\sqrt{L})$. We observe the following

$$\begin{aligned} \|A_{21}\| &= \Bigg\| \frac{1}{2} \frac{L}{n_i} \sum_i \frac{1}{(i-1)} \int \frac{1}{h_g^2} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \\ &\quad \times \omega_{i,R,j}^g K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \Bigg\| \\ &\leq \Bigg\| \sum_i \int \frac{1}{h_g^2} \frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)} \frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)} \omega_{i,R,j}^g K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) \\ &\quad \times K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \Bigg\| \end{aligned}$$

$$
\begin{aligned}
\leq & \left\| \sum_i \int \frac{1}{h_g^2} m_1(\xi(B_{p\ell}, X_\ell, i), X_\ell, i;\theta_0)\omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) \right. \\
& \left. \times g_0(B_{p\ell}, X_\ell, i) g_0(B_{qj}, X_j, i) dY_{p\ell} dY_{qj} \right\| \\
= & \left\| \sum_i \int m_1(\xi(uh_g + Y_{qj}, i), u_2 + X_j, i;\theta_0)\omega^g_{i,R,j} K_{1g,h_g}(-u_2) K_{2g,h_g}(-u_1) \right. \\
& \left. \times g_0(uh_g + Y_{qj}, i) du dY_{qj} \right\| \\
\leq & \left\| \sum_i \int m_1(\xi(uh_g + Y_{qj}, i), u_2 + X_j, i;\theta_0) K_{1g,h_g}(-u_2) K_{2g,h_g}(-u_1) \right. \\
& \left. \times g_0(uh_g + Y_{qj}, i) du dY_{qj} \right\|,
\end{aligned}
$$

where we have used that $(1/2)(L/n_i)1/(i-1) \leq \infty$ and also that densities are bounded. The last equality uses the change of variable $u = (Y_{p\ell} - Y_{qj})/h_g$ and the last inequality comes from observing that $\omega^g_{i,R,j} = O_p(1)$.

We observe that A_{21} can be expanded as a Taylor series of order R in the bandwidth h_g. Moreover, $A_{21} = 0|_{h_g=0}$ by A7-(i) as we have already shown above for $A_{11} = 0|_{h_G=0}$.

Therefore, by A7-(i) we can apply a Taylor expansion around h_g to obtain

$$
\|A_{21}\| \leq \sum_{I_j} \left\| c_1 h_g + c_2 \frac{h_g^2}{2} + \ldots + c_{R-1}\frac{h_g^{R-1}}{(R-1)!} + O(h_g^R) \right\|.
$$

We note that the remainder term vanishes, i.e. $\sqrt{L}h_g^R = o(1)$ by A4.AN-(ii). The remaining $R-1$ terms also vanish by A3-(iii). To see this observe that the kth coordinate of c_ρ, $\rho = 1, \ldots, R-1$ is

$$
\begin{aligned}
c_{k_\rho} &= \frac{\partial^\rho}{\partial h_g^\rho} \int [H_k(uh_g + \overline{Y}) - H_k(uh_g + \underline{Y})] K_{1g}(-u_2) K_{2g}(-u_1) du|_{h_g=0} \\
&= \sum_{k_1,\ldots,k_\rho=1}^{2} \int (u_{k_1}\ldots u_{k_\rho}) K_{1g}(-u_2) K_{2g}(-u_1) \frac{\partial^\rho}{\partial Y_{k_1}\ldots \partial Y_{k_\rho}} H_k(\overline{Y}) du \\
&\quad - \sum_{k_1,\ldots,k_\rho=1}^{2} \int (u_{k_1}\ldots u_{k_\rho}) K_{1g}(-u_2) K_{2g}(-u_1) \frac{\partial^\rho}{\partial Y_{k_1}\ldots \partial Y_{k_\rho}} H_k(\underline{Y}) du \\
&= 0,
\end{aligned}
$$

where $dH_k/dY(y) = m_{1,k}(\xi(y,i), x, i;\theta_0) g_0(y,i)$. The third equality uses A3-(iii), that is since $K_{1g}(\cdot)$ and $K_{2g}(\cdot)$ are higher order kernels, all moments of order strictly smaller than $R-1$ vanish.

We consider next B_{12} in (A.9)

$$\begin{aligned}
\|B_{12}\| &= \Bigg\| \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell} m_1(V_{p\ell,Z_\ell;\theta_0})\sqrt{L}\Bigg[\frac{1}{\hat{g}(B_{p\ell}|Z_\ell)g_0(B_{p\ell}|Z_\ell)}\Bigg(\frac{G_0(B_{p\ell}|Z_\ell)}{g_0(B_{p\ell}|Z_\ell)} \\
&\quad \times [\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)]^2 - \Big[\hat{G}(B_{p\ell}|Z_\ell)-G_0(B_{p\ell}|Z_\ell)\Big][\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)]\Bigg)\Bigg]\Bigg\| \\
&\le \left(\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell}\|m_1(V_{p\ell},Z_\ell;\theta_0)\|^2\right)^{\frac{1}{2}} \\
&\quad \times\Bigg\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell} L\Bigg[\frac{1}{\hat{g}(B_{p\ell}|Z_\ell)g_0(B_{p\ell}|Z_\ell)}\Bigg(\frac{G_0(B_{p\ell}|Z_\ell)}{g_0(B_{p\ell}|Z_\ell)}[\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)] \\
&\quad - \Big[\hat{G}(B_{p\ell}|Z_\ell)-G_0(B_{p\ell}|Z_\ell)\Big]\Big[\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)\Big]\Bigg)\Bigg]^2\Bigg\}^{\frac{1}{2}} \\
&= B_{121}B_{122},
\end{aligned}$$

where the inequality comes from Cauchy-Schwartz.

First we show that $B_{121}^2 < \infty$.

$$\begin{aligned}
B_{121}^2 &= \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell}\|m_1(V_{p\ell},Z_\ell;\theta_0)\|^2 \\
&\le \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell}\sup_{\theta\in\Theta}\|m_1(V_{p\ell},Z_\ell;\theta)\|^2 \\
&\le \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell}K_7(V_{p\ell},Z_\ell)^2 \\
&\le \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}K_7(V_{p\ell},Z_\ell)^2 \\
&= \mathrm{E}[K_7(V,Z)^2]+o_{as}(1)<\infty
\end{aligned}$$

where the second inequality comes from A6-(vi) and the fact that $0 < 1/(I_\ell - 1) \le 1$.

Next we show that $B_{122} = o(1)$.

$$\begin{aligned}
B_{122} = &\Bigg\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell} L\Bigg[\frac{1}{\hat{g}(B_{p\ell}|Z_\ell)g_0(B_{p\ell}|Z_\ell)}\Bigg(\frac{G_0(B_{p\ell}|Z_\ell)}{g_0(B_{p\ell}|Z_\ell)}\Big[\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)\Big]^2 \\
&- \Big[\hat{G}(B_{p\ell}|Z_\ell)-G_0(B_{p\ell}|Z_\ell)\Big]\Big[\hat{g}(B_{p\ell}|Z_\ell)-g_0(B_{p\ell}|Z_\ell)\Big]\Bigg)\Bigg]^2\Bigg\}^{\frac{1}{2}} \\
< &\sqrt{L}\left\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell(I_\ell-1)}\sum_{p=1}^{I_\ell}\left[\kappa_1\left(\kappa_2 O\left(\frac{1}{{r_g}^2}\right)-O\left(\frac{1}{r_G}\right)O\left(\frac{1}{r_g}\right)\right)\right]^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{L}\kappa_1 \left[\kappa_2 O\left(\frac{1}{{r_g}^2}\right) - O\left(\frac{1}{r_G}\right) O\left(\frac{1}{r_g}\right)\right] \\
&= \kappa_1 \left[\kappa_2 O\left(\frac{\sqrt{L}}{{r_g}^2}\right) - O\left(\frac{\sqrt{L}}{r_G r_g}\right)\right] = o(1),
\end{aligned}
$$

where we have used

- $\left|\dfrac{1}{\hat{g}(B_{p\ell}|Z_\ell)g_0(B_{p\ell}|Z_\ell)}\right| < \kappa_1 < \infty,$

  since densities are bounded away from zero and $\hat{g}(B_{p\ell}|Z_\ell) \xrightarrow{a.s} g_0(B_{p\ell}|Z_\ell)$
- $\left|\dfrac{G_0(B_{p\ell}|Z_\ell)}{g_0(B_{p\ell}|Z_\ell)}\right| < \kappa_2 < \infty,$
  since $g_0(\cdot,\cdot)$ is bounded away from zero.
-

$$
O\left(\frac{1}{r_g^2}\right) = \left|\hat{g}(B_{p\ell}|Z_\ell) - g_0(B_{p\ell}|Z_\ell)\right|^2 = O\left(h_{1g}^{2R} + h_{2g}^{2R} + \frac{\log L}{L h_{1g} h_{2g}}\right)
$$

$$
\begin{aligned}
O\left(\frac{1}{r_G}\right) O\left(\frac{1}{r_g}\right) &= \left|\hat{G}(B_{p\ell}|Z_\ell) - G_0(B_{p\ell}|Z_\ell)\right|\left|\hat{g}(B_{p\ell}|Z_\ell) - g_0(B_{p\ell}|Z_\ell)\right| \\
&= O\left(h_G^{R+1} + \sqrt{\frac{\log L}{L h_G}}\right) O\left(h_{1g}^R + h_{2g}^R + \sqrt{\frac{\log L}{L h_{1g} h_{2g}}}\right)
\end{aligned}
$$

- For each ℓ, $0 < 1/(I_\ell - 1) \leq 1$.

Therefore, the second term in (A.9) is $o(1)$, i.e. $B_{12} = o(1)$

Step 3.B2

We consider B_2 in (A.8)

$$
\begin{aligned}
\|B_2\| &\leq \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell}\|m_1(V_{p\ell}, Z_\ell;\theta_0) - m_1(V_{p\ell}^*, Z_\ell;\theta_0)\||\hat{V}_{p\ell} - V_{p\ell}| \\
&\leq \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell} K_6(Z_\ell)|V_{p\ell} - V_{p\ell}^*||\hat{V}_{p\ell} - V_{p\ell}| \\
&\leq \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{I_\ell}\sum_{p=1}^{I_\ell} K_6(Z_\ell)(\hat{V}_{p\ell} - V_{p\ell})^2 \\
&\leq \sqrt{L}\sup_{p,\ell}(\hat{V}_{p\ell} - V_{p\ell})^2 \frac{1}{L}\sum_{\ell=1}^{L} K_6(Z_\ell) \\
&= \sqrt{L} O_{as}\left(\frac{1}{r^2}\right) O_{as}(1)
\end{aligned}
$$

$$= O_{as}\left(\frac{\sqrt{L}}{r^2}\right)O_{as}(1)$$
$$= O_{as}\left(\frac{L^{1/4}}{r}\right)O_{as}(1) = o_{as}(1),$$

where the second inequality comes from A6-(v), the third uses the fact that $\hat{V}_{p\ell} \leq V^*_{p\ell} \leq V_{p\ell}$. The last equality follows from A4.AN.

Therefore, the desired result follows. *Q.E.D.*

LEMMA A1: *Let $\hat{\theta}$ be defined as in (2.4). Then, under A1-A3, A4.AN and A5-A.7, we have*

$$\begin{aligned}
\sqrt{L}(\hat{\theta}-\theta_0) &= \sqrt{L}(\tilde{\theta}-\theta_0) - \frac{L(L-1)}{L^2}\frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\Bigg\{\sum_i \frac{1}{i(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)\\
&\quad \times g_0(Y_{p\ell},i) + \mathrm{E}\left[\sum_i \frac{1}{i(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\right]\Bigg\} + o_p(1)\\
&= -\frac{1}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\Bigg\{(C^T\Omega C)^{-1}C\Omega m(V_{p\ell},X_\ell;\theta_0,i) + 2\frac{L(L-1)}{L^2}\\
&\quad \times \sum_i \frac{1}{i(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\\
&\quad - \mathrm{E}\left[\sum_i \frac{1}{i(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\right]\Bigg\} + o_p(1)
\end{aligned}$$

where $C = \mathrm{E}[m_3(V,X;\theta_0,i)]$, and

$N(Y_{p\ell},i) = [m_1(V_{p\ell},X_\ell,i;\theta_0)/g_0(B_{p\ell}|X_\ell,i)^2]G_0(B_{p\ell}|X_\ell,i).$

Proof

From proposition 2 we have

$$\sqrt{L}(\hat{\theta}-\theta_0) - \sqrt{L}(\tilde{\theta}-\theta_0) = \frac{L(L-1)}{L^2}\frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}[r_L(Y_{p\ell},i)-\theta_L]$$

where

$Y_{p\ell} = (B_{p\ell}, X_\ell)$

$r_L(Y_{p\ell},i) = \mathrm{E}[p_L((Y_{p\ell},i),(Y_{qj},i))|(Y_{p\ell},i)]$

$\theta_L = \mathrm{E}[r_L(Y_{p\ell},i)] = \mathrm{E}[p_L((Y_{p\ell},i),(Y_{qj},i))].$

First we show that $r_L(Y_{p\ell}, i) = -\sum_i \frac{1}{i(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) + t_L(Y_{p\ell}, i)$.
We observe that

$$r_L(Y_{p\ell}, i) = \mathrm{E}[p_L((Y_{p\ell}, i), (Y_{qj}, i))|(Y_{p\ell}, i)]$$
$$= \begin{cases} \int p_L((B_{p\ell}, X_\ell, i), (B_{qj}, X_j, i)) g_0(B_{qj}, X_j, i) dY_{qj} & \text{if } \ell \neq j \\ \int p_L((B_{pj}, X_j, i), (B_{qj}, X_j, i)) g_0((B_{pj}, X_j, i), (B_{qj}, X_j, i)|(B_{pj}, X_j, i)) dY_{qj} & \text{if } \ell = j. \end{cases}$$

We consider first the case $\ell \neq j$.

$$\begin{aligned}
& r_L(Y_{p\ell}, i) = \\
= \; & \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\Bigg\{\int\Bigg[\frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\frac{1}{h_G}\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!1(B_{qj} \leq B_{p\ell}) \\
& + \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)}\frac{1}{h_G}\omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!1(B_{p\ell} \leq B_{qj})\Bigg] g_0(Y_{qj}, i) dY_{qj}\Bigg\} \\
& - \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\Bigg\{\int\Bigg[\frac{m_1(V_{p\ell}, X_\ell, i; \theta_0)}{g_0(B_{p\ell}|X_\ell, i)}\frac{G_0(B_{p\ell}|X_\ell, i)}{g_0(B_{p\ell}|X_\ell, i)}\frac{1}{h_g^2}\omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) \\
& \times K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) + \frac{m_1(V_{qj}, X_j, i; \theta_0)}{g_0(B_{qj}|X_j, i)}\frac{G_0(B_{qj}|X_j, i)}{g_0(B_{qj}|X_j, i)}\frac{1}{h_g^2}\omega^g_{i,R,j} K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) \\
& \times K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right)\Bigg] g_0(Y_{qj}, i) dY_{qj}\Bigg\} \\
= \; & \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\Bigg\{\int\Bigg[M(Y_{p\ell}, i)\frac{1}{h_G}\omega^G_{i,R+1,j} K_G\left(\frac{X_j - X_\ell}{h_G}\right) 1\!\!1(B_{qj} \leq B_{p\ell}) \\
& + M(Y_{qj}, i)\frac{1}{h_G}\omega^G_{i,R+1,j} K_G\left(\frac{X_\ell - X_j}{h_G}\right) 1\!\!1(B_{p\ell} \leq B_{qj})\Bigg] g_0(Y_{qj}, i) dY_{qj}\Bigg\} \\
& - \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\Bigg\{\int\Bigg[N(Y_{p\ell}, i)\frac{1}{h_g^2}\omega^g_{i,R,j} K_{1g}\left(\frac{X_j - X_\ell}{h_g}\right) K_{2g}\left(\frac{B_{qj} - B_{p\ell}}{h_g}\right) \\
& + N(Y_{qj}, i)\frac{1}{h_g^2}\omega^g_{i,R,j} K_{1g}\left(\frac{X_\ell - X_j}{h_g}\right) K_{2g}\left(\frac{B_{p\ell} - B_{qj}}{h_g}\right)\Bigg] g_0(Y_{qj}, i) dY_{qj}\Bigg\} \\
= \; & \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int h_G\Big[M(Y_{p\ell}, i)\omega^G_{i,R+1,j} K_G(u_2) 1\!\!1(B_{p\ell} \leq u_1 h_G + B_{p\ell}) + M(u h_G + Y_{p\ell}, i) \\
& \times \omega^G_{i,R+1,j} K_G(-u_2) 1\!\!1(u_1 h_G + B_{p\ell} \leq B_{p\ell})\Big] g_0(u h_G + Y_{p\ell}, i) du \\
& - \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int\Big[N(Y_{p\ell}, i)\omega^g_{i,R,j} K_{1g}(u_2) K_{2g}(u_1) + N(u h_g + Y_{p\ell}, i) \\
& \times \omega^g_{i,R,j} K_{1g}(-u_2) K_{2g}(-u_1)\Big] g_0(u h_g + Y_{p\ell}, i) du.
\end{aligned}$$

We note that as $h = (h_G, h_g) \to 0$ we have

$$\begin{aligned} r_L(Y_{p\ell}, i) \quad &\to \quad -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)} \int \Big[N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) K_{1g}(u)] K_{1g}(u_2) K_{2g}(u_1) \\ &+ N(Y_{p\ell}, I_j) f_m^{-1}(X_\ell, i) K_{1g}(-u_2) K_{2g}(-u_1) \Big] g_0(Y_{p\ell}, i) du \\ &= -\frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) \end{aligned}$$

where we have used the following

$$\begin{aligned} \omega^g_{i,R,j} \quad &= \quad e_1^T \left[\frac{1}{n_i h_g} \sum_{s=1}^{n_i} \mathbf{x_s} \mathbf{x_s}^T K_{1g}\left(\frac{X_s - X_\ell}{h_g} \right) \right]^{-1} [1 \quad (X_j - X_\ell) \ldots (X_j - X_\ell)^{R-1}]^T \\ &= \quad e_1^T \left[\frac{1}{n_i h_g} \sum_{s=1}^{n_i} \mathbf{x_s} \mathbf{x_s}^T K_{1g}\left(\frac{X_s - X_\ell}{h_g} \right) \right]^{-1} [1 \quad (-u_2 h_g) \ldots (-u_2 h_g)^{R-1}]^T \\ &\xrightarrow{p} \quad e_1^T \left[\mathrm{E}\left(\mathbf{x_s} \mathbf{x_s}^{\mathrm{T}} \mathrm{K_{1g}} \left(\frac{\mathrm{X_s - X_\ell}}{\mathrm{h_g}} \right) \right) \right]^{-1} e_1 = f_m^{-1}(X_\ell, i) \end{aligned}$$

therefore we define

$$r_L(Y_{p\ell}, I_j) = -\frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) + t_L(Y_{p\ell}, i)$$

We consider now the reminder term $t_L(Y_{p\ell}, i)$

$$\begin{aligned} t_L(Y_{p\ell}, i) \quad &= \quad r_L(Y_{p\ell}, i) + \frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) \\ &= \quad \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)} \int h_G \Big[M(Y_{p\ell}, i) \omega^G_{i,R+1,j} K_G(u_2) \mathbb{1}(B_{p\ell} \le u_1 h_G + B_{p\ell}) \\ &+ M(u h_G + Y_{p\ell}, i) \omega^G_{i,R+1,j} K_G(-u_2) \mathbb{1}(u_1 h_G + B_{p\ell} \le B_{p\ell}) \Big] g_0(u h_G + Y_{p\ell}, i) du \\ &- \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)} \int \Big[N(Y_{p\ell}, i) \omega^g_{i,R,j} K_{1g}(u_2) K_{2g}(u_1) + N(u h_g + Y_{p\ell}, i) \\ &\times \omega^g_{i,R,j} K_{1g}(-u_2) K_{2g}(-u_1) \Big] g_0(u h_g + Y_{p\ell}, i) du \\ &+ \frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) \end{aligned}$$

Now, using $\int K_{1g}(u_2) K_{2g}(u_1) du = 1$ we can write

$$\begin{aligned} &\frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) \\ &= \quad \frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)} \int K_{1g}(u_2) K_{2g}(u_1) N(Y_{p\ell}, i) f_m^{-1}(X_\ell, i) g_0(Y_{p\ell}, i) du \end{aligned}$$

$$+\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int K_{1g}(-u_2)K_{2g}(-u_1)N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)du$$

therefore we can write the reminder term as follows

$$\begin{aligned}
t_L(Y_{p\ell},I_j) &= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int h_G\Big[M(Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(u_2)1\!\!1(B_{p\ell}\le u_1h_G+B_{p\ell}) \\
&\quad +M(uh_G+Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(-u_2)1\!\!1(u_1h_G+B_{p\ell}\le B_{p\ell})\Big]g_0(uh_G+Y_{p\ell},i)du \\
&\quad -\frac{1}{2}\sum_i \frac{1}{(i-1)}\int \frac{L}{iL_i}N(Y_{p\ell},i)\omega^g_{i,R,j}K_{1g}(u_2)K_{2g}(u_1)g_0(uh_g+Y_{p\ell},i)du \\
&\quad \frac{1}{2}\sum_i \frac{1}{(i-1)}\int \frac{L}{iL_i}N(uh_g+Y_{p\ell},i)\omega^g_{i,R,j}K_{1g}(-u_2)K_{2g}(-u_1)g_0(uh_g+Y_{p\ell},i)du \\
&\quad +\frac{1}{2}\sum_i \frac{1}{(i-1)}\int \frac{1}{i}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)K_{1g}(u_2)K_{2g}(u_1)g_0(Y_{p\ell},i)du \\
&\quad +\frac{1}{2}\sum_i \frac{1}{(i-1)}\int \frac{1}{i}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)K_{1g}(-u_2)K_{2g}(-u_1)g_0(Y_{p\ell},i)du \\
&= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int h_G\Big[M(Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(u_2)1\!\!1(B_{p\ell}\le u_1h_G+B_{p\ell}) \\
&\quad +M(uh_G+Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(-u_2)1\!\!1(u_1h_G+B_{p\ell}\le B_{p\ell})\Big]g_0(uh_G+Y_{p\ell},i)du \\
&\quad -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int N(Y_{p\ell},i)K_{1g}(u_2)K_{2g}(u_1) \\
&\quad \times\left[\frac{L}{L_i}\omega^g_{i,R,j}g_0(uh_g+Y_{p\ell},i)-f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\right]du \\
&\quad -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int K_{1g}(-u_2)K_{2g}(-u_1) \\
&\quad \times\left[\frac{L}{L_i}\omega^g_{i,R,j}N(uh_g+Y_{p\ell},i)g_0(uh_g+Y_{p\ell},i)-f_m^{-1}(X_\ell,i)N(Y_{p\ell},i)g_0(Y_{p\ell},i)\right]du \\
&= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int h_G\Big[M(Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(u_2)1\!\!1(B_{p\ell}\le u_1h_G+B_{p\ell}) \\
&\quad +M(uh_G+Y_{p\ell},i)\omega^G_{i,R+1,j}K_G(-u_2)1\!\!1(u_1h_G+B_{p\ell}\le B_{p\ell})\Big]g_0(uh_G+Y_{p\ell},i)du \\
&\quad -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int N(Y_{p\ell},i)K_{1g}(u_2)K_{2g}(u_1)\left[f_m^{-1}(X_\ell,i)+o_{as}(1)\right] \\
&\quad \times[g_0(uh_g+Y_{p\ell},i)-g_0(Y_{p\ell},i)]du \\
&\quad -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int K_{1g}(-u_2)K_{2g}(-u_1)\left[f_m^{-1}(X_\ell,i)+o_{as}(1)\right] \\
&\quad \times[N(uh_g+Y_{p\ell},i)g_0(uh_g+Y_{p\ell},i) \\
&\quad -N(Y_{p\ell},i)g_0(Y_{p\ell},i)]du
\end{aligned}$$

thus using the above expression we have

$$
\begin{aligned}
&\frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}[r_L(Y_{p\ell},i)-\theta_L]=\\
&= \frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\Big\{-\frac{1}{i}\sum_{i}\frac{1}{(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\\
&\quad +\mathrm{E}\Big[\frac{1}{i}\sum_{i}\frac{1}{(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\Big]+t_L(Y_{p\ell},i)-\mathrm{E}[t_L(Y_{p\ell},i)]\Big\}\\
&= -\frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\Big\{\frac{1}{i}\sum_{i}\frac{1}{(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\\
&\quad -\mathrm{E}\Big[\frac{1}{i}\sum_{i}\frac{1}{(i-1)}N(Y_{p\ell},i)f_m^{-1}(X_\ell,i)g_0(Y_{p\ell},i)\Big]\Big\}\\
&\quad +\frac{2}{\sqrt{L}}\sum_{\{\ell:I_\ell=i\}}^{L}\frac{1}{i}\sum_{p=1}^{i}\Big[t_L(Y_{p\ell},i)-\mathrm{E}[t_L(Y_{p\ell},i)]\Big]
\end{aligned}
$$

We denote the second term above by T_L and we observe that $\mathrm{E}[T_L]=0$. We now show that $\mathrm{var}[T_L]=o_{as}(1)$.

$$
\begin{aligned}
\mathrm{var}[T_L] &= 4\frac{L_i}{L}\mathrm{var}\left[\frac{1}{i}\sum_{p=1}^{i}t_1(Y_{p1},i)\right]\\
&= 4\frac{L_i}{L}\mathrm{E}\left\{\mathrm{var}\left[\frac{1}{i}\sum_{p=1}^{i}t_1(Y_{p1},i)\middle|i\right]\right\}+4\frac{L_i}{L}\mathrm{var}\left\{\mathrm{E}\left[\frac{1}{i}\sum_{p=1}^{i}t_1(Y_{p1},i)\middle|i\right]\right\}\\
&= 4\frac{L_i}{L}\mathrm{E}\left\{\frac{1}{i}\mathrm{var}\left[t_1(Y_{p1},i)\middle|i\right]\right\}+4\frac{L_i}{L}\mathrm{var}\Big\{\mathrm{E}\left[t_1(Y_{p1},i)|i\right]\Big\}\\
&= A+B
\end{aligned}
\tag{A.24}
$$

We consider first the kth coordinate of the conditional variance inside the A term above, namely

$$
\begin{aligned}
\mathrm{var}\left[t_{1_k}(Y_{p1},i)\middle|i\right] &\le \mathrm{E}\left[t_{1_k}(Y_{p1},i)^2\middle|i\right]\\
&\le O\left(h_G^2\right)+O\left(h_g^{2(R-1)}\right)
\end{aligned}
$$

where the last inequality comes from observing that

$$
\begin{aligned}
t_{1_k}(Y_{p1},i) &= \frac{1}{2}\frac{L}{iL_i}\sum_{i}\frac{1}{(i-1)}\int h_G\Big[M(Y_{p1},i)\omega^G_{i,R+1,j}K_G(u_2)\mathbb{1}(B_{p1}\le u_1h_G+B_{p1})\\
&\quad +M(uh_G+Y_{p1},i)\omega^G_{i,R+1,j}K_G(-u_2)\mathbb{1}(u_1h_G+B_{p1}\le B_{p1})\Big]g_0(uh_G+Y_{p1},i)du
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int N(Y_{p1},i)K_{1g}(u_2)K_{2g}(u_1)f_m^{-1}(X_1,i) \\
&\times[g_0(uh_g+Y_{p1},i)-g_0(Y_{p1},i)]du \\
&-\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)}\int K_{1g}(-u_2)K_{2g}(-u_1)f_m^{-1}(X_1,i)[N(uh_g+Y_{p1},i) \\
&\times g_0(uh_g+Y_{p1},i)-N(Y_{p1},i)g_0(Y_{p1},i)]du+o_{as}(1) \\
= \quad & a+b+c+o_{as}(1)
\end{aligned}
$$

therefore, twice application of $(a+b)^2 \leq 2(a^2+b^2)$ yields $(a+b+c)^2 \leq \kappa(a^2+b^2+c^2)$, thus

$$
\begin{aligned}
\mathrm{E}\left[t_{1_k}(Y_{p1},i)^2\middle|i\right] &= \mathrm{E}[(a+b+c)^2|i]+o_{as}(1) \\
&\leq 4\mathrm{E}[a^2+b^2+c^2|i]+o_{as}(1) \\
&= O\left(h_G^2\right)+O\left(h_g^{2(R-1)}\right)+O\left(h_g^{2(R-1)}\right)
\end{aligned}
$$

where the order of the last two terms after the last equality above follows from an $(R-1)$th Taylor Expansion around Y_{p1} and the fact that kernels are of order $R-1$ by A.3-(iii).

We consider now the B term in (A.24) and more precisely we consider the following

$$
\begin{aligned}
\frac{B}{4} &= \frac{L_i}{L}\mathrm{var}\Big\{\mathrm{E}\left[t_1(Y_{p1},i)|i\right]\Big\} \\
&\leq \mathrm{E}\Big\{\mathrm{E}\left[t_1(Y_{p1},i)|i\right]^2\Big\} \\
&\leq \mathrm{E}\Big\{\mathrm{E}\left[t_1(Y_{p1},i)^2|i\right]\Big\} \\
&\leq O\left(h_G^2\right)+O\left(h_g^{2(R-1)}\right)+O\left(h_g^{2(R-1)}\right)
\end{aligned}
$$

where the last inequality follows from the same argument used above. Hence, by Chebyshev Inequality $T_L = o_p(1)$.

We consider now the case $\ell = j$ and observe the following

$$
\begin{aligned}
&r_L(Y_{pj},i) \\
&= \mathrm{E}[p_L((B_{pj},X_j,i),(B_{qj},X_j,i))|(B_{pj},X_j,i)] \\
&= \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int \frac{1}{h_G}\bigg[M(Y_{pj},i)\omega^G_{i,R+1,j}K_G(0)\mathbb{1}(B_{qj}\leq B_{pj}) \\
&\quad +M(Y_{qj},i)\omega^G_{i,R+1,j}K_G(0)\mathbb{1}(B_{pj}\leq B_{qj})\bigg]g_0((Y_{pj},i),(Y_{qj},i)|(Y_{pj},i))dY_{qj} \\
&\quad -\frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)}\int \frac{1}{h_g^2}\bigg[N(Y_{pj},i)\omega^g_{i,R,j}K_{1g}(0)K_{2g}\left(\frac{B_{qj}-B_{pj}}{h_g}\right) \\
&\quad +N(Y_{qj},i)\omega^g_{i,R,j}K_{1g}(0)K_{2g}\left(\frac{B_{pj}-B_{qj}}{h_g}\right)\bigg]g_0((Y_{pj},i),(Y_{qj},i)|(Y_{pj},i))dY_{qj}
\end{aligned}
$$

We now use the change of variable $u = (Y_{qj} - Y_{pj})/h_G$ and $\tilde{u} = (Y_{qj} - Y_{pj})/h_g$ to obtain

$$\begin{aligned}
& r_L(Y_{pj}, i) \\
= \;& \frac{1}{2}\frac{L}{n_i}\sum_i \frac{1}{(i-1)} \int h_G \Big[M(Y_{pj}, i)\omega^G_{i,R+1,j} K_G(0) \mathbb{1}(u_1 h_G + B_{pj} \leq B_{pj}) + M(uh_G + Y_{pj}, i) \\
& \times \omega^G_{i,R+1,j} K_G(0) \mathbb{1}(B_{pj} \leq u_1 h_G + B_{pj}) \Big] g_0((Y_{pj}, i), (uh_G + Y_{pj}, i) | (Y_{pj}, i)) du \\
& - \frac{1}{2}\frac{L}{iL_i}\sum_i \frac{1}{(i-1)} \int \Big[N(Y_{pj}, i)\omega^g_{i,R,j} K_{1g}(0) K_{2g}(\tilde{u}_1) + N(\tilde{u}h_g + Y_{pj}, i) \\
& \times \omega^g_{i,R,j} K_{1g}(0) K_{2g}(-\tilde{u}_1) \Big] g_0((Y_{pj}, i), (uh_g + Y_{pj}, i) | (Y_{pj}, i)) d\tilde{u}
\end{aligned}$$

next, we observe that as $h = (h_G, h_g) \to 0$ we have

$$\begin{aligned}
r_L(Y_{pj}, i) \;\; \to \;\; & -\frac{1}{2}\frac{1}{i}\sum_i \frac{1}{(i-1)} \int \Big[N(Y_{pj}, i) f_m^{-1}(X_j, i) K_{1g}(0) K_{2g}(\tilde{u}_1) \\
& + N(Y_{pj}, i) f_m^{-1}(X_j, i) K_{1g}(0) K_{2g}(-\tilde{u}_1) \Big] g_0((Y_{pj}, i), (Y_{pj}, i) | (Y_{pj}, i)) d\tilde{u} \\
= \;\; & -\frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{pj}, i) f_m^{-1}(X_j, i) g_0(Y_{pj}, i)
\end{aligned}$$

as before we define

$$r_L(Y_{pj}, i) = -\frac{1}{i}\sum_i \frac{1}{(i-1)} N(Y_{pj}, i) f_m^{-1}(X_j, i) g_0(Y_{pj}, i) + t_L(Y_{pj}, i)$$

The rest of the proof is analogous to the one for the case $\ell \neq j$. Therefore the desired result follows. *Q.E.D.*

Appendix B

Choices of Kernels and Bandwidths

As it is well known in the nonparametric econometric literature, the choice of kernel is not crucial in practice. The estimators in chapter 3 are multivariate kernels which are computed as the product of univariate kernels. That is

$$K_m\left(\frac{a-A_k}{h_g},\frac{b-B_k}{h_g},\frac{n-N_k}{h_{gn}}\right)=K_a\left(\frac{a-A_k}{h_g}\right)K_b\left(\frac{b-B_k}{h_g}\right)K_n\left(\frac{n-N_k}{h_{gn}}\right),$$

where $K_m(\cdot)$ refers to the multivariate kernel, $K_a(\cdot)$ and $K_b(\cdot)$ denote the univariate kernels corresponding to the continuous variables A and B, say, and $K_n(\cdot)$ is the kernel for the discrete variables. Recall that $K_n \equiv K_{n_0}K_{n_1}K_{n_2}$.

The econometric procedure follows closely that of Guerre, Perrigne, and Vuong (2000). Accordingly, the kernels for continuous variables are required to be symmetric with bounded supports (see Assumption A3 in Guerre, Perrigne, and Vuong (2000)). Thus, we decide to use the triweight kernel function defined as $K(u) = 35/32(1-u^2)^3\,1\!\!1(|u|\leq 1)$ for these variables, namely d, x and c. The compact support of this function implies that only non–trimmed private costs are used in the second step to obtain the corresponding latent densities.

For the kernels involving discrete variables we use a Gaussian kernel. The main reason to change the kernel functions for the discrete variables has to do with the nature of these variables. That is, given that relatively small variation in the number of bidders it is desirable to give more weight to observations farther from the point at which estimation takes place. This is best achieved with a kernel with unbounded support.[1]

The smoothness of the distribution of private values is denoted by R, we assume $R = 1$. The bandwidths' choice is critical in nonparametric estimation. To ensure the uniform consistency at the optimal convergence rates of the estimators the bandwidths for the continuous variables are of the following form: $h_g = 1.06\times 2.978\times\hat{\sigma}\times(T)^{-1/(2R+4)}$, $h_G = 1.06\times 2.978\times\hat{\sigma}\times(T)^{-1/(2R+3)}$, $h_f = 1.06\times 2.978\times\hat{\sigma}\times(T_\tau)^{-1/(2R+4)}$ $h_q = 1.06\times 2.978\times\hat{\sigma}\times(T_\tau)^{-1/(2R+3)}$. The constant term

[1]There are no theoretical restrictions to the kernels applied to discrete variables.

comes from the so–called rule of thumb and the factor 2.978 is the one corresponding to the use of triweight kernels instead of Gaussian kernels (see Härdle (1991)) and T_τ denotes the number of observations kept after trimming.

There are 47 bandwidths involved in the whole estimation procedure, with 27 being used in the first step and 20 in the second step. Some bandwidths correspond to the continuous variables variables, while others to the discrete variables. The following tables summarize the values of the different bandwidths outlined above.

Table B.1. Bandwidths used in the triplet–case

First Step					
Continuous Variables		Discrete Variables			
h_{gd_0}	0.276	$h_{g_0n_0}$	0.624	$h_{G_0n_0}$	0.481
h_{gx_0}	0.272	$h_{g_0n_1}$	0.417	$h_{G_0n_1}$	0.321
h_{Gx_0}	0.209	$h_{g_0n_2}$	0.624	$h_{G_0n_2}$	0.481
h_{gd_1}	0.372	$h_{g_1n_0}$	0.826	$h_{G_1n_0}$	0.676
h_{gx_1}	0.382	$h_{g_1n_1}$	0.735	$h_{G_1n_1}$	0.601
h_{Gx_1}	0.313	$h_{g_1n_2}$	0.826	$h_{G_1n_2}$	0.676
h_{gd_2}	0.400	$h_{g_2n_0}$	0.894	$h_{G_2n_0}$	0.732
h_{gx_2}	0.394	$h_{g_2n_1}$	0.734	$h_{G_2n_1}$	0.600
h_{Gx_2}	0.323	$h_{g_2n_2}$	0.894	$h_{G_2n_2}$	0.732
Second Step					
Continuous Variables		Discrete Variables			
h_{f_0c}	0.246	$h_{f_0n_0}$	0.628		
h_{f_0x}	0.224	$h_{f_0n_1}$	0.426		
$h_{f_{1A}c}$	0.334	$h_{f_0n_2}$	0.628		
$h_{f_{1A}x}$	0.326	$h_{f_{1A}n_0}$	0.852		
$h_{f_{1B}c}$	0.334	$h_{f_{1A}n_1}$	0.979		
$h_{f_{1B}x}$	0.316	$h_{f_{1A}n_2}$	0.852		
h_{f_2c}	0.360	$h_{f_{1B}n_0}$	0.854		
h_{f_2x}	0.339	$h_{f_{1B}n_1}$	0.726		
		$h_{f_{1B}n_2}$	0.854		
		$h_{f_2n_0}$	0.932		
		$h_{f_2n_1}$	0.730		
		$h_{f_2n_2}$	0.932		

Table B.2. Bandwidths used in the pair–case

First Step					
Continuous Variables		Discrete Variables			
h_{gd_0}	0.276	$h_{g_0n_0}$	0.624	$h_{G_0n_0}$	0.481
h_{gx_0}	0.272	$h_{g_0n_1}$	0.417	$h_{G_0n_1}$	0.321
h_{Gx_0}	0.209	$h_{g_0n_2}$	0.624	$h_{G_0n_2}$	0.481
h_{gd_1}	0.369	$h_{g_1n_0}$	0.963	$h_{G_1n_0}$	0.791
h_{gx_1}	0.379	$h_{g_1n_1}$	0.441	$h_{G_1n_1}$	0.362
h_{Gx_1}	0.311	$h_{g_1n_2}$	0.963	$h_{G_1n_2}$	0.791
h_{gd_2}	0.396	$h_{g_2n_0}$	1.049	$h_{G_2n_0}$	0.856
h_{gx_2}	0.392	$h_{g_2n_1}$	0.539	$h_{G_2n_1}$	0.439
h_{Gx_2}	0.319	$h_{g_2n_2}$	1.049	$h_{G_2n_2}$	0.856
Second Step					
Continuous Variables		Discrete Variables			
h_{f_0c}	0.246	$h_{f_0n_0}$	0.628		
h_{f_0x}	0.225	$h_{f_0n_1}$	0.426		
$h_{f_{1A}c}$	0.332	$h_{f_0n_2}$	0.628		
$h_{f_{1A}x}$	0.324	$h_{f_{1A}n_0}$	0.973		
$h_{f_{1B}c}$	0.342	$h_{f_{1A}n_1}$	0.597		
$h_{f_{1B}x}$	0.316	$h_{f_{1A}n_2}$	0.973		
h_{f_2c}	0.353	$h_{f_{1B}n_0}$	0.978		
h_{f_2x}	0.336	$h_{f_{1B}n_1}$	0.449		
		$h_{f_{1B}n_2}$	0.978		
		$h_{f_2n_0}$	1.086		
		$h_{f_2n_1}$	0.548		
		$h_{f_2n_2}$	1.086		

Bibliography

Asker, J. (2008): "A Study of the Internal Organisation of a Bidding Cartel," *American Economic Reveiw, Forthcoming.*

Bajari, P. (1997): "The First Price Auction With Asymmetric Bidders: Theory and Applications," *University of Minnesota Ph.D. thesis.*

——— (2001): "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory*, 18, 187–205.

Bajari, P., and L. Ye (2003): "Deciding Between Competition and Collusion," *Review of Economics and Statistics*, 85, 971–989.

Baldwin, L., R. Marshall, and J. Richard (1997): "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy*, 4, 657–699.

Bierens, H. J. (1983): "Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions," *Journal of the American Statistical Association*, 78, 699–707.

Campo, S., E. Guerre, I. Perrigne, and Q. Vuong (2006): "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders," *PSU-Working Paper.*

Campo, S., I. Perrigne, and Q. Vuong (1998): "Asymmetry and Joint Bidding in OCS Wildcat Auctions," *Mimeo, University of Southern California.*

Donald, S., and H. Paarsch (1993): "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, 121–148.

Elyakime, B., J. Laffont, P. Loisel, and Q. Vuong (1994): "First-Price Sealed-Bid Auctions with Secret Reservation Prices," *Annales d'Economie et de Statistiques*, 34, 115–141.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall.

Fan, J., and J. Marron (1993): "Comments on 'Local regression: automatic kernel carpentry' by Hastie and Loader," *Statistical Science*, 8, 129–134.

Flambard, V., and I. Perrigne (2006): "Asymmetry In Procurement Auctions: Evidence From Snow Removal Contracts," *The Economic Journal*, 116, 10141036.

Graham, D., and R. Marshall (1987): "Collusive Bidder Behavior at Single–Object Second–Price and English Auctions," *Journal of Political Economy*, 95, 1217–1239.

Graham, D., R. Marshall, and J. Richard (1990): "Differential Payments within a Bidder Coalition and the Shapley Value," *American Economic Review*, 80, 493–510.

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

Härdle, W. (1991): *Smoothing Techniques with Implementation in S.* Springer-Verlag New York, Inc.

Hendricks, K., and R. Porter (1989): "Collusion in Auctions," *Annales d'Économie et de Statistique*, (15/16), 217–230.

Honoré, B., and E. Kryriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variable," *Econometrica*, 68, 839–874.

Horowitz, J. (2000): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

Ibragimov, I., and R. Has'minskii (1981): *Statistical Estimation. Asymptotic Theory.* Springer-Verlag New York, Inc.

Jofre-Bonet, M., and M. Pesendorfer (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489.

Kelman, S. (1990): *Procurement and Public Management: The Fear of Discretion and the Quality of Public Performance.* American Enterprise Institute Press.

KOROSTELEV, A., AND A. TSYBAKOV (1993): *Minimax Theory of Image Reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag New York, Inc.

KRISHNA, V. (2002): *Auction Theory*. Academic Press.

KRYRIAZIDOU, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

LAFFONT, J., H. OSSARD, AND Q. VUONG (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63, 953–980.

LAFFONT, J., AND Q. VUONG (1996): "Structural Analysis of Auction Data," *American Economic Review*, 86, 414–420.

LAVERGNE, P., AND Q. VUONG (1996): "Nonparametric Selection of Regressors: The Nonnested Case," *Econometrica*, 64, 207–219.

LEBRUN, B. (1996): "Existence of an Equilibrium in First Price Auctions," *Economic Theory*, 7, 421–443.

——— (1999): "First–Price Auction in the Asymmetric N Bidder Case," *International Economic Review*, 40, 125–142.

LEVIN, D., AND J. SMITH (1994): "Equilibrium in Auctions with Entry," *American Economic Review*, 84, 585–599.

LI, T., AND I. PERRIGNE (2003): "Timber Sale Auctions with Random Reserve Prices," *The Review of Economics and Statistics*, 85, 189–200.

LI, T., I. PERRIGNE, AND Q. VUONG (1999): "Structural Estimation of Affiliated Private Values with an Application to OCS Auctions," *Mimeo, University of Southern California*.

LI, T., AND Q. VUONG (1997): "Using all Bids in Parametric Estimation of First–Price Auctions," *Economic Letters*, 55, 321–325.

MANSKI, C. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 205–228.

MARSHALL, R., AND M. MEURER (2001): *The Economics of Auctions and Bidder Collusion*, Game Theory and Business Applications. Kluwer Academic Publishers, Norwell.

——— (2004): "Bidder Collusion and Antitrust Law: Refining the Analysis of Price Fixing to Account for the Special Features of Auction Markets," *Antitrust Law Journal*, 72(1), 83–118.

Marshall, R., M. Meurer, J. Richard, and W. Stromquist (1994): "Numerical Analysis of Asymmetric First Price Auctions," *Games and Economic Behaviour*, 7, 193–220.

Maskin, E., and J. Riley (2000a): "Asymmetric Auctions," *Review of Economic Studies*, 67, 413–438.

——— (2000b): "Equilibrium in Sealed High Bid Auctions," *Review of Economic Studies*, 67, 439–454.

——— (2003): "Uniqueness of Equilibrium in Sealed High–Bid Auctions," *Games and Economic Behavior*, 45, 395–409.

McAfee, R., and J. McMillan (1987): "Auction and Bidding," *Journal of Economic Literature*, 25, 699–738.

——— (1992): "Bidding Rings," *American Economic Review*, 82, 579–599.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

Newey, W., and D. McFadden (1994): *Large Sample Estimation and Hypothesis Testing*, vol. IV of *Handbook of Econometrics*. Amsterdam-North Holland.

Paarsch, H. (1992): "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191–215.

Paarsch, H., and H. Hong (2006): *An Introduction to the Structural Econometrics of Auction Data*. MIT Press.

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*, Themes in Modern Econometrics. Cambridge University Press.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Parzen, E. (1962): "On Estimation of a Probability Density and Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.

Perrigne, I., and Q. Vuong (1999): "Structural Econometrics of First-Price Auctions: A Survey of Methods," *Canadian Journal of Agricultural Economics*, 47, 203–223.

——— (2008): *Auctions: Empirics*. Palgrave McMillan, second edn.

Pesendorfer, M. (2000): "A Study of Collusion in First–Price Auctions," *Review of Economic Studies*, 67, 381–411.

Porter, R., and D. Zona (1993): "Detection of Bid–Rigging in Procurement Auctions," *Journal of Political Economy*, 101, 518–538.

——— (1999): "Ohio School Milk Markets: An Analysis of Bidding," *Rand Journal of Economics*, 30, 263–288.

Powell, J. (1994): *Estimation of Semiparametric Models*, vol. IV of *Handbook of Econometrics*. Amsterdam-North Holland.

Powell, J., J. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Riley, J., and W. Samuelson (1981): "Optimal Auctions," *American Economic Review*, 71, 381–392.

Robinson, M. (1985): "Collusion and the Choice of Auction," *Rand Journal of Economics*, 16, 141–145.

Robinson, P. (1988): "Root–N–Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Rosenblatt, M. (1956): "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 642–669.

Ruppert, D., and M. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *Annals of Statistics*, 22, 1346–1370.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

Stone, C. (1982): "Optimal Rates of Convergence for Nonparametric Regressions," *Annals of Statistics*, 10, 1040–1053.

Wilson, R. (1992): *Strategic analysis of auctions*, vol. I of *Handbook of Game Theory with Economic Applications*. New York-North Holland.

Vita

María Florencia Gabrielli

M. Florencia Gabrielli was born in Mendoza, Argentina on May 11, 1976. She received a B.A. degree from Universidad Nacional de Cuyo, Mendoza, Argentinta and an M.A. degree in Economics and Public Policy from Universidad Torcuato Di Tella, Buenos Aires, Argentina, in 1998 and 2000 respectively. She worked for the Central Bank of Argentina as a research economist in Buenos Aires, Argentina, during the period 2000-2004. She enrolled in the Ph.D. program in Economics at the Pennsylvania State University in August 2004.