NORTHWESTERN UNIVERSITY

Essays in Empirical Auctions and Partially Identified Econometric Models

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Xun Tang

EVANSTON, ILLINOIS

December 2008

© Copyright by Xun Tang 2008

All Rights Reserved

ABSTRACT

Essays in Empirical Auctions and Partially Identified Econometric Models

Xun Tang

Chapter 1: (Bounds on the Counterfactual Revenue Distributions in Auctions with Reserve Prices) In first-price auctions with interdependent bidder values, the distributions of private signals and values cannot be uniquely recovered from bids in Bayesian Nash equilibria. Non-identification invalidates structural analyses that rely on exact identification of the model primitives. In this paper I introduce tight, informative bounds on the distribution of revenues in counterfactual first- and second-price auctions with binding reserve prices. These robust bounds are identified from distributions of equilibrium bids in first-price auctions under minimal restrictions where I allow for affiliated signals and both private- and common-value paradigms. The bounds can be used to compare auction formats and to select optimal reserve prices. I propose consistent nonparametric estimators of the bounds. I extend the approach to account for observed heterogeneity across auctions, as well as endogenous participation due to binding reserve prices. I use a recent data of 6,721 first-price auctions of U.S. municipal bonds to estimate bounds on counterfactual revenue distributions. I then bound optimal reserve prices for sellers with various risk attitudes.

Chapter 2: (Semiparametric Estimation of Binary Response Models under Inequality Quantile Restrictions) In this paper I study the estimation of a class of binary response models where conditional medians of disturbances are bounded between known functions of regressors. This class of models incorporates several interesting micro-econometric sub-models with wide empirical applications. These include binary response with interval data on regressors, simultaneous discrete games with incomplete information, and Markovian binary choice processes. I characterize the identification region of linear coefficients in payoff functions, and give fairly general restrictions on the distribution of regressors that are sufficient for point identification. I also show how these restrictions are satisfied by primitive conditions in some of the motivating sub-models. I then define a two-step extreme estimator, and show it is consistent regardless of point identification, and converges to a normal distribution at the rate of $\sqrt{n}$ under point identification. This is possible because point identification can be attained even when the regressors have bounded supports. Monte Carlo evidence on the estimator's performance in finite samples when the model is partially identified is reported.

Chapter 3: (Identification of Dynamic Binary Choice Processes) In this paper, we study the identification of structural parameters in a class of dynamic binary choice processes where transitions to future states are independent from unobservable disturbances conditional on current actions and observable states. We give a full characterization of the set of single-period payoffs and disturbance distributions that generate the same choice probabilities as observed in a given process. We show with knowledge of the disturbance distribution, the differences in payoffs from two trivial policies of choosing the same action forever can be uniquely recovered from choice probabilities. Furthermore, we analyze the identifying power of various stochastic restrictions such as the statistical independence and conditional symmetry of the disturbance distributions. For models with finite spaces of observable states, we characterize the identification region of single-period payoffs under these restrictions by checking the feasibility of a system of linear equations in the nuisance parameters, subject to inequality constraints implied by observational equivalence and the restrictions imposed. This approach of identification through linear programming can be readily extended to cases where single-period payoffs are known to satisfy any form of restrictions.

Acknowledgements

I am deeply indebted to my wife Jie Liu and my parents, Yueqing Tang and Dunrong Liu, for their silent and everlasting support. I am grateful to my dissertation committee, Rosa Matzkin, Robert Porter and Elie Tamer, for their patient advice throughout the course of my work on this dissertation. I also benefit from discussions with Victor Aguirregabiria, Federico Bugni, John Chen, Xiaohong Chen, Benjamin Handel, Joel Horowitz, Yongbae Lee, Charles Manski, Aviv Nevo, Mallesh Pai, James Roberts, William Rogerson, Artyom Shneyerov, Kevin Song, Viktor Subbotin, Quang Vuong, Siyang Xiong, and participants of Northwestern's econometrics and industrial organization seminars. Financial support from the Center of Industrial Organizations is gratefully acknowledged.

Contents

LIST OF TABLES

LIST OF GRAPHS AND FIGURES

Chapter 1

Chapter 2

Chapter 3

CHAPTER 1

Bounds on the Counterfactual Revenue Distributions in Auctions with Reserve Prices

1.1. Introduction

In a structural auction model, a potential bidder does not know his own valuation of the auctioned object, but has some noisy private signal about its value. Bidders make their decisions conditional on these signals and their knowledge of the distribution of their competitors' private signals and values. A structural approach for empirical studies of auctions posits the distribution of bids observed can be rationalized by a joint distribution of bidder values and signals in Bayesian Nash equilibria, and defines this joint distribution as the model primitive. The objective is to extract information about this primitive from the distribution of bids, and to use it to answer policy questions such as the choice of optimal reserve prices or auction formats. (See Hendricks and Porter (2007) for a survey.) Depending on whether bidders would find rivals' signals informative about their own values conditional on their own signals, an auction belongs to one of the two mutually exclusive types : *private values* (PV), and *common values* (CV).[1] These two types have distinct implications for revenue distributions under a given auction format.

In this paper I propose tight, informative bounds on counterfactual revenue distributions that can be constructed from the distribution of bids in a general class of first-price auctions

[1] I use the term "interdependent values" for a larger class of auctions that encompass both PV and CV auctions. The formal definition of a PV auction is one in which bidders' values are mean independent from rival signals conditional on their own signals.

with interdependent values and affiliated signals. The counterfactual formats considered in this paper include both first- and second-price auctions with reserve prices.[2] Thus I introduce a unified approach of policy analyses for both PV and CV auctions that does not require exact identification of model primitives. My method is motivated by several empirical challenges related to structural CV models. First, several policy questions have not been addressed outside the restrictive case of PV auctions due to difficulties resulting from non-identification of signal and value distributions.[3] For a fixed reserve price, theory ranks expected revenue for general interdependent value auctions with affiliated signals, but the magnitude of expected revenue differences remains an empirical question.[4] Another open issue is the choice of optimal reserve prices in general interdependent value auctions with affiliated signals and finite number of bidders.[5] Since model primitives cannot be recovered from equilibrium bids in CV auctions, these questions cannot be addressed as in PV auctions, where point identification of signal distributions helps exactly recover revenue distributions in counterfactual formats.[6] Second, it is difficult to distinguish PV and CV auctions from the distribution of bids alone under a given auction format, even though the two have distinct implications in counterfactual revenue analyses. Laffont and Vuong (1996) proved for a given number of potential bidders, distributions of equilibrium bids in CV auctions can always be

[2]In this paper, I use the term "second-price auctions" exclusively for the sealed-bid format. This does not include the open formats, or "English auctions".

[3]For a proof of non-identification, see Laffont and Vuong (1996).

[4]The only exception is the case with i.i.d. signals, where expected revenue from first-price, second-price and English auctions are the same regardless of value interdependence.

[5]An exception is symmetric, independent private value auctions, where the optimal reserve price is identified from the distribution of equilibrium bids. Levin and Smith (1994) also showed in symmetric first-price auctions, where signals are affiliated and values are interdependent through a common unobserved component, the optimal reserve price converges to the seller's true value as the number of potential bidders n goes to infinity. Yet the theory is otherwise silent about identifying optimal reserve prices with a finite n.

[6]See Guerre et.al (2000), Li, Perrigne and Vuong (2002) and Li, Perrigne and Vuong (2003) for details. Also note in PV auctions, the distribution of signals $\{X_i\}_{i=1}^n$ are equivalent to the distribution of values $\{V_i\}_{i=1}^n$ under the normalization $E(V_i|X_i = x) = x$.

rationalized by certain PV structures. Empirical methods that have been proposed for distinguishing between the two types often have practical limitations. They either rely on assumptions that may not be valid (such as exogenous variations of number of bidders, as in Haile, Hong and Shum (2003)), or they may entail strong data requirements (an ex post measure of bidder values as in Hendricks, Pinkse and Porter (2003), or many bids near a binding reserve price as in Hendricks and Porter (2007)).[7] Third, the empirical auction literature has not considered the magnitude of the bias if a CV environment is analyzed with a PV model in counterfactual revenue analyses.

I propose a structural estimation method through partial identification of revenue distributions to address the questions above. First, the bounds on revenue distributions are constructed directly from the bids, and do not rely on pinpointing the underlying signal and value distributions. Second, the bounds only require a minimum set of general restrictions on value and signal distributions that encompass both PV and CV paradigms. Third, the bounds are tight and sharp within the general class of first-price auctions. The lower bound is the true counterfactual revenue distribution under a PV structure, while the upper bound can be close to the truth for certain types of CV models. Hence the distance between the bounds can be interpreted as a measure of maximum error possible when a CV structure is analyzed as PV in counterfactual analyses. The bounds can be nonparametrically estimated consistently. Although I do not provide point estimates of revenue distributions, the bounds are informative for answering policy questions, for they can be used to compare auction formats, or to bound revenue maximizing reserve prices. The analysis can be extended to risk-averse sellers immediately given the sellers' utility functions.

[7] A reserve price is *binding* if it is high enough to have a positive probability for screening bidders.

My paper is related to the literature on robust inference in auction models. Haile and Tamer (2003) use incomplete econometric models to bound the optimal reserve price in independent PV English auctions, where the equilibrium bidding assumption is replaced with two intuitive behavior assumptions. In contrast, my paper focuses on first-price CV auctions. Incompleteness here arises from the range of possible rationalizing signal and value distributions, instead of a flexible interpretation of bids. Hendricks, Pinkse and Porter (2003) introduce nonparametric structural analyses to CV auctions. They use an ex post measure of bidder values to test the assumption of equilibrium bidding. They also provide evidence that the winner's curse effect dominated the competition effect, leading to less aggressive bidding in equilibrium as the number of bidders increase. Shneyerov (2006) introduces an approach for counterfactual revenue analyses in common-value auctions without the need to identify model primitives. In particular, he shows that for any given reserve price, equilibrium bids from first-price auctions can be used to identify the expected revenues in second-price auctions with the same reserve price. He also shows how to bound the expected gains in revenues from English auctions under the general restriction of monotone value functions and affiliated signals.

My paper makes three novel contributions. First, the focus on revenue distributions, as opposed to distributional parameters such as expectations, allows more general revenue analyses. Auction theory usually uses expected revenue as a criterion to compare auction designs, but central tendency may not be justifiable in practice, say if the seller is not risk-neutral. Knowledge of distributions is necessary for other criteria, such as maximizing expected seller utility. (A seller may choose a design to maximize the probability that revenue falls in a certain range.) Second, bounds on revenue distributions can be constructed for hypothetical reserve prices. One can then compare reserve prices within first- or second-price formats. In CV auctions, a counterfactual, non-binding reserve price r creates serious challenges in policy analyses. The probability that no one bids higher than r in equilibrium can not be pinpointed from bids in the data, since the screening level can not be identified without further restrictions.[8] Moreover, the mapping between equilibrium strategies under r and those in the data cannot be uniquely recovered. I address this issue by bounding the bid that a marginal bidder under a counterfactual binding r actually places in equilibrium under the data-generating auction format.[9] These bounds in turn lead to bounds on the revenue distribution under r. Finally, the bounds on revenue distributions are robust and independent of the exact form of signal affiliations and value interdependence, and are identified from the distribution of equilibrium bids alone. This robustness comes with the price of partial identification of revenue distributions. Nonetheless, one can obtain informative answers for some policy questions.

The remainder of the paper proceeds as follows. Section 2 introduces bounds on counterfactual revenue distributions in a benchmark model where data is collected from homogenous auctions with exogenous participation. Section 3 defines nonparametric estimators for bounds and proves their pointwise consistency. Section 4 provides Monte Carlo evidence about the performance of the bound estimators. Section 5 extends the benchmark model to allow for observable auction heterogeneity and endogenous participation under binding reserve prices in the data. Section 6 applies the proposed method to U.S. municipal bond auctions on the primary market. Section 7 concludes.

[8] A screening level under r is the value of signal such that only bidders with signals higher than the screening level will choose to submit bids above r in equilibrium. See Section 2 below for a formal definition.
[9] A marginal bidder under r is the one whose signal is exactly equal to the screening level.

1.2. Bounds on Counterfactual Revenue Distributions in the Benchmark Model

This section focuses on a benchmark case where bids are observed from increasing, symmetric pure-strategy Bayesian Nash Equilibria ($PSBNE$) in homogenous, single-unit first-price auctions with a non-binding reserve price. I use distributions of these bids (denoted $G^0_{\mathbf{B}}$) to construct tight bounds on counterfactual revenue distributions for both first-price and second-price auctions with reserve price $r > 0$ (denoted $F_{R^I(r)}$ and $F_{R^{II}(r)}$ respectively). Extensions to cases where bids are observed from heterogenous auctions or auctions with endogenous participation due to binding reserve prices are discussed in Section 5.

1.2.1. Model specifications

Consider a single-unit first-price auction with n potential risk-neutral bidders and a non-binding reserve price. Each bidder receives a private signal X_i but cannot observe his own valuation V_i. The distribution of all bids submitted in equilibrium (denoted $\{B_i\}_{i=1,..,n}$) is observed from a random sample of independent, identical auctions, but neither X_i nor V_i can be observed. For simplicity, X_i and V_i are both scalars.[10] The following assumptions are maintained throughout the paper.

A1 (Symmetric, Affiliated Signals) Private signals $\mathbf{X} \equiv \{X_i\}_{i=1,..,n}$ are affiliated with support $[x_L, x_U]^n$, and the joint distribution $F_{\mathbf{X}}$ is exchangeable in all arguments.[11]

[10]Throughout the paper I use upper case letters to denote random variables and lower case letters for corresponding realized values.

[11]Let Z be a random vector in $\mathbb{R}^K$ with joint density f. Let $\vee$ and $\wedge$ denote respectively component-wise maximum and minimum of any two vectors in $\mathbb{R}^K$. Variables in Z are *affiliated* if, for all z and z' in $\mathbb{R}^K$, $f(z \vee z')f(z \wedge z') \geq f(z)f(z')$. For a more formal definition, see Milgrom and Weber (1982).

A2 (Interdependent Values) A bidder's valuation satisfies $V_i = \theta(X_i, \mathbf{X}_{-i})$, where $\theta(.)$ is a nonnegative, bounded, continuous function that is exchangeable in $\mathbf{X}_{-i} \equiv \{X_1, .., X_{i-1}, X_{i+1}, ...X_n\}$, non-decreasing in all signals, and increasing in own signal X_i over $[x_L, x_U]$.[12]

Note *A2* implies private signals are drawn from identical marginal distributions on $[x_L, x_U]$, and *A1* includes private values (PV) as a special case, where $\theta(x_i, \mathbf{x}_{-i}) = \theta(x_i)$ $\forall (x_i, \mathbf{x}_{-i}) \in [x_L, x_U]^n$. Common values (CV) correspond to value functions that are non-degenerate in rival signals $\mathbf{X}_{-i}$. A pure strategy for a bidder is a function $b_i : X_i \rightarrow \mathbb{R}^1_+$ and a pure-strategy Bayesian Nash Equilibria is a portfolio $\{b_{0,i}(.)\}_{i=1,..,n}$ such that $\forall\ i, b_{0,i}(.)$ is the best response to $\{b_{0,j}(.)\}_{j \in \{1,..,n\}\backslash\{i\}}$, i.e. $\forall i$,

$$b_{0,i}(x) = \arg\max_{b_{0,i}} E(V_i - b_{0,i} | \max_{j \neq i} b_{0,j}(X_j) \leq b_{0,i}, X_i = x) \Pr(\max_{j \neq i} b_{0,j}(X_j) \leq b_{0,i} | X_i = x)$$

The regularity conditions for existence of such a $PSBNE$ is collected in *A3* below. These restrictions are otherwise inessential for the partial identification result in this paper.

A3 (Regularity Conditions) (i) $\theta(.)$ is twice continuously differentiable; *(ii)* The joint density of $\{X_i\}_{i=1,..,n}$ exists on $[x_L, x_U]^n$, is continuously differentiable, and $\exists f_{low}, f_{high} > 0$ such that $f(\mathbf{x}) \in [f_{low}, f_{high}]$ $\forall \mathbf{x} \in [x_L, x_U]^n$.

McAdams (2006) proved *A1, A2,* and *A3* are sufficient for the existence of unique symmetric, increasing $PSBNE$ in first-price auctions.

Definition 1. *A joint distribution of bids* $\{b_{0,i}\}_{i \in N}$ *in first-price auctions with zero reserve prices (denoted* $G^0_{\mathbf{B}}$*) is* ***rationalized*** *by an auction defined by the structure* $\{\theta, F_{\mathbf{X}}\}$*) if* $G^0_{\mathbf{B}}$ *is the distribution of bids in a symmetric, increasing PSBNE in this auction. Two*

[12]Throughout the paper I use bold letters for vectors of random variables or functions (e.g., $\mathbf{X} = (X_1, X_2, ..., X_n)$ and $\boldsymbol{\theta}(\mathbf{X}) = (\theta(X_1, X_{-1}), \theta(X_2, X_{-2}), ..., \theta(X_n, X_{-n}))$).

structures $\{\theta, F_{\mathbf{X}}\}$ *and* $\{\tilde{\theta}, \tilde{F}_{\mathbf{X}}\}$ *are* ***observationally equivalent*** *if they generate the same distribution* $G^0_{\mathbf{B}}$ *in PSBNE of first-price auctions.*

Let $b_{0,n}(.)$ denote the equilibrium bidding strategy in a first-price auction with zero reserve price and $N = n$ potential bidders. The first-order condition of such a *PSBNE* is characterized by the first order condition:

$$b'_{0,n}(x) = [v_{h,n}(x,x) - b_{0,n}(x)]\frac{f_{Y|X;n}(x|x)}{F_{Y|X;n}(x|x)} \tag{1.1}$$

for all $x \in [x_L, x_U]$, where $Y_i \equiv \max_{j\neq i} X_j$, $v_{h,n}(x,y) \equiv E(V_i|X_i = x, Y_i = y, N = n)$, $F_{Y|X;n}(t|x) \equiv \Pr(Y_i \leq t|X_i = x, N = n)$, and $f_{Y|X;n}(t|x)$ denotes the corresponding conditional density. The equilibrium boundary condition is $b_{0,n}(x_L) = v_{h,n}(x_L, x_L)$. Subscripts for bidder indices are dropped in $v_{h,n}, F_{Y|X}$ and $f_{Y|X}$ due to the symmetry in $F_{\mathbf{X}}$ and θ. In an increasing *PSBNE* where $b'_{0,n}(x) > 0\ \forall x \in [x_L, x_U]$, Guerre et.al (2000) showed a link between the primitives and $G^0_{\mathbf{B}}$ by manipulating (1.1) using change-of-variable :

$$v_{h,n}(x,x) = b_{0,n}(x) + G^0_{M|B;n}(b_{0,n}(x)|b_{0,n}(x))/g^0_{M|B;n}(b_{0,n}(x)|b_{0,n}(x)) \equiv \xi(b_{0,n}(x); G^0_{\mathbf{B}}) \tag{1.2}$$

where $G^0_{M|B;n}(\tilde{b}|b) = \Pr(\max_{j\neq i} b_{0,n}(X_j) \leq \tilde{b}|b_{0,n}(X_i) = b)$ and $g^0_{M|B;n}(\tilde{b}|b)$ is the corresponding density.[13] Bidder indices on M and B are dropped due to symmetry of $F_{\mathbf{X}}$ and θ.

1.2.2. Review of literature on PV auctions

In this subsection, I review the literature on identification of signal distributions and optimal reservation prices in private value auctions. The objective is to highlight how unique identification of the bidder signal distributions and screening levels leads to exact identification

[13] Following convention in the literature, I assume the second order conditions are always satisfied and thus first-order conditions are sufficient for characterizing the equilibrium.

of the optimal reserve price, and to motivate my incomplete approach when the screening level can not be point identified in the more general case of interdependent values.

Guerre, Perrigne and Vuong (2000) and Li, Perrigne and Vuong (2002) showed the joint distribution of bidder values are nonparametrically identified from distribution of equilibrium bids in first-price, PV auctions with no reserve prices, regardless of the form of dependence between private signals. The main idea is that in private value auctions, the left-hand side of (1.2) can be normalized to the signal x itself, and thus the inverse bidding function is recovered from $G^0_{\mathbf{B}}$, for both independent and affiliated signals.[14] Another simplification peculiar to PV auctions is that the screening level under a binding reserve price r is equal to r itself. That is, bidders choose not to bid above r in equilibrium if and only if their private signals are below r. To see this, note the screening level under r in a general interdependent value auction is defined as :

$$x^*(r) \equiv \inf\{x \in [x_L, x_U] : E(V_i|X_i = x, \max_{j\neq i} X_j \leq x) \geq r\}$$

In PV auctions, $E(V_i|X_i = x, \max_{j\neq i} X_j \leq x) = E(V_i|X_i = x)$ and the normalization $E(V_i|X_i = x) = x$ implies $x^*(r) = r$. Thus in private value auctions, both signal distribution $F_{\mathbf{X}}$ and $x^*(r)$ are exactly recovered from $G^0_{\mathbf{B}}$.

In principle, knowledge of $F_{\mathbf{X}}$ in private value auctions is sufficient for finding counterfactual revenue distributions under a binding reserve price r. It follows that the optimal r which maximizes expected revenue is also identified. Yet in reality it can be impractical to implement this fully nonparametric estimation due to data deficiencies, especially when the signals are affiliated. Li, Perrigne and Vuong (2003) proposed a nonparametric algorithm for estimating optimal reserve prices that is implemented with less intensive computations. The

[14]In private value auctions, the conventional normalization of the signals is $E(V_i|X_i = x) = x$ for all x.

idea is to express expected seller revenue under r as a functional of r itself and the observed distribution of equilibrium bids. Then optimizing a sample analog of this objective function over reserve prices gives a consistent estimator of the optimal reservation price. Again the assumption of private values is indispensable for two reasons. First, it implies $x^*(r) = r$ under appropriate normalizations, which is used for defining the objective function; Second, it ensures full nonparametric identification of the distributions of counterfactual equilibrium bidding strategies.

This approach can not be applied to CV auctions with affiliated signals immediately because of two non-identification results. First, the screening level cannot be pinned down without further restrictions on how bidders' signals and valuations are correlated. Second, inverse bidding functions can not be recovered without knowledge of θ. Hence underlying structure $\{\theta, F_{\mathbf{X}}\}$ can not be identified. These pose a major challenge for identifying counterfactual revenue distributions in CV auctions.

1.2.3. Observational equivalence of PV and CV

In this subsection I prove the observational equivalence of PV and CV paradigms when only $G^0_{\mathbf{B}}$ is observed. In other words, any $G^0_{\mathbf{B}}$ is rationalized by a PV paradigm *if and only if* it is also rationalized by a CV paradigm. Laffont and Vuong (1996) already proved the sufficiency in their well-known argument for the non-identification of CV auctions. Below I complete the proof of observational equivalence by showing the converse (necessity).

Besides being a main motivation for the bounds in this paper, observational equivalence of PV and CV auctions also has an important implication about the efficiency of the bounds. That is, they effectively exhaust all information that can be extracted from $G^0_{\mathbf{B}}$ and equilibrium conditions alone for counterfactual revenue analyses. To see this, note the true counterfactual revenue distribution is exactly equal to the lower bound if G^0_B is rationalized by a PV structure. On the other hand, it can lie anywhere within the open interval between the bounds for certain types of CV auctions, depending on how a bidder weighs his own signals while calculating his expected value conditional on winning. Therefore, observational equivalence of PV and CV auctions given a rationalizable distribution $G^0_{\mathbf{B}}$ implies that the possible range of counterfactual revenue distributions can not be reduced to any strict subsets of the interval between the bounds.

Let $\mathcal{F}$ denote the set of joint signal distributions that satisfy *A1*, and Θ denote the set of value functions that satisfy *A2*. Let Θ_{CV} denote a subset of Θ that is non-degenerate in rival signals $\mathbf{X}_{-i}$. The following proposition gives *necessary and sufficient* conditions for $G^0_{\mathbf{B}}$ to be rationalized by some element of $\Theta_{CV} \otimes \mathcal{F}$.

Proposition 2. *A joint distribution of bids $G^0_{\mathbf{B}}$ observed in first-price auctions with non-binding reserve prices can be rationalized by some $\{\theta, F_{\mathbf{X}}\} \in \Theta_{CV} \otimes \mathcal{F}$ if and only if (i) $G^0_{\mathbf{B}}$ is affiliated and exchangeable in all arguments; and (ii) $\xi(b; G^0_{\mathbf{B}}) = b + G^0_{M_i|B_i}(b|b)/g^0_{M_i|B_i}(b|b)$ is strictly increasing on the support of the marginal distribution of bids $[b_L, b_U]$.*

The intuition of the proof is as follows.[15] Let Θ_P be a subset of Θ that only depends on bidders' own signals. Laffont and Vuong (1996) proved if $G^0_{\mathbf{B}}$ is rationalized by some $\{\theta, F_{\mathbf{X}}\} \in \Theta \otimes \mathcal{F}$, then it must also be rationalizable by some $\{\tilde{\theta}, \tilde{F}_{\mathbf{X}}\} \in \Theta_P \otimes \mathcal{F}$. Li, Perrigne and Vuong (2002) showed *(i)* and *(ii)* in Proposition 1 are necessary conditions for $G^0_{\mathbf{B}}$ to be rationalized by some $\{\tilde{\theta}, \tilde{F}_{\mathbf{X}}\} \in \Theta_P \otimes \mathcal{F}$. A combined argument proves the necessity. Sufficiency is proven by constructing examples how any $G^0_{\mathbf{B}}$ satisfying *(i)* and *(ii)* can be rationalized by a certain type of CV auctions where bidders' values only depend on his own

[15]Formal proofs for all lemmae and propositions in this paper are included in the Appendix.

and the highest rival signals. Li, Perrigne and Vuong (2002) showed the conditions in Proposition 1 are also necessary and sufficient for $G^0_{\mathbf{B}}$ to be rationalized by PV auctions, and it follows that a $G^0_{\mathbf{B}}$ is rationalized by some $\theta \in \Theta_P$ *if and only if* it is also rationalized by some $\theta \in \Theta_{CV}$. A corollary of Proposition 1 is that any private value auction is observationally equivalent to a certain CV auction.

Note the observational equivalence of PV and CV auctions above is restricted in this benchmark environment under a non-binding reserve price and a fixed number of bidders. Several recent works have showed ways to derive different testable implications of the two paradigms with augmented data. These include exogenous variations in the number of bidders as in Haile, Hong and Shum (2003), ex post measures of bidder values as in Hendricks, Pinkse and Porter (2003), and bid distributions under a strictly binding reserve price as in Hendricks and Porter (2007). In such case, the lower bound point identifies the real revenue distribution for the PV paradigm, while the open interval between the bounds are efficient in the sense that they are tight and sharp within the class of CV auctions.

1.2.4. Bounds on $F_{R^I(r)}$

The conventional criterion for choosing optimal reserve prices is the expected revenue for the seller. The Revenue Equivalence Theorem states that in auctions with independent private values, optimal reserve prices are the same for both 2nd- and 1st-price auctions, and are independent from the number of potential bidders. On the other hand, there is no theoretical result about the choice of optimal reserve prices in general 1st-price auctions with affiliated signals and finite number of bidders. The answer depends on the specifics of model primitives and is left open for empirical analyses. Besides, the criteria of expected revenue itself can be hard to justify if the seller is not risk neutral. Knowledge of counterfactual revenue distributions will be useful for addressing both concerns. For a binding reserve price r, I propose informative bounds on $F_{R^I(r)}$ that can be constructed from $G^0_{\mathbf{B}}$ (the distribution of equilibrium bids in first-price auctions with a non-binding reserve price).

1.2.4.1. Relations between $G^0_{\mathbf{B}}$ and $F_{R^I(r)}$. The equilibrium strategy in first-price auctions under a reserve price $r \geq 0$ has a closed form:

$$\begin{aligned} b_r(x;\theta,F_{\mathbf{X}}) &= rL(x^*(r)|x;F_{\mathbf{X}}) + \int_{x^*(r)}^{x} v_h(s,s;\theta,F_{\mathbf{X}})dL(s|x;F_{\mathbf{X}}) \;\; \forall x \geq x^*(r) \\ b_r(x;\theta,F_{\mathbf{X}}) &< r \;\; \forall x < x^*(r) \end{aligned}$$

where $L(s|x;F_{\mathbf{X}}) \equiv \exp\{-\int_s^x \Lambda(u;F_{\mathbf{X}})du\}$ and $\Lambda(x;F_{\mathbf{X}}) \equiv f_{Y|X}(x|x)/F_{Y|X}(x|x)$.[16] For any given x on the closed interval $[x_L, x_U]$, $L(s|x;F_{\mathbf{X}})$ is a well-defined distribution function with support $[x_L, x]$ and is first-order stochastically dominated by the distribution of the second highest signal (i.e. $F_{Y|X}(s|x)/F_{Y|X}(x|x)$). When signals are i.i.d., the two distributions are identical. [17]

The range of r for nontrivial counterfactual analyses is $[\xi_L, \xi_U]$, where $\xi_k \equiv \xi(x_k; G^0_{\mathbf{B}})$ for $k = L, U$. For $r < \xi_L$, $x^*(r) = x_L$ and there is no effective screening of bidders. For $r > \xi_U$, $x^*(r) = x_U$ and all bidders are screened out with probability one. Let v_0 denote the seller's own reserve value of the auctioned object. For all $r > v_0$, the distribution of revenue

[16] This section focuses on a benchmark model with fixed n. Hence the superscript n is suppressed for notational ease.

[17] That $L(s|x)$ is a well-defined distribution on $[x_L, x]$ is shown in Krishna (2002). Furthermore $L(s|x) \geq \exp\{-\int_s^x \frac{f_{Y|X}(u|x)}{F_{Y|X}(u|x)}du\} = \frac{F_{Y|X}(s|x)}{F_{Y|X}(x|x)}$, where the inequality follows from the fact that $F_{Y|X}(x|z)/f_{Y|X}(x|z)$ is decreasing in z when signals are affiliated. The equality holds when signals are i.i.d.

in first-price auction with r is [18]:

$$
\begin{aligned}
F_{R^I(r)}(t) &= 0 \quad \forall t < v_0 \\
&= \Pr(X^{(1)} < x^*(r)) \quad \forall t \in [v_0, r) \\
&= \Pr(X^{(1)} \leq b_r^{-1}(t)) \quad \forall t \in [r, +\infty)
\end{aligned}
$$

where $X^{(k)}$ denotes the k-th highest order statistic out of n signals. As b_0 and b_r are both monotone in signals above the screening level $x^*(r)$, there exists a one-to-one mapping between them for $x \geq x^*(r)$. Counterfactual revenue distribution $F_{R^I(r)}$ would be exactly identified from $G^0_{\mathbf{B}}$ if the mapping between b_0 and b_r for $x \geq x^*(r)$ can be recovered from $G^0_{\mathbf{B}}$. The following lemma gives a closed form for such a mapping which is also a functional of $G^0_{\mathbf{B}}$.

Lemma 3. *In first-price auctions, for all $r > 0$ and $x \geq x^*(r)$,*

$$
b_r(x) = \delta_r(b_0(x); G^0_{\mathbf{B}}) \equiv r\tilde{L}(b_0(x^*(r))|b_0(x); G^0_{\mathbf{B}}) + \int_{b_0(x^*(r))}^{b_0(x)} \xi(\tilde{b}; G^0_{\mathbf{B}}) d\tilde{L}(\tilde{b}|b_0(x); G^0_{\mathbf{B}})
$$

where $\tilde{L}(b|b'; G^0_{\mathbf{B}}) \equiv \exp\left(-\int_b^{b'} \tilde{\Lambda}(u; G^0_{\mathbf{B}})du\right)$ *and* $\tilde{\Lambda}(u; G^0_{\mathbf{B}}) \equiv \frac{g^0_{M|B}(u|u)}{G^0_{M|B}(u|u)}$.

By construction, δ_r is increasing in b for $b \geq b_0(x^*(r))$. Despite its closed form, δ_r can not be exactly recovered from $G^0_{\mathbf{B}}$ as $b_0(x^*(r))$ is not point identified only under *A1* and *A2*. The line of reasoning for bounds on $F_{R^I(r)}$ constructed from $G^0_{\mathbf{B}}$ is as follows. First, value interdependence and signal affiliations imply tight bounds on $b_0(x^*(r))$ that can be identified from $G^0_{\mathbf{B}}$. Second, these bounds on $b_0(x^*(r))$ lead to envelops on the mapping δ_r

[18] *Proof of this claim:* By definition of v_0, $\Pr(R(r) \leq t) = 0$ for all $t < v_0$. For all $t \in [v_0, r)$, $\Pr(R(r) \leq t) = \Pr(R(r) = V_0) = \Pr(X^{(1)} < x^*(r))$. Note $b_r'(x) > 0$ $\forall (r, x)$ such that $r \geq 0$ and $x > x^*(r)$. Since $b_r(x^*(r)) = r$, $b_r(x)$ is invertible on $[r, +\infty)$. Then for $t \geq r$, $\Pr(R(r) \leq t) = \Pr(R(r) < r) + \Pr(r \leq b_r(X^{(1)}) \leq t) = \Pr(X^{(1)} < x^*(r)) + \Pr(x^*(r) \leq X^{(1)} \leq b_r^{-1}(t)) = \Pr(X^{(1)} \leq b_r^{-1}(t))$ for all $t \in [r, +\infty)$.

that can also be constructed from $G^0_{\mathbf{B}}$. Next, inverting these envelops at revenue level t gives bounds on $b_0(b_r^{-1}(t))$. Finally, evaluating the distribution of winning bids at these inverses gives bounds on $F_{R^I(r)}(t)$.

1.2.4.2. Bounds on the screening level $x^*(r)$ and $b_0(x^*(r))$. Let $v(x,y) \equiv E(V_i|X_i = x, Y_i \leq y)$, and $v_l(x,y) \equiv \int_{x_L}^{y} v_h(s,s)\frac{f_{Y|X}(s|x)}{F_{Y|X}(y|x)}ds = E(v_h(Y,Y)|X_i = x, Y_i \leq y)$. (For the rest of the paper I will use $v(x)$ and $v_k(x)$ as shorthand notations for $v(x,x)$ and $v_k(x,x)$ for $k = l,h$.) In symmetric equilibria, $v(x)$ denotes a bidder's expected value conditional on winning with signal x in both 1st-price and 2nd-price auctions, and $v_l(x)$ denotes a bidder's expected payment conditional on winning in a 2nd-price auction with a non-binding reserve price. Affiliated signals and interdependent values implies $v_h(x) \geq v(x)$ for all x, and the equilibrium condition in second-price auctions guarantees $v(x) \geq v_l(x)$ for all x. For all $r > 0$, the screening level $x^*(r)$ is defined as the inverse of $v(x)$ at r. Hence inverting $v_h(x)$ and $v_l(x)$ at r gives bounds on $x^*(r)$. The following lemma formalizes this idea.

Lemma 4. *(i) For all $(\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$ (satisfying A1,A2), $v_l(x,y) \leq v(x,y) \leq v_h(x,y)$ for all $x_L \leq y \leq x \leq x_U$, and both $v_l(x)$ and $v_h(x)$ are increasing in x on $[x_L, x_U]$. (ii) For $r \in [v_{h,L}, v_{h,U}]$ where $v_{h,k} \equiv v_h(x_k)$, define $x_l(r) \equiv \arg\min_{x\in[x_L,x_U]}[v_h(x) - r]^2$ and $x_h(r) \equiv \arg\min_{x\in[x_L,x_U]}[v_l(x) - r]^2$. Then for all $(\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$, $x_l(r) \leq x^*(r) \leq x_h(r)$ for all r in the range above.*

The screening level $x^*(r)$ can not fall outside this bound, provided bidders' values are non-decreasing in both own and competitors' signals, and private signals are affiliated. Another desirable property of this bound on $x^*(r)$ is its tightness, in the sense that it has exhausted all information in the restricted set $\Theta \otimes \mathcal{F}$.

Lemma 5. *(i) $\exists\ \psi \equiv (\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$ such that $x_l(r;\psi) = x^*(r;\psi)$ for all r; and (ii) $\forall \varepsilon > 0$, $\exists\ \psi \in \Theta \otimes \mathcal{F}$ such that $\sup_{r\in[v_{h,L},v_{h,U}]} |x_h(r;\psi) - x^*(r;\psi)| \leq \varepsilon$.*

The upper bound is reached if, conditional on winning in a first-price auction, a bidder finds his rivals' signals completely uninformative about his own value. This includes *PV* auctions as a special case. On the other hand, the lower bound is attained if the margin between a bidder's own signal and the highest competing signal reveals no additional information about his own value conditional on winning. In other words, lower and upper bounds of screening levels correspond to two extreme cases of weights (0 and 1 respectively) that a bidder assigns to his own signals while calculating his expected value conditional on winning.

Both v_l and v_h are related to $G^0_{\mathbf{B}}$ through equilibrium bidding condition in first-price auctions. The non-negativity of θ suggests $x^*(0) = x_L$.[19] Hence $\forall x \geq x_L$, $v_h(x) = \xi(b_0(x); G^0_{\mathbf{B}})$ and

$$v_l(x) = \xi_l(b_0(x); G^0_{\mathbf{B}}) \equiv \int_{b_0(x_L)}^{b_0(x)} \xi(\tilde{b}; G^0_{\mathbf{B}}) \frac{g^0_{M|B}(\tilde{b}|b_0(x))}{G^0_{M|B}(b_0(x)|b_0(x))} d\tilde{b}$$

It follows from *Lemma 2* above that $\xi(b; G^0_{\mathbf{B}}) \geq \xi_l(b; G^0_{\mathbf{B}})$ for all $b \in [b^0_L, b^0_U]$ (where $b^0_k \equiv b_0(x_k)$ for $k = L, U$), and $\xi(b^0_L; G^0_{\mathbf{B}}) = \xi_l(b^0_L; G^0_{\mathbf{B}})$. Furthermore, both $\xi(b; G^0_{\mathbf{B}})$ and $\xi_l(b; G^0_{\mathbf{B}})$ are increasing over $[b^0_L, b^0_U]$ by the monotonicity of $b_0(.)$. Define $\xi^0_k \equiv \xi(b_0(x_k); G^0_{\mathbf{B}}) = v_{h,k}$ for $k = L, U$. For $r \in [\xi^0_L, \xi^0_U]$, define $b_0(x_l(r)) = \arg\min_{b\in[b^0_L,b^0_U]}[\xi(b; G^0_{\mathbf{B}}) - r]^2$ and $b_0(x_h(r)) = \arg\min_{b\in[b^0_L,b^0_U]}[\xi_l(b; G^0_{\mathbf{B}}) - r]^2$. Then $b_0(x^*(r))$ is bounded between $b_0(x_l(r))$ and $b_0(x_h(r))$ for all $r \in [\xi_L, \xi_U]$. Note that $\{x_k(r)\}_{k=l,h}$ are tight bounds on $x^*(r)$ implies $\{b_0(x_k(r))\}_{k=l,h}$ are tight bounds on $b_0(x^*(r))$, as θ and v_h are bounded.

[19] The non-negativity of θ is testable in equilibrium, for $\xi_L \equiv \xi(b_L; G^0_{\mathbf{B}}) = v_h(x_L, x_L; \theta, F_{\mathbf{X}}) = \theta(\mathbf{x}_L)$.

1.2.4.3. Envelops on the δ_r-mapping and $F_{R^I(r)}$. The δ_r-mapping in *Lemma 1* turns out to be a solution for a differential equation :

$$\delta_r'(b;G^0_{\mathbf{B}}) = [\xi(b;G^0_{\mathbf{B}}) - \delta_r(b;G^0_{\mathbf{B}})]\tilde{\Lambda}(b;G^0_{\mathbf{B}})$$

with the boundary condition : $\delta_r(b_0(x^*(r))) = b_r(x^*(r)) = r$. The lemma below shows that replacing $b_0(x^*(r))$ with $\{b_0(x_k(r))\}_{k=l,h}$ in boundary conditions will lead to new solutions $\{\delta_{r,k}(.;G^0_{\mathbf{B}})\}_{k=l,h}$ respectively, which can be constructed from $G^0_{\mathbf{B}}$ and are envelops of δ_r from above for all $x \geq x^*(r)$ and from below for all $x \geq x_h(r)$.

Lemma 6. *For $k \in \{l,h\}$, define for $b \geq b_0(x_k(r))$,*

$$\delta_{r,k}(b;G^0_{\mathbf{B}}) \equiv r\tilde{L}(b_0(x_k(r))|b;G^0_{\mathbf{B}}) + \int_{b_0(x_k(r))}^{b} \xi(\tilde{b};G^0_{\mathbf{B}})d\tilde{L}(\tilde{b}|b;G^0_{\mathbf{B}})$$

Under A1-A3, $\delta_{r,k}(.;G^0_{\mathbf{B}})$ are increasing on $[b_0(x_k(r)), b_U]$ for $k = l,h$, and $\delta_{r,h}(b;G^0_{\mathbf{B}}) \leq \delta_r(b;G^0_{\mathbf{B}})$ $\forall b \geq b_0(x_h(r))$ and $\delta_{r,l}(b;G^0_{\mathbf{B}}) \geq \delta_r(b;G^0_{\mathbf{B}})$ $\forall b \geq b_0(x^*(r))$.

Intuitively, for bidders with signals above the screening level, $b_r(x)$ is the expectation of a function $h(t)$ with respect to the distribution $L(t|x)$, where $h(t)$ is defined as r for $t < x^*(r)$ and $v_h(t,t)$ for $t \in [x^*(r), x]$. By the definition of $x_l(r)$ and the monotonicity of v_h, $b_r(x)$ is smaller than the expectation of $h_u(t)$ conditional on $t \leq x$ for $x \geq x^*(r)$, where $h_u(t)$ is defined as r for $t < x_l(r)$ and $v_h(t,t)$ for $x \in [x_l(r), x]$. Likewise $b_r(x)$ for $x \geq x_h(r)$ is greater than the expectation of $h_l(t)$, where $h_l(t)$ is defined as r for $t < x_h(r)$ and $v_h(t,t)$ for $x \in [x_h(r), x]$. The lemma proves a version of these inequalities, with structural elements h_l, h_u, $L(s|x)$ and $x_k(r)$ replaced by corresponding functionals of $G^0_{\mathbf{B}}$ and $b_0(x)$ through the manipulation of the equilibrium condition in (1.2).

The fact that $L(s|x)$ is stochastically increasing in x has important implications on the performance of $\delta_{r,k}$. Specifically, their differences $\delta_{r,l} - \delta_{r,h}$ is non-increasing in b for $b \geq b_0(x_h(r))$. [20] This is a desirable property, for it implies the difference between $\delta_{r,l}^{-1}(t; G_{\mathbf{B}}^0)$ and $\delta_{r,h}^{-1}(t; G_{\mathbf{B}}^0)$ is decreasing in the revenue level t as long as both $\delta_{r,l}$ and $\delta_{r,h}$ increase at a moderate rate.

Given the lemma above and the identification of $b_0(x_k(r))$, the bounds on $F_{R^I(r)}$ are derived immediately.

Proposition 7. *Suppose $r > v_0$. Under A1, A2 and A3, $F_{R^I(r)}^l \succeq_{F.S.D.} F_{R^I(r)} \succeq_{F.S.D.} F_{R^I(r)}^u$, where $\succeq_{F.S.D.}$ denotes first-order stochastic dominance, and*

$$
\begin{aligned}
F_{R^I(r)}^l(t) &= 0 \quad \forall t < v_0 \\
&= \Pr(b_0(X^{(1)}) < b_0(x_l(r)) \quad \forall t \in [v_0, r) \\
&= \Pr(b_0(X^{(1)}) \leq \delta_{r,l}^{-1}(t; G_{\mathbf{B}}^0)) \quad \forall t \in [r, +\infty)
\end{aligned}
$$

and

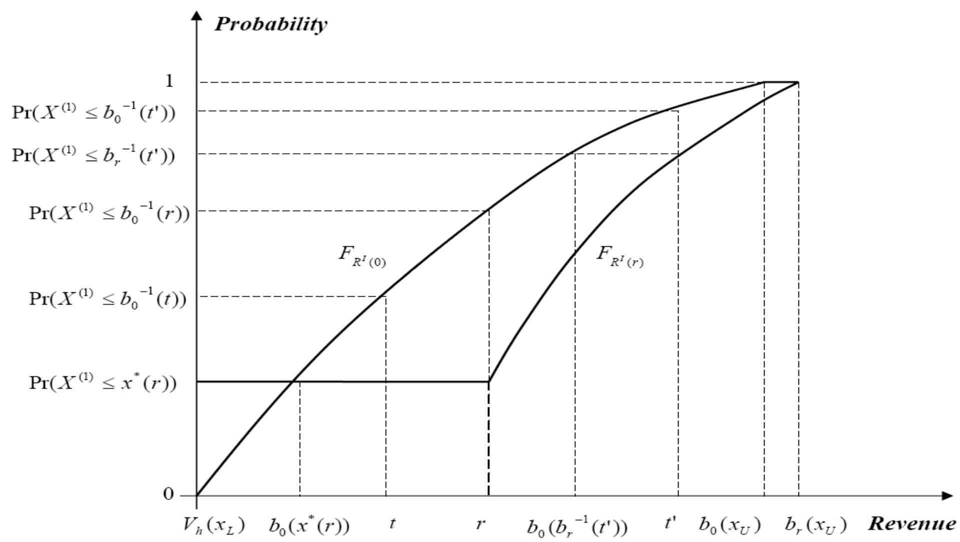
$$
\begin{aligned}
F_{R^I(r)}^u(t) &= 0 \quad \forall t < v_0 \\
&= \Pr(b_0(X^{(1)}) < b_0(x_h(r))) \quad \forall t \in [v_0, r) \\
&= \Pr(b_0(X^{(1)}) \leq \delta_{r,h}^{-1}(t; G_{\mathbf{B}}^0)) \quad \forall t \in [r, +\infty)
\end{aligned}
$$

where $\delta_{r,k}^{-1}(t; G_{\mathbf{B}}^0) \equiv \arg\min_{b \in [b_0(x_k(r)), b_U]} [\delta_{r,k}(b; G_{\mathbf{B}}^0) - t]^2$.

[20]Proof of the claim : $\delta'_{r,l}(b; G_{\mathbf{B}}^0) - \delta'_{r,h}(b; G_{\mathbf{B}}^0) = \tilde{\Lambda}(b; G_{\mathbf{B}}^0) \left[\int_{b_0(x_l(r))}^{b_0(x_h(r))} r - \xi(\tilde{b}; G_{\mathbf{B}}^0) d\tilde{L}(\tilde{b}|b; G_{\mathbf{B}}^0) \right] \leq 0$ for $r \leq \xi(b; G_{\mathbf{B}}^0)$ $\forall b \geq b_0(x_h(r))$ in equilibrium. The inequality is strict if $b_0(x_h(r)) > b_0(x_l(r))$.

1.2.4.4. A simpler upper bound of $F_{R^I(r)}$. Below I propose a simpler upper bound on $F_{R^I(r)}$ (denoted $\tilde{F}^u_{R^I(r)}$) that is constructed from $F_{R^I(0)}$ directly, rather than from $G^0_{\mathbf{B}}$. This simpler upper bound coincides with $F^u_{R^I(r)}$ when private signals are i.i.d., but is first-order stochastically dominated by $F^u_{R^I(r)}$ otherwise. The following lemma relates b_r to equilibrium strategies with no reserve price, and also helps relate $F_{R^I(r)}$ to $F_{R^I(0)}$ later.

Lemma 8. *Under A1-A2 and for $r > 0$, equilibrium strategies b_0 and b_r satisfy: (i) $b_0(x) \leq b_r(x)$ $\forall x \geq x^*(r)$ and (ii) the difference $b_r(x) - b_0(x)$ is decreasing in x $\forall x \geq x^*(r)$.*



Graph 1

Graph 1 depicts $F_{R^I(0)}$ and $F_{R^I(r)}$ for the case where $v_0 \leq v_h(x_L, x_L)$, reflecting the analytical results in *Lemma 5*. It shows $F_{R^I(0)}$ crosses $F_{R^I(r)}$ only once from below at $b_0(x^*(r)) < r$. In principle, the distance between $F_{R^I(0)}(t)$ and $F_{R^I(r)}(t)$ for $t > r$ can be non-monotone due to the distribution of $X^{(1)}$.

While $F_{R^I(0)}$ as such does not suggest any lower bound on $F_{R^I(r)}$, it does suggest a simple upper bound of $F_{R^I(r)}$. Define $\tilde{F}^u_{R^I(r)}(t) = 0$ for $t < v_0$, $\tilde{F}^u_{R^I(r)}(t) = \Pr(b_0(X^{(1)}) < r)$ for $t \in [v_0, r)$ and $\tilde{F}^u_{R^I(r)}(t) = \Pr(b_0(X^{(1)}) \leq t)$ for $t \geq r$. In general $\tilde{F}^u_{R^I(r)}$ is first-order stochastically dominated by $F^u_{R^I(r)}$, but the two are equivalent if signals are i.i.d. To see this, note $F_{Y|X}(s|x)/F_{Y|X}(x|x) \succeq_{F.S.D.} L(s|x)$ when signals are affiliated, and the two are identical when signals are i.i.d.. It follows $v_l(x) \geq b_0(x)$ for all x and therefore $x_h(r) \leq b_0^{-1}(r)$ for $r \in [b^0_L, b^0_U]$. Furthermore, $b_0(x_h(r)) \leq r = v_l(x_h(r))$. Then an argument similar to the proof of *Lemma 5* shows $\delta_{r,h}(b_0(x)) = rL(x_h(r)|x) + \int_{x_h(r)}^{x} v_h(s,s)dL(s|x) \geq b_0(x)$ for all $x \geq x_h(r)$. Hence $\delta^{-1}_{r,h}(t; G^0_{\mathbf{B}}) \leq t$ for $t \geq r$. All inequalities above hold with equality when signals are i.i.d.

The alternative upper bound $\tilde{F}^u_{R^I(r)}$ is very easy to construct. When signals are strictly affiliated, it is less efficient than $F^u_{R^I(r)}$ in the sense that it is not a tight bound on $F_{R^I(r)}$. This is not surprising as $F^u_{R^I(r)}$ is constructed using full information from $G^0_{\mathbf{B}}$, while $\tilde{F}^u_{R^I(r)}$ is a only a functional of $F_{R^I(0)}$. On the other hand, $\tilde{F}^u_{R^I(r)}$ is equivalent to $F^u_{R^I(r)}$ when signals are i.i.d.. As this restriction has testable implications on $G^0_{\mathbf{B}}$, $\tilde{F}^u_{R^I(r)}$ can be useful in practice.

1.2.5. Bounds on $F_{R^{II}(r)}$

This subsection proposes bounds on counterfactual revenue distributions in 2nd-price auctions under reserve price r (denoted $F_{R^{II}(r)}$) that are constructed from $G^0_{\mathbf{B}}$. Theory predicts for any given reserve price r, the expected revenues in 2nd-price auctions are at least as high as those in 1st-price auctions provided bidder signals are affiliated. However, the size of this difference is an open empirical question. In addition, within the format of 2nd-price auctions, theory is silent about the choice of optimal reserve price r that maximizes expected revenue in 2nd-price auctions when signals are not independent. Knowledge of $F_{R^{II}(r)}$ helps address these open questions.

The equilibrium strategy in a second-price auction with reserve price r is $\beta_r(x) = v_h(x)$ for $x \geq x^*(r)$ and $\beta_r(x) < r$ for $x < x^*(r)$. For all $r > v_0$, the revenue distribution in a second-price auction with reserve price r is : [21]

$$\begin{aligned} F_{R^{II}(r)}(t) &= 0 \quad \forall t < v_0 \\ &= \Pr(X^{(1)} < x^*(r)) \quad \forall t \in [v_0, r) \\ &= \Pr(X^{(2)} < x^*(r)) \quad \forall t \in [r, v_h(x^*(r))) \\ &= \Pr(v_h(X^{(2)}) \le t) \quad \forall t \in [v_h(x^*(r)), +\infty) \end{aligned}$$

The following proposition derives bounds on $F_{R^{II}(r)}(t)$ that are constructed from $G^0_{\mathbf{B}}$.

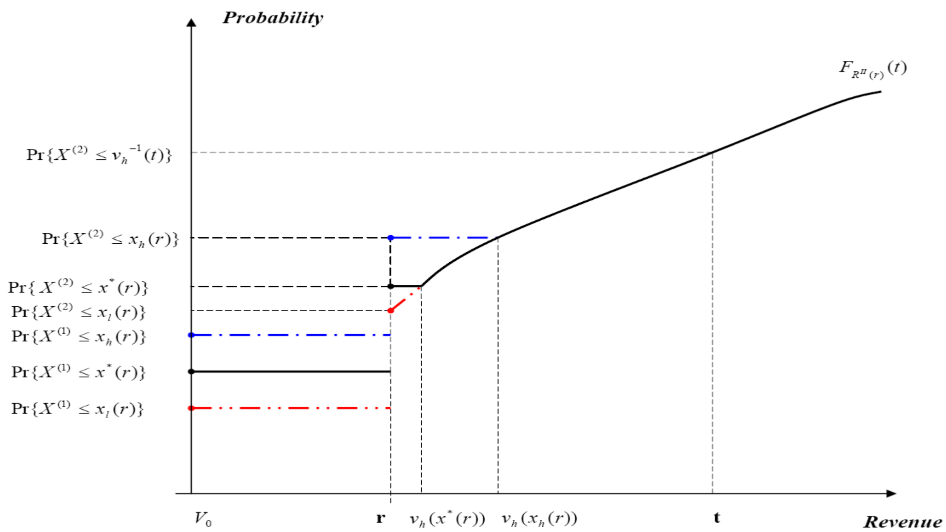
Proposition 9. *Suppose $v_0 < r$. Under A1 and A2, $F^u_{R^{II}(r)}(t) \preceq_{F.S.D.} F_{R^{II}(r)}(t) \preceq_{F.S.D.} F^l_{R^{II}(r)}(t)$, where*

$$\begin{aligned} F^l_{R^{II}(r)}(t) &= 0 \quad \forall t < v_0 \\ &= \Pr(b_0(X^{(1)}) < b_0(x_l(r))) \quad \forall t \in [v_0, r) \\ &= \Pr(v_h(X^{(2)}) \le t) \quad \forall t \in [r, +\infty) \end{aligned}$$

and

$$\begin{aligned} F^u_{R^{II}(r)}(t) &= 0 \quad \forall t < v_0 \\ &= \Pr(b_0(X^{(1)}) < b_0(x_h(r))) \quad \forall t \in [v_0, r) \\ &= \Pr(b_0(X^{(2)}) < b_0(x_h(r))) \quad \forall t \in [r, v_h(x_h(r))) \\ &= \Pr(v_h(X^{(2)}) \le t) \quad \forall t \in [v_h(x_h(r)), +\infty) \end{aligned}$$

[21]See proof of the following proposition for details.



Graph 2

The idea of the proof is better explained by *Graph 2*. Both $b_0(x_l(r))$ and $b_0(x_h(r))$ are identified from $G^0_{\mathbf{B}}$, and $b_0(X^{(1)})$ and $b_0(X^{(2)})$ are order statistics of $G^0_{\mathbf{B}}$. Furthermore in equilibrium $v_h(X^{(2)}) = \xi(b_0(X^{(2)}))$ and $v_h(x_h(r))$ can be identified as $\xi(b_0(x_h(r)))$. Therefore $F^l_{R^{II}(r)}$ and $F^u_{R^{II}(r)}$ can be constructed from $G^0_{\mathbf{B}}$, and the boundwidth depends on the distance between $b_0(x_l(r))$ and $b_0(x_h(r))$, and distributions of $X^{(1)}$ and $X^{(2)}$. In the Monte Carlo section below, I will experiment with different designs to study how this distance changes with the affiliation between signals.

1.3. Nonparametric Estimation of Bounds

1.3.1. Three-step estimators $\{\hat{F}^k_{R^I(r)}\}_{k=l,u}$

This section defines three-step estimators for the bounds on $F_{R^I(r)}(t)$. Throughout the section, we suppose data contains all bids submitted in L_n independent, homogenous auctions, each with n potential bidders and no reserve price.[22] Let $C(B) \equiv [b_L, b_U]$ denote the

[22]"Independence" here has both economic and statistical interpretations. First, it requires there is no strategic interaction or learning across the auctions so that the first-order condition characterizes equilibria in all auctions. Second, the random vector of bidders' private information is independent across auctions. "Homogeneity" means all auctions share the same commonly observed characteristics of the auctioned asset.

support of equilibrium bids in 1st-price auctions with non-binding reserve prices. For all $(m, b) \in C^2(B)$, define the following kernel estimators:

$$\hat{G}_{M,B}(m,b) = \frac{1}{L_n h_G} \sum_{l=1}^{L_n} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(m_{il} \leq m) K_G(\frac{b - b_{il}}{h_G})$$

$$\hat{g}_{M,B}(m,b) = \frac{1}{L_n h_g^2} \sum_{l=1}^{L_n} \frac{1}{n} \sum_{i=1}^{n} K_g(\frac{m - m_{il}}{h_g}, \frac{b - b_{il}}{h_g})$$

where b_{il} and m_{il} are respectively bidder i's own bid and the highest competing bid against him in auction l, L_n is the total number of auctions with n potential bidders, K_G and K_g are symmetric kernel functions with bounded hypercube supports with each side equal to 2, and h_g and h_G and corresponding bandwidths. It is well known that density estimators are asymptotically biased near boundaries of the support for $b \in [b_L, b_L + h_g) \cup (b_U - h_g, b_U]$. Let $\delta \equiv \max(h_g, h_G)$ and $C_\delta(B) = [b_L + \delta, b_U - \delta]$ be an expanding subset of $C(B)$ where $\hat{G}_{M,B}$ and $\hat{g}_{M,B}$ are asymptotically unbiased. Natural estimators for the boundaries of $C_\delta(B)$ are:

$$\tilde{b}_L = \hat{b}_L + \delta, \qquad \tilde{b}_U = \hat{b}_U - \delta$$

where $\hat{b}_L = \min_{i,l} b_{il}$ and $\hat{b}_U = \max_{i,l} b_{il}$ converge almost surely to b_L and b_U respectively. Nonparametric estimators for ξ and ξ_l are defined as:

$$\hat{\xi}(b) = b + \frac{\hat{G}_{M,B}(b,b)}{\hat{g}_{M,B}(b,b)}, \qquad \tilde{G}_{M,B}(b,b) = \int_{\tilde{b}_L}^{b} \hat{g}_{M,B}(t,b)dt + \hat{G}_{M,B}(\tilde{b}_L, b)$$

$$\hat{\xi}_l(b) = \hat{\xi}(\tilde{b}_L) \frac{\hat{G}_{M,B}(\tilde{b}_L, b)}{\hat{G}_{M,B}(b,b)} + \int_{\tilde{b}_L}^{b} \hat{\xi}(t) \frac{\hat{g}_{M,B}(t,b)}{\tilde{G}_{M,B}(b,b)} dt$$

where $\tilde{G}_{M,B}$ and $\hat{\xi}_l$ are defined over the random support $\hat{C}_\delta(B) \equiv [\tilde{b}_L, \tilde{b}_U]$. Let $b^0_{k,r}$ be a short-hand notation for $b_0(x_k(r))$. The first-step estimators for $b^0_{l,r}$ *and* $b^0_{h,r}$ are defined as:

$$\hat{b}^0_{l,r} = \arg\min_{b \in \hat{C}_\delta(B)} [\hat{\xi}(b) - r]^2, \quad \hat{b}^0_{h,r} = \arg\min_{b \in \hat{C}_\delta(B)} [\hat{\xi}_l(b) - r]^2$$

With estimates $\hat{b}^0_{l,r}$ and $\hat{b}^0_{h,r}$, we construct kernel estimator for $\delta_{r,l}(b)$ and $\delta_{r,h}(b)$ on $[\tilde{b}_L, \tilde{b}_U]$. For $k = \{l, h\}$, define:

$$\begin{aligned} \hat{\delta}_{r,k}(b; \hat{b}^0_{k,r}) &\equiv r\hat{L}(\hat{b}^0_{k,r}|b) + \int_{\hat{b}^0_{k,r}}^b \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt \quad \forall b \in (\hat{b}^0_{k,r}, b_U - \delta] \\ &\equiv r \quad \forall b \in [b_L + \delta, \hat{b}^0_{k,r}] \end{aligned}$$

where $\hat{\Lambda}(t) \equiv \hat{g}_{M,B}(t,t)/\hat{G}_{M,B}(t,t)$ and $\hat{L}(t|b) \equiv \exp(-\int_t^b \hat{\Lambda}(s)ds)$.

The second-step estimators for $\delta^{-1}_{r,l}(t)$ *and* $\delta^{-1}_{r,h}(t)$ are defined as:

$$\hat{\delta}^{-1}_{r,l}(t) = \arg\min_{b \in \hat{C}_\delta(B)} [\hat{\delta}_{r,l}(b) - t]^2, \qquad \hat{\delta}^{-1}_{r,h}(t) = \arg\min_{b \in \hat{C}_\delta(B)} [\hat{\delta}_{r,h}(b) - t]^2$$

Note that by definition, $\hat{C}_\delta(B) \subseteq C_\delta(B)$ with probability one. As a final step, the bounds on the counterfactual revenue distribution under reserve price r are estimated as:

$$\hat{F}^l_{R^I(r)}(t) = \frac{1}{L_n} \sum_{l=1}^{L_n} 1(B^{\max}_l \leq \hat{\delta}^{-1}_{r,l}(t)), \qquad \hat{F}^u_{R^I(r)}(t) = \frac{1}{L_n} \sum_{l=1}^{L_n} 1(B^{\max}_l \leq \hat{\delta}^{-1}_{r,h}(t))$$

where $B^{\max}_l = \max_{i=1,..,n} b_{il}$ is the highest bid in auction l.

The three-step estimators defined above converge in probability to $F^l_{R(r)}(t) \equiv \Pr(b_0(X^{(1)}) \leq \delta^{-1}_{h,r}(t))$ and $F^u_{R(r)}(t) \equiv \Pr(b_0(X^{(1)}) \leq \delta^{-1}_{l,r}(t))$ respectively over all r and t. Below I strengthen restrictions *A1* ,*A2* and *A3* to include all regularity conditions needed and state the main proposition.

S1 For $n \geq 2$, (i) The n-dimensional vectors of private signals $(x_{1l}, x_{2l}, ..x_{nl})_{l=1}^{L_n}$ are independent, identical draws from the joint distribution $F(x_1, .., x_n)$, which is exchangeable in all n arguments and affiliated with support $[x_L, x_U]$; (ii) $F(x_1, .., x_n)$ has $R + n$, $R \geq 2$, continuous bounded partial derivatives on $[x_L, x_U]^n$, with density $f(\mathbf{x}) \geq c_f > 0$ for all $\mathbf{x} \in [x_L, x_U]^n$.

S2 (i) The value function $\theta_n(.) : [x_L, x_U]^n \to \mathbb{R}_+$ is nonnegative, bounded, and continuous on the support; (ii) $\theta_n(.)$ is exchangeable in rival signals $\mathbf{X}_{-i}$, non-decreasing in all signals, and increasing in own signal X_i over $[x_L, x_U]$. (iii) $\theta_n(.)$ is at least R times continuously differentiable and $\theta(x_L) > 0$; (iv) $v_h(x_U) < \infty$ and $v_h'(x_U) < \infty$.

S3 (i) The kernels $K_G(.)$ and $K_g(.)$ are symmetric with bounded hypercube supports of sides equal to 2, and continuous bounded first derivatives; (ii) $\int K_G(b) = 1$, and $\int K_g(\tilde{B}, b) d\tilde{B} db = 1$; (iii) $K_G(.)$ and $K_g(.)$ are both of order $R + n - 2$.

Proposition 10. *Let $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ and $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$, where c's are constants.[23] Suppose S1, S2 and S3 are satisfied and $R > 2n - 1$, then for all $r \geq v(x_L)$ and $t \geq r$, $\hat{F}^k_{R^I(r)}(t) \xrightarrow{p} F^k_{R^I(r)}(t)$ for $k = l, u$.*

The proof proceeds in several steps. First, I prove smoothness of bid distributions in equilibrium, using regularity conditions of smoothness of signal distributions. Second, I show the kernel estimators $\hat{\xi}_l$ and $\hat{\xi}$ defined above converge in probability to ξ_l and ξ uniformly over $\hat{C}_\delta(B)$, and use a version of the Basic Consistency Theorem (which is generalized for objective functions defined on random support) to prove $\hat{b}^0_{k,r} \xrightarrow{p} b^0_{k,r}$ for $k = l, h$ and relevant $r > 0$. Next, I prove $\hat{\delta}_{k,r}(.; \hat{b}^0_{k,r}) \xrightarrow{p} \delta_{k,r}(.; b^0_{k,r})$ uniformly over $\hat{C}_\delta(B)$ and again use the generalized *BCT* to prove $\hat{\delta}^{-1}_{r,k}(t) \xrightarrow{p} \delta^{-1}_{r,k}(t)$ for all relevant t. Finally, the Glivenko-Cantelli

[23]The choice of constant c's will be discussed in the following section on Monte Carlo experiments.

uniform law of large numbers are used to show the empirical distributions of $B_l^{\max}$ evaluated at $\hat{\delta}_{r,k}^{-1}(t)$ for $k = l, h$ are consistent estimators for bounds on $F_{R^I(r)}(t)$.

1.3.2. Two-step estimators $\{\hat{F}^k_{R^{II}(r)}\}_{k=l,u}$

Given work on $\hat{F}^k_{R^I(r)}$ above, definition of two-step estimators for $\{F^k_{R^{II}(r)}\}_{k=l,u}$ is straightforward. The estimator for $F^u_{R^{II}(r)}$ is :

$$\begin{aligned}
\hat{F}^l_{R^{II}(r)}(t) &= \frac{1}{L_n}\sum\nolimits_{l=1}^{L_n} 1(B_l^{(1:n)} < \hat{b}^0_{l,r}) \qquad \forall t \in [v_0, r) \\
&= \frac{1}{L_n}\sum\nolimits_{l=1}^{L_n} 1(B_l^{(2:n)} < \hat{\xi}^{-1}(t)) \qquad \forall t \in [r, +\infty)
\end{aligned}$$

and the estimator for $F^u_{R^{II}(r)}$ is :

$$\begin{aligned}
\hat{F}^u_{R^{II}(r)}(t) &= \frac{1}{L_n}\sum\nolimits_{l=1}^{L_n} 1(B_l^{(1:n)} < \hat{b}^0_{h,r}) \qquad \forall t \in [v_0, r) \\
&= \frac{1}{L_n}\sum\nolimits_{l=1}^{L_n} 1(B_l^{(2:n)} < \hat{b}^0_{h,r}) \qquad \forall t \in [r, \hat{\xi}(\hat{b}^0_{h,r})) \\
&= \frac{1}{L_n}\sum\nolimits_{l=1}^{L_n} 1(B_l^{(2:n)} < \hat{\xi}^{-1}(t)) \qquad \forall t \in [\hat{\xi}(\hat{b}^0_{h,r}), +\infty)
\end{aligned}$$

where $\hat{b}^0_{k,r}$ is defined as above and $\hat{\xi}^{-1}(t) = \arg\min_{b \in \hat{C}_\delta(B)}[\hat{\xi}(b) - t]^2$ for $t \geq r$. Pointwise consistency of $\hat{F}^k_{R^{II}(r)}(t)$ for $r \geq v(x_L)$ and $t \geq r$ follows from similar arguments for consistency of $\hat{F}^k_{R^I(r)}(t)$ and the fact that $\hat{\xi}(\hat{b}^0_{h,r}) \xrightarrow{p} \xi(b^0_{h,r}; G^0_{\mathbf{B}}) = v_h(x_h(r))$.

1.4. Monte Carlo Experiments

This section reports Monte Carlo experiments of the three step estimator of bounds on $F_{R^I(r)}$ and $F_{R^{II}(r)}$. The objective is to illustrate how estimates of bounds vary with structural parameters such as affiliation between private signals, number of potential bidders n and reserve price r.

1.4.1. Analytical impacts of signal correlations on bounds : the case with $n = 2$

Before discussing Monte Carlo performances of bound estimators, I study analytically how bounds on the probability that no one bids above the reserve price in equilibrium vary with signal correlations. For a given value function, dependence between signals affects these bounds through two channels: bounds on the screening level $x^*(r)$, and bounds on equilibrium strategy $b_r(.)$. To capture these impacts, I use a parametric design where signal affiliations can be controlled.

Design 1 *($n = 2$ with pure common values (PCV) and affiliated signals) Two potential bidders compete in an auction with $V_i = (X_1 + X_2)/2$ for $i = 1, 2$. Private signals are noisy estimates of a common random variable, i.e., $X_i = X_0 + \varepsilon_i$ for $i = 1, 2$. For either bidder, his noise ε_i is independent from(X_0, ε_{-i}), and distributed uniformly on $[-c, c]$ for some $0 \leq c \leq 0.5$. The common random term X_0 is distributed uniformly on $[c, 1 - c]$.*

The signals have triangular marginal densities on $[0, 1]$.[24] Their correlation coefficient is:

$$corr(X_1, X_2) = \frac{var(X_0)}{var(X_0) + var(\varepsilon_i)} = \frac{(1 - 2c)^2}{(1 - 2c)^2 + 4c^2}$$

By definition, $v_h(x) = x$, $v_l(x) = E[X_2|X_2 \leq x, X_1 = x]$, and $v(x) = \frac{x + v_l(x)}{2}$. In this design, $v_l(x)$ has a closed form, and the impacts of correlations on boundwidth can be studied analytically.

[24]The density function is $f(x) = 4x$ for $0 \leq x \leq 0.5$ and $4 - 4x$ for $0.5 \leq x \leq 1$. For details, see Simon (2000).

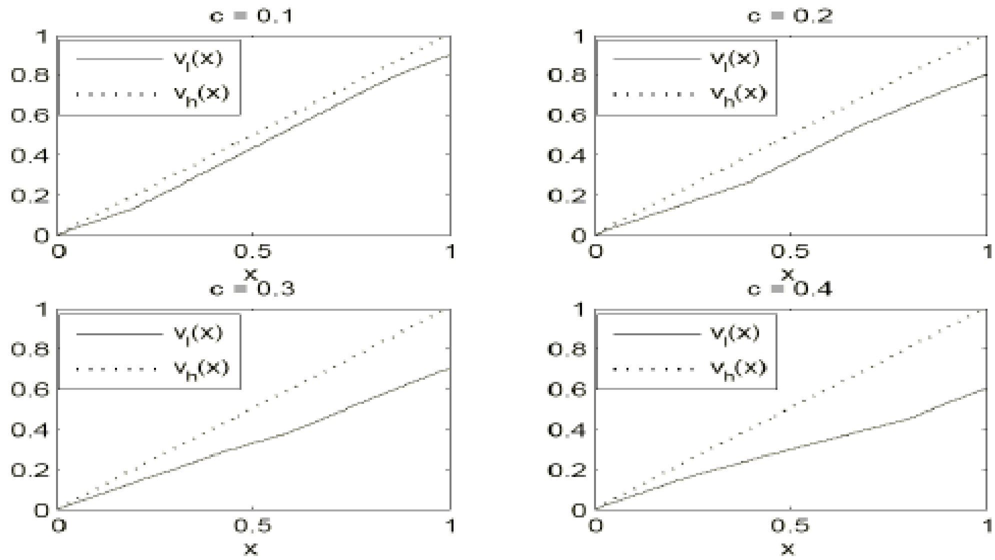


Figure 1 (a)

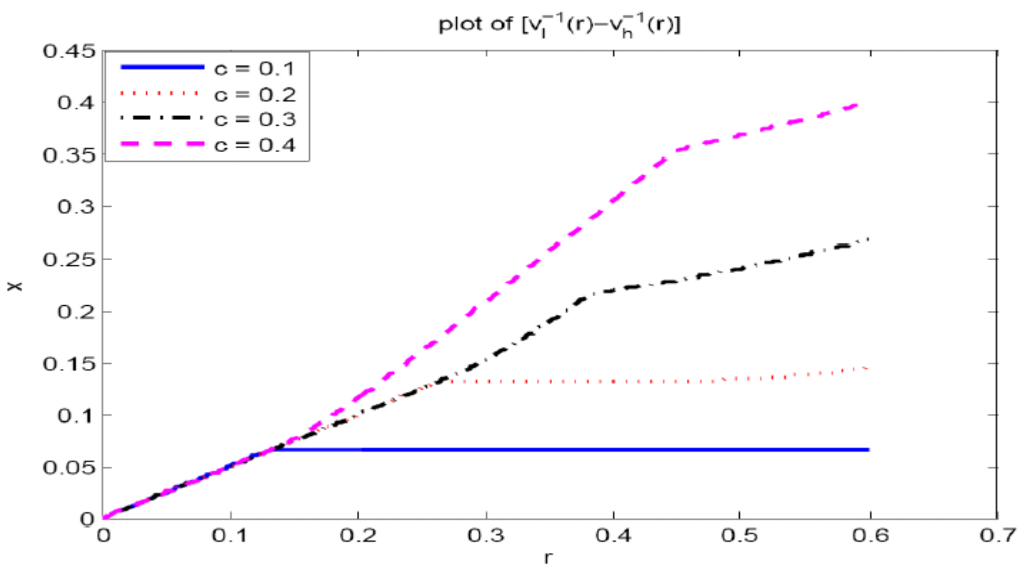


Figure 1 (b)

Figure 1(a) plots $v_l(x)$ and $v_h(x)$ for $c = [0.1 \quad 0.2 \quad 0.3 \quad 0.4]$. The distance between v_h and v_l is non-decreasing in signals, as $v_l(x)$ is a truncated expectation and therefore cannot increase faster than the threshold x. *Figure 1(b)* plots the boundwidth $x_h(r) - x_l(r)$ as a function of reserve prices for each c. For any given reserve price, bounds on screening levels are narrower as c decreases and correlation increases. Besides, boundwidths increase at a rate slower than r for high correlations. When $c = 0.1$ and $c = 0.2$, the boundwidths are invariant over some range of r.

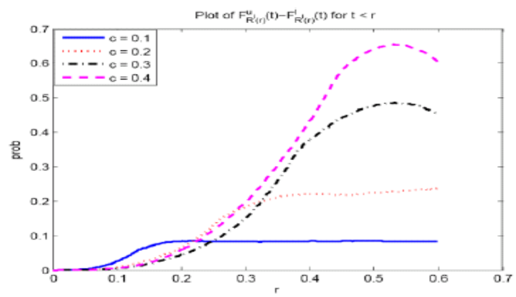


Figure 1 (c)

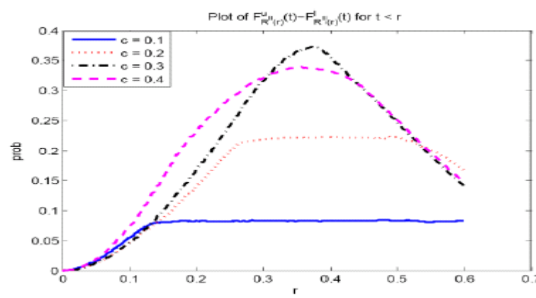


Figure 1 (d)

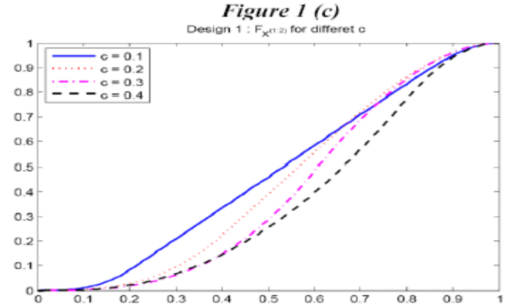


Figure 1 (e)

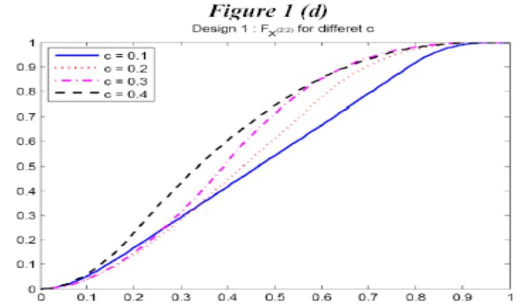


Figure 1 (f)

For different signal correlations, *Figure 1(c)* plots the boundwidths of the probability that neither bidder bids above r in a 1st-price auction. That is, $F_{X^{(1:2)}}(x_h(r)) - F_{X^{(1:2)}}(x_l(r))$, where $X^{(1:2)}$ is the higher of two signals. For a lower r, the boundwidth can be slightly wider when correlations are high. But as the reserve price increases, the boundwidths are unambiguously smaller for higher correlations. This is explained by the pattern in *Figure 1(b)* and the distribution of $X^{(1:2)}$. As *Figure 1(e)* shows, probability mass of $X^{(1:2)}$ is more skewed to the left when signals are closer to being uncorrelated. For smaller r, $x_l(r)$ and $x_h(r)$ are small and $x_h(r) - x_l(r)$ are close for all c, while $X^{(1:2)}$ has more mass close to 0 for more correlated signals. Hence $F_{X^{(1:2)}}(x_h(r)) - F_{X^{(1:2)}}(x_l(r))$ is bigger for $c = 0.1$. As r, $x_l(r)$ and $x_h(r)$ all increase, the boundwidths become wider for higher c because $x_h(r)$ $-$ $x_l(r)$ increases faster and the density of $X^{(1:2)}$ is higher in the relevant range as signals become less correlated.

Figure 1(d) plots boundwidths for the probability that neither bids above r in 2nd-price auctions, i.e. $F_{X^{(2:2)}}(x_h(r)) - F_{X^{(2:2)}}(x_l(r))$, for different c. In this case, the boundwidths associated with a smaller c is almost unambiguously smaller than those with smaller correlations. Likewise, the pattern is explained by similar arguments as demonstrated in *Figure 1(b)* and the distribution of $X^{(2:2)}$ plotted in *Figure 1(f)*.

1.4.2. Performance of $\hat{F}^k_{R^I(r)}$ under i.i.d. signals

This subsection focuses on the performance of three-step estimators $\hat{F}^k_{R^I(r)}$ when private signals are identically and independently distributed. The i.i.d. restriction has testable implications on observed bid distributions, and helps simplify the estimation procedures. In this subsection, I vary n, r and distributional parameters and study their impacts on estimator performances.

Design 2 *($n \geq 3$ with PCV and i.i.d. uniform signals) Private signals* $\{X\}_{i=1,..,n}$ *are identically, independently distributed as uniform on* $[0,1]$*. The pure common value is* $V_i = \sum_{j=1}^{n} X_j/n$.

Design 3 *($n \geq 3$ with PCV and i.i.d. truncated normal signals) Private signals* $\{X_i\}_{i=1,..,n}$ *are identically, independently distributed as truncated normal on* $[0,1]$ *with underlying parameters* (μ, σ^2)*. The pure common value is* $V_i = \sum_{j=1}^{n} X_j/n$.[25]

Independence is a special case of affiliation, and these two designs satisfy restrictions for a general symmetric, interdependent value auctions (*A1* and *A2*). Besides, changes in the number of bidders are not exogenous to the value distribution, as the distribution of the average of signals depends on n. Therefore both designs do not meet the necessary

[25]This form of value functions introduces a restriction/normalization on the signals, since it requires the support of signals is the same as the support of the values.

restrictions for tests distinguishing PV and CV auctions in Haile, Hong and Shum (2003). This makes the partial identification approach more interesting in both designs.

I experiment with different numbers of potential bidders and reserve prices for *Design 2*. For each pair of fixed n and r, I replicate the nonparametric estimators of $\hat{F}^k_{R^I(r)}$ for $1,000$ times, with each estimate calculated from equilibrium bids in 500 simulated first-price auctions. For *Design 2*, it can be shown $b_{0,n}(x) = \frac{n-1}{n}(\frac{1}{n}+\frac{1}{2})x$, and bids are simulated as random draws from a uniform distribution on $[0, \frac{n-1}{n}(\frac{1}{n}+\frac{1}{2})]$. For *Design 3*, I also vary distributional parameters μ and σ in addition to n and r. Likewise for each value of (n, r, μ, σ), I replicate the estimator for $1,000$ times, each based on 500 simulated auctions. For *Design 3*, $b_{0,n}(x) = \int_{x_L}^{x} \frac{2}{n}s + \frac{n-2}{n}\varphi(s) d\frac{F_X^{n-1}(s)}{F_X^{n-1}(x)}$, where $\varphi(x) = \mu - \sigma\frac{\phi(\frac{x-\mu}{\sigma})-\phi(\frac{x_L-\mu}{\sigma})}{\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x_L-\mu}{\sigma})}$ and $\frac{F_X(s)}{F_X(x)} = \frac{\Phi(\frac{s-\mu}{\sigma})-\Phi(\frac{x_L-\mu}{\sigma})}{\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x_L-\mu}{\sigma})}$. Hence equilibrium bids are simulated by first drawing $500 * n$ signals x_{il} randomly from the truncated distribution, and then calculate $b_{0,n}(x_{il})$ through numerical integrations.[26] For both designs, the true counterfactual distribution $F_{R^I(r)}$ can be recovered through inverting $b_r(.)$, which can be calculated using the closed form above.

In the symmetric equilibria above, bids in both designs are i.i.d.. This testable implication can be verified from the distribution of bids observed, and simplifies the estimation as $\xi(b; G^0_{\mathbf{B}_n}) = b + \frac{1}{n-1}\frac{G^0_{B_n}(b)}{g^0_{B_n}(b)}$ and $\xi_l(b; G^0_{\mathbf{B}_n}) = b$. The simplified estimator is $\hat{\xi}(b) \equiv b + \frac{1}{n-1}\frac{\hat{G}^0_{B_n}(b)}{\hat{g}^0_{B_n}(b)}$, where $\hat{G}_{B_n}(b) = \frac{1}{L_n}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n} 1(b_{il} \le b)$, $\hat{g}_n(b) = \frac{1}{L_n h_g}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n} K(\frac{b_{il}-b}{h_g})$ and L_n is the number of auctions with n bidders. For estimation, I use the tri-weight kernel $K(u) = \frac{35}{32}(1-u^2)1(|u| \le 1)$.[27] Bandwidths h_g is $2.98 * 1.06\hat{\sigma}_b(nL_n)^{-\frac{1}{4n-4}}$, where $\hat{\sigma}_b$ is the empirical standard deviation of bids in the data. The bandwidths are chosen in line with the consistency proposition in the appendix, while the constant factor $1.06\hat{\sigma}_b$ is chosen by

[26] I use the midpoint approach for numerical intergrations in this paper.

[27] The triweight kernel is of order 2. In principle when $n \ge 3$, kernels used in $\hat{g}_{Bn}$ should be of higher order. But can lead to the issue of negative density estimates. Therefore empirical literature typically ignore this requirement and use kernels with order 2.

the "rule of thumb" (Li, Perrigne and Vuong 2002). The multiplicative factor 2.98 is due to the use of tri-weight kernels (Hardle 1991).

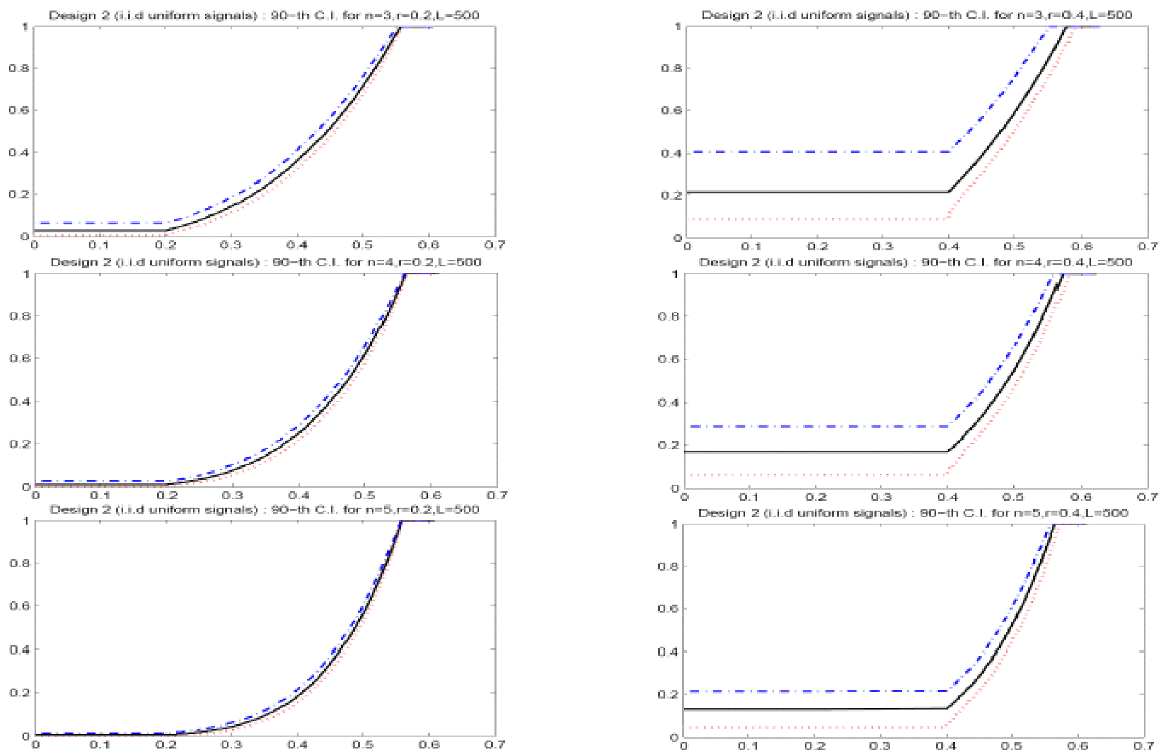


Figure 2

Figure 2 plots the true revenue distribution $F_{R^I(r)}$ in *Design 2* and, for different n and r, reports the 5th percentile of $\hat{F}^l_{R^I(r)}$ and the 95th percentile of $\hat{F}^u_{R^I(r)}$ out of $1,000$ pairs of estimates. The two percentiles form an estimate of the conservative 90% pointwise confidence interval for the bounds $[F^l_{R^I(r)}, F^u_{R^I(r)}]$.[28] The true revenue distribution always falls within the 90% point-wise confidence interval for the bound. The confidence intervals for lower r are narrower, holding n constant. On the other hand, more potential bidders correspond to tighter confidence regions ceteris paribus. To understand the pattern, note the boundwidth of the probability that no one bids above r in equilibrium is $\Pr(b_{0,n}(X^{(1:n)}) \leq r) - \Pr(b_{0,n}(X^{(1:n)})$

[28]This approach for constructing a pointwise confidence region *for the bounds* was used in Haile and Tamer (2003). An alternative way to report the performance of our estimator would be to construct a pointwise confidence region *for the true distribution* introduced by Manski and Imbens (2004).

$\leq \frac{n-1}{n}r) = F_{X^{(1:n)}}(\frac{n}{n-1}\frac{2n}{n+2}r) - F_{X^{(1:n)}}(\frac{2n}{n+2}r)$, which is increasing in r for a given n. For a given r, $\frac{1}{n-1}\frac{2n}{n+2}r$ decreases in n and this offsets the impacts of a rising $\frac{2n}{n+2}r$ and a more left-skewed $F_{X^{(1:n)}}$ as competition increases. The simulations suggest changes in the width of confidence intervals are mostly due to impacts of n and r on the boundwidths of $F_{R^I(r)}$.

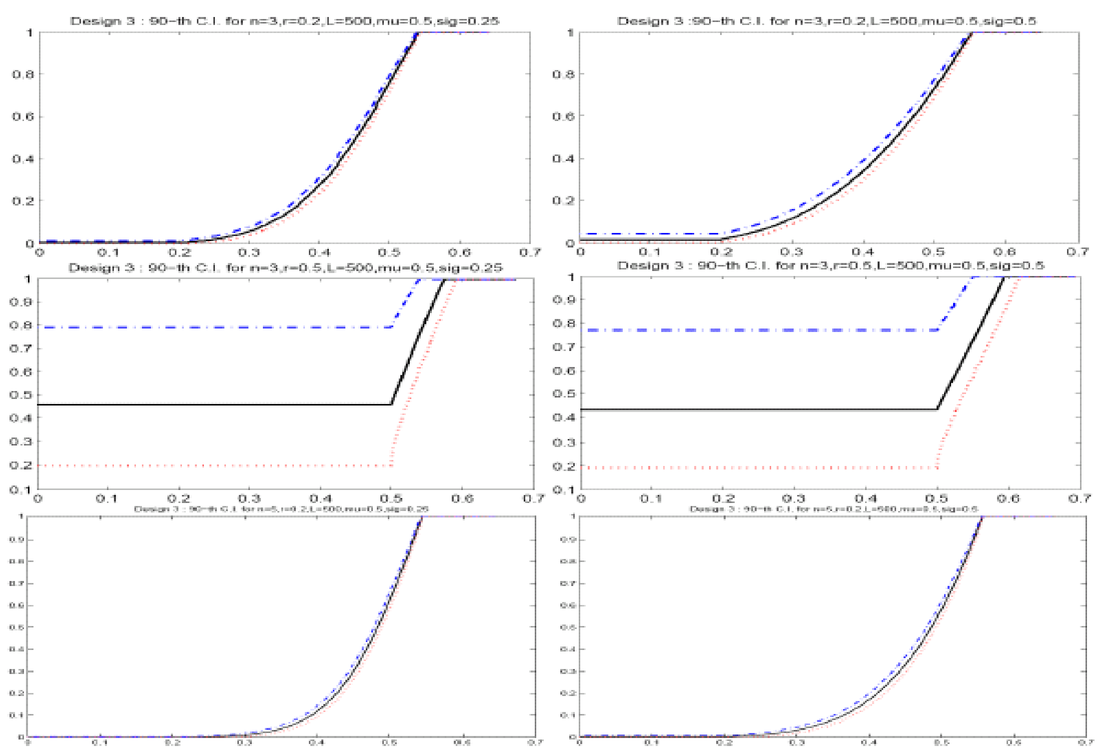


Figure 3(a)

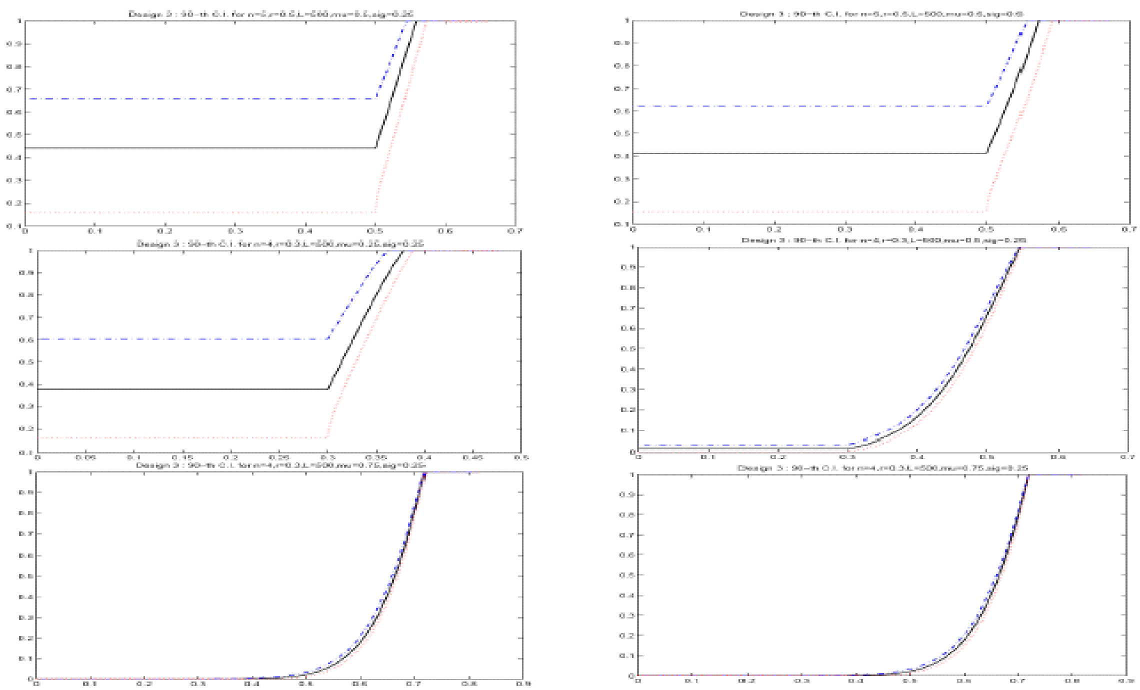


Figure 3(b)

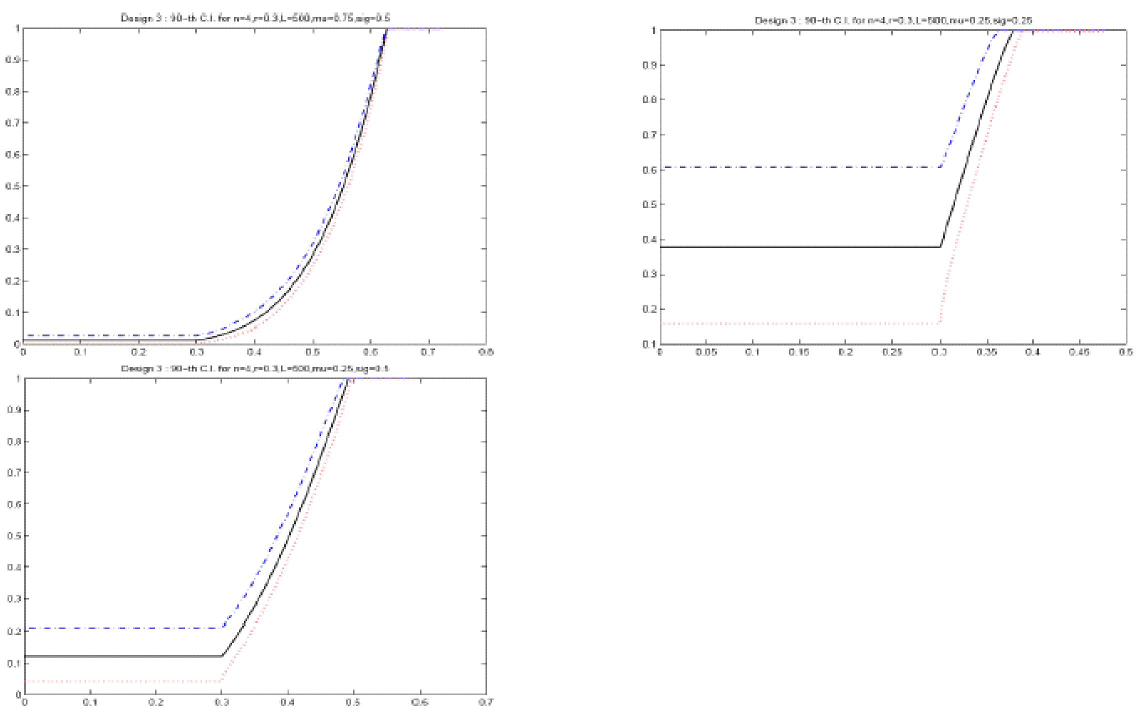


Figure 3(c)

Figure 3 reports $F_{R^I(r)}$ and the estimated 90% confidence interval for *Design 3*. Again, the true revenue distribution falls within the 90% point-wise confidence intervals for the parameters considered. The impacts of n and r on the estimated confidence intervals in *Design 3* are the same as those for *Design 2* in *Figure 2*. In addition, *Figure 3* also shows impacts of distributional parameters μ and σ on confidence intervals. First, holding n, r and σ fixed, the confidence intervals become narrower as μ increases. This is because for all t, $E(X|X \leq t)$ gets closer to t as the distribution of X is more skewed to the left. Consequently, $x^*(r)$ decreases for a given r, while the distance between v_h and v_l also becomes smaller. As a result, the bound on the screening probability is shifted to the left and becomes tighter. Second, the impact of σ on confidence intervals depends on μ, holding n and r fixed. A higher standard deviation increases the width of confidence intervals for signal distributions sufficiently skewed to the left, but reduce the width of confidence intervals for signal distributions sufficiently skewed to the right. The impacts are more obvious for distributions skewed to the right. This pattern is explained by similar reasoning above. Again, simulations suggest changes in the width of confidence intervals are mostly due to impacts of n and r on the boundwidths of $F_{R^I(r)}$.

1.4.3. Performance of $\hat{F}^k_{R^I(r)}$ with affiliated signals

When signals are not i.i.d., there are no simplified forms for $\hat{\xi}$ and $\hat{\xi}_l$, and the full nonparametric estimates in Section 3 applies. In this subsection I extended *Design 1* for $n \geq 3$ so that $V_i = \sum_{j=1}^n X_j/n$, and experiment with the correlation parameter c to study its impact on the performance of estimators.

With $n \geq 3$, it is impractical to derive the closed form of the inverse hazard rate $f_{Y|X,n}(u|u)/F_{Y|X,n}(u|u)$. To find out the true revenue distribution, I replace $v_h(x)$ and $L(s|x)$

with their kernel estimates in a simulated sample of $5 * 10^5$ auctions, and calculate the equilibrium bidding strategies using these estimates and numerical integrations. The true $F_{R^I(r)}$ is then recovered with knowledge of the distribution of the highest signal $X^{(1:n)}$.

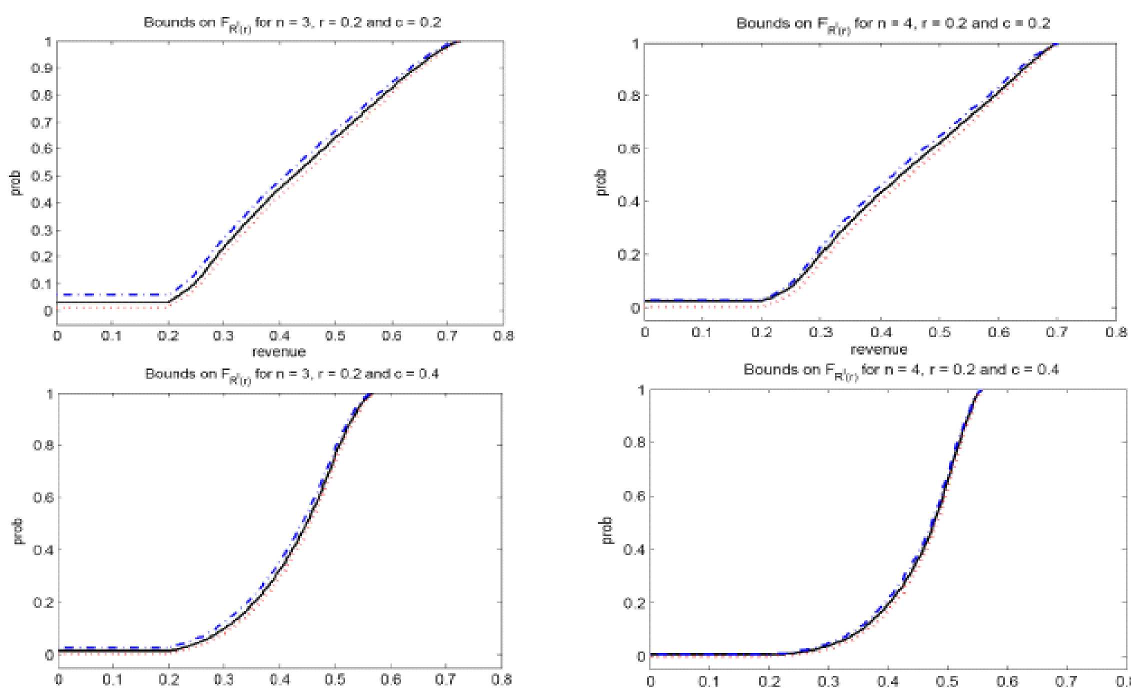


Figure 4(a)

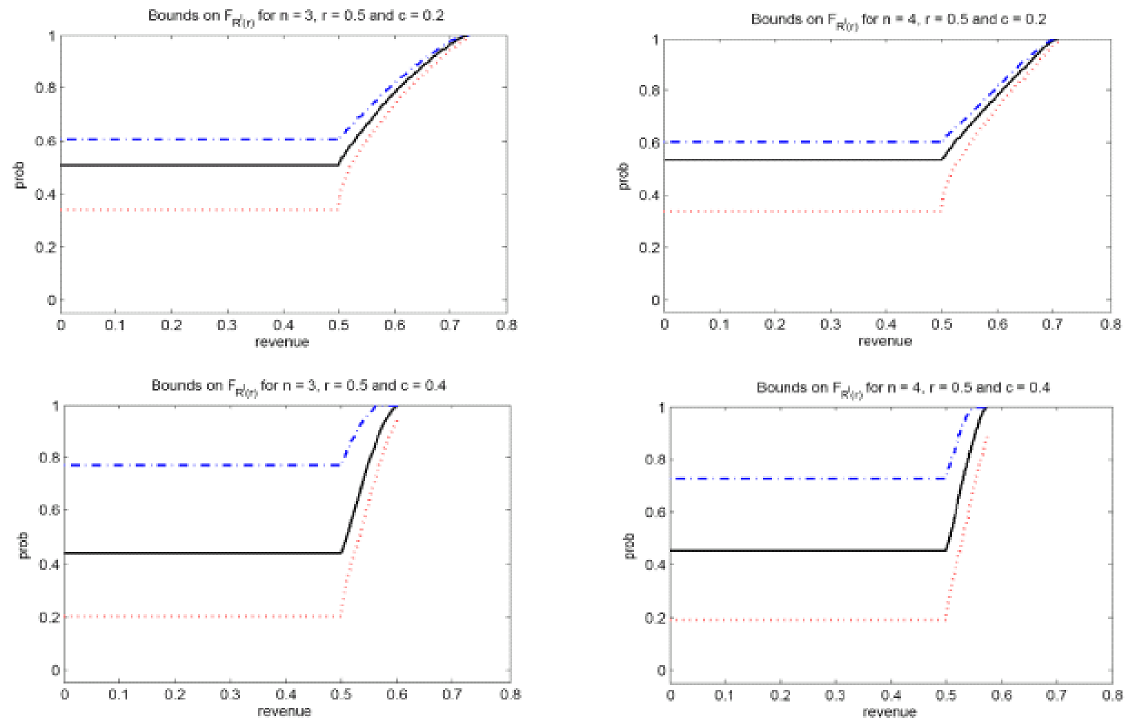


Figure 4(b)

For each (c, n), I simulate 200 samples, with each containing $1,000$ simulated first-price auctions. For each r and revenue level t, *Figure 4* reports the 5-th percentile of $\hat{F}^{l}_{R^I(r)}(t)$ and the 95-th percentile for $\hat{F}^{u}_{R^I(r)}(t)$ out of 200 pairs of estimates. This forms an estimate for a conservative 90% confidence interval for the bounds on $F_{R^I(r)}$. *Figure 4* shows the true $F_{R^I(r)}$ lies within the estimated confidence interval for $r = 0.2$ or 0.5, $c = 0.2$ or 0.4 and $n = 3$ or 4. Holding r and c constant, the widths of the estimated confidence intervals decrease slightly as n increases. For $r = 0.2$, higher correlation leads to slightly wider confidence intervals, whereas for $r = 0.5$ higher signal correlation leads to obviously narrower confidence intervals. Smaller correlations among signals implies the distribution of $X^{(1:n)}$ is more skewed to the left, and the distance between v_l and v_h are bigger. These explain why a higher c leads to wider confidence intervals when r is high at 0.5. On the other hand, when r is low at 0.2, the left-skewness of $F_{X^{(1:n)}}$ offsets the impact of a wider bound $[x_l(r), x_h(r)]$ due to a higher

c, and may lead to a narrower confidence interval. Furthermore, the theory also states for $x \geq x^*(r)$ the bounds on $\delta_r(b_0(x))$ is tighter as $b_0(x)$ increases. For $t > r$, this counteracts the left skewness of $F_{X(1:n)}$ due to lower correlations. This prediction is consistent with patterns in *Figure 4* where confidence intervals on $F_{R^I(r)}$ never broaden substantially as revenue level t increases.

1.5. Extensions

1.5.1. Heterogenous Auctions

In practice, bidding data are often collected from heterogenous auctions that report different characteristics of auctioned objects. If commonly observed by all bidders, such heterogeneity affects bidders' strategies, and distributions of counterfactual revenues. When heterogeneity across auctions is completely observed in the data, the logic for bounds in homogenous auctions extends in principle to bounds on revenue distributions conditional on specific values of auction features. Auctions are homogenous within subsets of the data where such features (denoted $\mathbf{Z}$) are controlled for, and the same algorithm in the benchmark model extends immediately to bounds on the conditional revenue distribution given these characteristics $F_{R^I(r)|\mathbf{Z}=\mathbf{z}}$. Such bounds are constructed from conditional bid distribution $G^0_{\mathbf{B}|\mathbf{Z}=\mathbf{z}}$.

The real challenge posed by observed auction heterogeneity is empirical. The construction of bounds on conditional revenue distributions requires a large cross-sectional data of homogenous auctions for fixed $\mathbf{z}$ and n. The issue of data deficiency aggravates as the dimension of $\mathbf{z}$ becomes higher. The rest of this subsection shows if signals are independent from auction characteristics conditional on n, and are additively separable from them in value functions, then it is possible to "homogenize" bids across heterogenous auctions, thus alleviating the data deficiency problem.

A1' (Interdependent Values) $V_{i,N} = h(\mathbf{Z}'\boldsymbol{\gamma}) + \theta_N(X_i, \mathbf{X}_{-i})$, where $h(.)$ is differentiable, and θ_N is bounded, continuous, exchangeable in its last $N-1$ arguments, non-decreasing in all arguments, and increasing in X_i.

A4 (Conditional Independence of $\mathbf{X}$ *and* $\mathbf{Z}$*)* Conditional on $N = n$, $\{X_i\}_{i=1,..,n}$ is independent from $\mathbf{Z}$.

Then a *PSBNE* in the auction with no binding reserve price is a profile of strategies that solve:

$$b_{0i}(x, \mathbf{z}; n) = \arg\max_b E[(V_i - b)1\{\max_{j \neq i} b_{0j}(X_j, \mathbf{Z}) \leq b\}|X_i = x, \mathbf{Z} = \mathbf{z}, N = n].$$

Under additional assumptions above, common knowledge of auction features impact strategies of all bidders in the same way. As the proposition below shows, the separability and the index specification of value functions are inherited by bidding strategies in equilibrium.

Proposition 11. *Under A1', A2, A3 and A4, bidders' equilibrium strategies satisfy :* $b_{0i}(x, \mathbf{z}; n) = h(\mathbf{z}'\boldsymbol{\gamma}) + \lambda(x; n)\ \forall x, z\ \forall i$, *where* $\lambda(x; n) \equiv \int_{x_L}^{x} \phi(s; n) dL(s|x; n)$, *and* $\phi(s; n) \equiv E[\theta(\mathbf{X})|X_i = Y_i = s; N = n]$.

Fix the number of potential bidders n, the proposition implies $E(b_{0i}|\mathbf{Z} = \mathbf{z}, N = n) = h(\mathbf{z}'\boldsymbol{\gamma}) + E(\lambda(X; N)|\ N = n)$, where the second term is a constant independent from $\mathbf{Z}$. This becomes a single index model, and both Powell, Stock and Stocker (1989) and Ichimura (1991) showed $\boldsymbol{\gamma}$ can be identified up to scale, and estimated consistently using average derivative estimator and semiparametric least square estimators respectively. In the special case where $h(.)$ is known to be the identity function, an OLS regression of bids from heterogenous auctions on $\mathbf{z}$ for a fixed n will estimate $\boldsymbol{\gamma}$ consistently. Alternatively, including dummies for the number of potential bidders in a pooled regression will also give consistent coefficient estimators for $\boldsymbol{\gamma}$.

A corollary of the proposition is that for any pair of different features of auctions $\mathbf{z}$ and $\bar{\mathbf{z}}$, the equilibrium strategies for a given signal x are related as $b_0(x, \mathbf{z}; n) = b_0(x, \bar{\mathbf{z}}; n) - h(\bar{\mathbf{z}}'\boldsymbol{\gamma}) + h(\mathbf{z}'\boldsymbol{\gamma})$. Thus when h is known, bids across heterogenous auctions can be "homogenized" at any specific reference level $\mathbf{z}$ so that more observations are available for estimating $G^0_{\mathbf{B}(\mathbf{Z})}$. Larger sample size leads to better performance of estimators of bounds on $F_{R^I(r)|\mathbf{Z}=\mathbf{z}}$.

1.5.2. Endogenous Participation

In practice, bidding data are often collected from homogenous auctions with a binding reserve price r known to all bidders.[29] Data from such auctions can depart from those with non-binding reserve prices in one or both of two aspects : First, bids from potential bidders that are screened out may not be observed. Second, data may only include auctions with at least one bid above r, and thus exclude those where all bidders are screened out (i.e. $X^{(1)} < x^*(r)$). In both cases, the algorithm in the benchmark model can not be applied immediately.

In addition, a binding reserve price r in the data also reduces the scope of reserve prices that are eligible for counterfactual analyses, for the logic underlying bounds in the benchmark case only applies to revenue distributions for $r' > r$. This is because bids below r reveal no information about underlying signals as the equilibrium condition linking $G^r_{\mathbf{B}}$ with θ and $F_{\mathbf{X}}$ only holds for $b_r(x) \geq r$. As a result, for all $r' < r$, $x^*(r')$ is lower than $x^*(r)$ and can

[29] In the case of heterogenous auctions, reserve prices are often set according to characteristics of the auctioned object. The subsection above showed observed heterogeneities can be controlled for. For the sake of highlighting challenges due to endogenous participation, I focus on homogenous auctions in this subsection.

not be bounded in its small neighborhoods using equilibrium conditions.[30] Throughout this subsection, I focus on the bounds for $F_{R^I(r)}$. Extensions to those for $F_{R^{II}(r)}$ is straightforward and omitted.

1.5.2.1. Unobserved screened bidders. Unobserved bids from bidders who are screened out matter in bounding $F_{R^I(r')}$ (where $r' > r$) only in the sense that they may make the number of potential bidders unobservable. (For now assume auctions with $X^{(1)} < x^*(r)$ are also observed in the data.) If the number of potential bidders is known (as is often the case in applications), then the algorithm for bounding $F_{R^I(r)}$ can be applied, even if data do not contain bids from bidders that are screened out. The following lemma generalizes the equilibrium condition (1.2) for distributions rationalized under a binding r.

Proposition 12. *Suppose a distribution of bids $G^r_{\mathbf{B}}$ in a first-price auction with reserve price $r \geq 0$ is rationalized by $(\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$. Then $\xi(b_r(x); G^r_{\mathbf{B}}) = v_h(x; \theta, F_{\mathbf{X}})$ $\forall$ $x \geq x^*(r)$.*

A complication due to binding reserve prices in the data is that the lower bound of $v(x)$ can no longer be identified from $G^r_{\mathbf{B}}$, for bids below r cannot be linked to signals through equilibrium conditions. The solution is to bound $v(x)$ from below with the expected payment of a winner in second-price auctions with reserve price r. For $x \geq x^*()$ define

$$v_{l,r}(x) \equiv r \tfrac{F_{Y|X}(x^*(r)|x)}{F_{Y|X}(x|x)} + \int_{x^*(r)}^{x} v_h(s) \tfrac{f_{Y|X}(s|x)}{F_{Y|X}(x|x)} ds$$

Then $v_{l,r}(x)$ is increasing in x by monotonicity of the value function and affiliations between signals, and $v(x) \geq v_{l,r}(x)$ for $x \geq x^*(r)$ by the equilibrium condition in second-price auctions with r. (The formal proof is similar to the benchmark case and omitted.) Hence for all $r' > r$, $x^*(r')$ is bounded by $x_{h,r}(r') \equiv \arg\min_{x \in [x^*(r), x_U]} (v_{l,r}(x) - r')^2$ and

[30] For $r' < r$, it can be shown that $F_{R^I(r')}$ is bounded below by $F_{R^I(r)}$ for $t \geq r$.

$x_{l,r}(r') \equiv \arg\min_{x\in[x^*(r),x_U]}(v_h(x) - r')^2$. Then $v_h(x)$ and $v_{l,r}(x)$ are identified from $G^r_{\mathbf{B}}$ for $x \geq x^*(r)$ respectively as $\xi(b_r(x); G^r_{\mathbf{B}})$ and

$$\xi_{l,r}(b_r(x); G^r_{\mathbf{B}}) \equiv r\frac{G^r_{M|B}(r|b_r(x))}{G^r_{M|B}(b_r(x)|b_r(x))} + \int_r^{b_r(x)} \xi(\tilde{b}; G^r_{\mathbf{B}})\frac{g^r_{M|B}(r|b_r(x))}{G^r_{M|B}(b_r(x)|b_r(x))}d\tilde{b}$$

By similar reasoning as in the benchmark case, bounds on the $\delta^r_{r'}$-mapping (which maps $b_r(x)$ into $b_{r'}(x)$ for $x \geq x^*(r')$) are identified as

$$\delta_{r,r',k}(b_r(x); G^r_{\mathbf{B}}) = r'\tilde{L}(b^r_{k,r'}|b; G^r_{\mathbf{B}}) + \int_{b^r_{k,r'}}^{b} \xi(\tilde{b}; G^r_{\mathbf{B}})d\tilde{L}(\tilde{b}|b; G^r_{\mathbf{B}})$$

where $b^r_{k,r'} \equiv b_r(x_{k,r}(r'))$ for $k = l, h$, and are identified as inverses of $\xi(.; G^r_{\mathbf{B}})$ and $\xi_{l,r}(.; G^r_{\mathbf{B}})$ over $[r, b_r(x_U)]$ respectively. It can be shown that $\delta_{r,r',k}(b_r(.); G^r_{\mathbf{B}})$ is increasing for $x \geq x_{k,r}(r')$, and inverting $\delta_{r,r',k}(.; G^r_{\mathbf{B}})$ at $t \geq r'$ gives bounds on $b_r(b^{-1}_{r'}(t))$. Thus bounds on $F_{R^I(r')}$ can be constructed from the distribution of $b_r(X^{(1)})$.

1.5.2.2. Unobserved screened auctions (with $X^{(1)} < x^*(r)$). When data exclude auctions with a reserve price r that screens out all bidders (i.e. $X^{(1)} < x^*(r)$), we observe the distribution of equilibrium bids b_r conditional on at least one bidder bids above r (denoted $G^r_{\mathbf{B}|B^{(1)}\geq r}$) rather than $G^r_{\mathbf{B}}$. For $b > r$, $G^r_{M|B}(b|b)$ and $g^r_{M|B}(b|b)$ can still be identified from $G^r_{\mathbf{B}|B^{(1)}>r}$, and thus bounds on $b_r(x^*(r'))$ and the $\delta_{r,r'}$-mapping can be constructed as above. However, $G^r_{\mathbf{B}|B^{(1)}\geq r}$ can only be used to construct bounds on $F_{R^I(r)|X^{(1)}\geq r}$. That is,

$$\begin{aligned} \Pr(\xi(B^{(1)}_r; G^r_{\mathbf{B}}) < r'|B^{(1)}_r \geq r) \leq \quad & \Pr(X^{(1)} < x^*(r')|X^{(1)} \geq x^*(r)) \\ \leq \quad & \Pr(\xi_{l,r}(B^{(1)}_r; G^r_{\mathbf{B}}) < r'|B^{(1)}_r \geq r) \end{aligned}$$

and for $t \geq r'$,

$$\begin{aligned}\Pr(B_r^{(1)} \leq \delta_{r,r',l}^{-1}(t;G_{\mathbf{B}}^r)|B_r^{(1)} \geq r) \leq \quad & \Pr(X^{(1)} \leq b_{r'}^{-1}(t)|X^{(1)} \geq x^*(r)) \\ \leq \quad & \Pr(B_r^{(1)} \leq \delta_{r,r',h}^{-1}(t;G_{\mathbf{B}}^r)|B_r^{(1)} \geq r)\end{aligned}$$

where $B_r^{(1)}$ is shorthand for $b_r(X^{(1)})$. The probability that r screens out all bidders $\Pr(X^{(1)} < r)$ is needed to bound the unconditional distribution $F_{R^I(r)}$. It is impossible to identify this probability solely from $G_{\mathbf{B}|B^{(1)}\geq r}^r$ without further restrictions on $F_{\mathbf{X}}$. However, the lemma below shows when bidder signals are i.i.d., $\Pr(X^{(1)} < x^*(r))$ can be recovered from $G_{\mathbf{B}|B^{(1)}\geq r}^r$ alone. [31]

Proposition 13. *Suppose signals $\{X_i\}_{i=1,..N}$ are i.i.d. in first-price auctions with N potential bidders and reservation price r. If both the number of active bidders and N are observed, then $\Pr(X^{(1)} < x^*(r))$ is identified even if auctions with $X^{(1)} < x^*(r)$ are not observed.*

1.5.2.3. About the number of potential bidders. That the number of potential bidders N is observed is key to our discussion of auctions with endogenous participations so far. This is not an issue in some applications where N is directly reported in the data, or where good proxies exist. In other applications, the issue is more subtle.

In some applications, neither bidders nor econometricians can observe N. Then strategic decisions can be modeled as based on subjective probability distributions of N given private signals (denoted $p(N = n|X = x)$). Bidders integrate $v_{h,N}$, $f_{Y|X,N}$ over N with respect to this distribution and make strategic decisions based on these integrated primitives, so the actual number of potential bidders becomes irrelevant in equilibria. The new equilibrium conditions

[31] Under *A1*, the auction model still has interdependent values even $\{X_i\}_{i\in N}$ are i.i.d..

can also be manipulated through change of variables to get an analog of (1.2) that links bid distributions observed to model primitives.[32] The logic of partial identification in benchmark models can be extended in principle to bound revenue distribution in such equilibria with unobserved potential bidders.

In other applications where bidder signals are i.i.d., the number of potential bidders can be identified even if data only report the number of actual bidders. This is because in equilibria, the number of actual bidders is distributed as Binomial(n, p) with p equal to the screening probability $\Pr(x \geq x^*(r))$. Provided the distribution of bids and actual bidders are rationalizable,[33] both n and p are uniquely identified.

1.6. Application : U.S. Municipal Bond Auctions

Municipal bonds are a chief means of debt-financing for U.S. state and county governments. They are usually issued to finance public projects such as construction or renovation of schools and public transportation facilities, etc. A main advantage of muni-bonds over corporate securities is that interest income from them are exempt from federal and local taxes. As a result, they appeal especially to investors in high tax brackets. In 2005, the total par amount of outstanding municipal bonds was $1.8 trillion. [34]

[32]One example of such an application is Hendricks, Pinkse and Porter (2003). In OCS auctions, potential bidders' decisions to submit bids take multi-stages. HPP endogenize participations by introducing multiple signals, each corresponding to a stage in the decision-making. Then only those still active in the last-stage and their signals are relevant to decisions on strategic bids. The additional restrictions in the model is that decisions to remain active till the last stage only depends on signals from previous stages, and that conditional on last-stage signals, signals in previous stages reveal no information about bidders' values.

[33]See Guerre et.al (2000) for conditions for rationalizability.

[34]Source of information : SIFMA(2005)

1.6.1. Institutional details

Muni-bonds are identified by their issuers and several basic features (coupon rates, maturity dates, and par amounts, etc).[35] Investors valuate muni-bonds based on this information and its implied risks (credit risks, interest rate risks, and liquidity risks, etc.). [36] On the primary market, muni-bonds are initially issued through first-price, sealed-bid auctions to underwriters (security firms such as investment banks). Notices of these competitive sales are posted on major industry publications such as *The Bondbuyer*. In practice, issuers usually package a series of bonds for one auction, and investment banks participate by bidding a single dollar price per $100 in par value for the whole series. The bidder with the highest dollar price wins the right to underwrite the entire series, and consequently resell the series to investors on secondary markets with some mark-up.

To decide whether and how to bid for a series, securities firms tap into their research and marketing staff to assess the creditworthiness of the municipality and the market prospects of the bonds. Typically managers meet with sales and research personnel on Monday mornings to review new issues on the week's calendar. Both in-house researchers and traders contribute to estimates about market trends and how the issues considered may trade on secondary markets. For issues with a large par amount, investment banks usually form bidding syndicates, where members share responsibilities for reselling the bonds as well as the liability for unsold bonds. A syndicate is usually clearly defined for each issuance, as

[35] A coupon rate is the interest rate stated on the bond and payable to the bondholder on a semi-annual basis. A maturity date is the date on which the bondholder will receive par value of the bond along with its final interest payment.

[36] *Credit risk* measures how likely the issuer is to default on its payment of interests and principals. *Interest rate risk* is due to flucuations in real interest rates that affect the market value of bonds (to both speculators and long-term investors). *Liquidity risk* refers to the situation where investors have difficulty finding buyers when they want to sell, and are forced to sell at a significant discount to market value.

underwriters traditionally stay in the group where they bid on the last occasion that the issuer came to market.

As of 2006, more than 2,100 securities firms are registered with the Municipal Securities Regulatory Board and authorized as legal underwriters. However, only a small number of these firms are active bidders in competitive sales. By 1990, 25 leading underwriters managed about 75 percent of the total volume of all new long-term issues either as a lone bidder or leaders of syndicates.

1.6.2. Bond values : private or common ?

The bounds proposed introduce an approach of partial identification for policy analyses which is applicable regardless of underlying paradigms (PV or CV). This is highly relevant in the context of muni-bond auctions, as institutional details do not suggest conclusive evidence for either paradigm and there are limitations in empirical methods available for differentiating between them using bid data.

The value of bonds for firms in these auctions are resale prices on secondary markets. On most occasions bidders on the primary market cannot foresee at what price they can resell the bonds, and therefore only have noisy estimates. These estimates capture the syndicates' expectation on how investors on secondary markets interpret bond features, and depend on their beliefs about the skills of their sales and trading staff. The estimates are also built on companies' perception of how investors view relevant uncertainties such as the creditworthiness of municipalities and fluctuations of future real interest rates.

The crucial question is whether a bidding syndicate can extract additional useful information about bond values from competitors' estimates if it had access to them. The auction is one with common values if and only if the answer is positive. On some occasions, all firms participating in an auction manage to pre-sell bonds to secondary investors prior to their actual participation. Such auctions are private value ones, as all bidders have perfect foresight of bond values. On other occasions pre-sales are not possible or limited in scope, and firms can have heterogenous source of information about municipalities' creditworthiness, or different interpretation of factors related to bond values. Unless all firms confidently believe their own source or interpretation dominates their competitors', they will find rival signals informative, and auctions are closer to common values.

While the informational environment per se does not justify either PV or CV conclusively, data limitations also deter empirical efforts to discriminate between them. First, there is strong evidence that the number of potential bidders is correlated with bond values. This nullifies the test proposed by Haile, Hong and Shum (2003), which requires the variation in the number of bidders to be exogenous with respect to the distribution of values. Second, the data does not have ex post measures of bond values that can be used to test whether $v_h(x,x) = E(V_i|B_i = b_0(x_i), B_{-i} = b_0(x_{-i}))$ is independent from B_{-i}. Finally muni-bond auctions often proceed with no explicit reserve prices and therefore the testable restrictions in Hendricks, Pinkse and Porter (2003) are not useful.

This paper focuses on an incomplete approach for policy analyses by bounding counterfactual revenue distributions under general restrictions that encompass both PV and CV paradigms. In this context, distinguishing PV and CV only matters for interpreting the tightness of bounds. The lower bound is a point estimate of counterfactual distributions when values are private. On the other hand, if nothing is known about the values except for their interdependence, then for a given level of revenue, any point within the bounds can be rationalized by some value function and signal distributions.

1.6.3. Data description

The data contains all bids submitted in 6,721 auctions of municipal bonds on the primary market in the United States between 2004 and 2006. They are downloaded from auction worksheets at a website of *Thompson Financial*. The data are from the same source as those in Shneyerov (2006), but are more recent and the sample size is larger.

The data reports bond features including the identity of issuers, the sale date, the date of the first coupon, par values of each bond in a series, coupon rates of each bond, S&P and Moody's ratings of each bond, the type of government credit support for the issuance (general obligation or revenue),[37] and whether the issuance is bank-qualified.[38] It also includes macroeconomic variables that measure opportunity costs of investing in bonds and affect bond values for investors on secondary markets.

There are 97,936 bonds in 6,721 series, with an average of 14.5 for each issuance. About 70% of the series have 10 to 20 bonds. The average coupon rate of all bonds is 4.06% and average number of semiannual payments is 19.6. I use the par-weighted averages of coupon rates and numbers of coupon payments as a measure of "overall" interest rates and maturity for a series. About 90% of all issuances have a weighted average coupon rate between 3% and 5%. The weighted average maturity is approximately normally distributed with mean 20.8 and standard deviation 9.5. The total par of a series ranges from $0.1 million to $809 million, and is skewed to the right with mean $21.4 million and median $6 million. About 64.5%

[37]Bonds are categorized into two groups by the degree of credit support from municipalities. General obligation bonds are endorsed by the full faith and credit of the issuer, whereas revenue bonds promise repayment from a specified stream of future income, such as that generated by the public project financed by the issue. The latter usually bears higher interest rates due to risk premium.

[38]The Tax Reform Act of 1986 eliminated the tax benefits for commercial banks from holding municipal bonds in general. But exceptions were made for "bank-qualified" bonds, for which commercial banks can still accrue interests that are tax-exempt. Hence banks have a strong appetite for bank qualified bonds that are in limited supply, and bank qualified bonds carry a lower rate than non-bank qualified bonds.

of the series are backed by full credit of municipalities, while the rest are backed by limited municipal support, such as revenue stream from public works financed by the issuance.

In practice, issuers have the option to include reserve prices in the notice of sale, but few issuers use this option and the data does not report any reserve prices.[39] For each auction, the number of bidding coalitions, the number of companies within each coalition and their identities are all reported in the data. The number of syndicates ranges from 1 to 20, with mean 5.6 and standard deviation 2.6. Series that received more than 3 but fewer than 7 bids account for 68% of all auctions.

The dollar prices tendered are not always reported. However, total interest costs for all bids are always reported.[40] I use the following formula to calculate and impute missing dollar bids :

$$B = (1+TIC)^{-t_f} * \frac{\sum_{q=1}^{Q}\left(\sum_{t=0}^{T_q-1}\frac{C_q/2}{(1+\frac{TIC}{2})^t} + \frac{P_q}{(1+\frac{TIC}{2})^{T_q}}\right)}{\sum_{q=1}^{Q} P_q} * 100$$

where q indexes bonds in a series of Q bonds, T_q is the number of semi-annual periods from the date of first coupon until maturity, C_q and P_q are coupon and principal payments respectively, t_f is the time until first coupon payment and B is the dollar bid per $100 of face value.

[39]Shneyerov (2006) interpreted the bids as generated in equilibria with no binding reserve prices and estimated expected revenue in second-price auctions. In my paper I choose the same interpretation for our counterfactual analyses for revenue distributions.

[40]Total interest cost (TIC) is the interest rate that equates dollar prices with discounted present value of future cashflows the series.

Table 1(a) : Descriptive Statistics

# of bidding syndicates	2	3	4	5	6	7	8	9	10	11
# of auctions	608	971	1075	1007	852	687	531	406	258	158
Average par value ($mil)	4.90	9.22	17.33	23.13	27.02	27.27	30.20	32.68	26.54	33.57
Average price (/ $100 par)	99.17	99.31	99.51	99.97	100.32	100.44	101.00	101.06	101.19	101.64
Average spread	0.87	1.15	1.18	1.20	1.18	1.20	1.07	1.04	1.00	1.01
Average bid	98.73	98.77	98.95	99.42	99.79	99.94	100.53	100.59	100.76	101.22
Std. dev.	1.67	1.84	1.95	2.54	2.58	2.98	2.97	2.90	3.07	2.97
Minimal bid	90.28	85.20	86.47	87.03	86.68	13.26	93.44	91.67	93.94	93.87
Maximal bid	113.93	110.59	108.85	119.16	111.66	114.55	111.45	111.87	128.00	111.45

Table 1 (b) : Descriptive Statistics

Prices		All Bids		# of bidders	# of auctions	
Min	91.17	Min	85.20	1	19	0.28
Percentiles		Percentiles		2	608	9.05
1	96.74	1	95.32	3	971	14.45
10	97.98	10	97.37	4	1075	15.99
20	98.49	20	98.07	5	1007	14.98
30	98.91	30	98.56	6	852	12.68
40	99.29	40	98.99	7	687	10.22
50	99.66	50	99.40	8	531	7.90
60	100.00	60	99.81	9	406	6.04
70	100.33	70	100.28	10	258	3.84
80	100.91	80	101.14	11	158	2.35
90	102.84	90	103.54	12+	149	2.22
99	109.06	99	109.30	Total	6721	100.00
Max	128.00	Max	128.00			
# of auctions	6,721	# of bids	37,547			

WA Coupon Rate		Total Par Value	(in $million)	SecType		
Min	0.0100	Min	0.105	Unlimited GO	4334	64.484
1	0.0214	1	0.385	Limited GO	1061	15.786
10	0.0322	10	1.275	Revenue	1326	19.729
20	0.0355	20	2.160			
30	0.0377	30	3.200			
40	0.0392	40	4.485			
50	0.0405	50	6.000			
60	0.0419	60	8.581			
70	0.0435	70	12.000			
80	0.0454	80	20.415			
90	0.0481	90	45.000			
99	0.0549	99	297.831			
Max	0.0671	Max	809.470			

Table 1 summarizes the distribution of all 37,547 bids submitted in 6,721 auctions. The 1st percentile is $95.32 and the 99th percentile is $109.30. The median is $99.40, the mean

is \$99.92, and the standard deviation is \$2.76. The median winning dollar bid is \$99.66, the average is \$100.01, and the standard deviation is \$2.36.

1.6.4. Homogenization of bids

The data reports a wide variation in bond features. In competitive sales, syndicates take these characteristics into account in their bidding decisions, and thus strategies across auctions are not homogenous as the benchmark model posits. In principle bounds in the benchmark model still apply to subsets of homogenous auctions where bond features are held fixed. The main empirical challenge is that large samples for auctions with these specific features are needed for constructing nonparametric bounds on conditional revenue distributions. In this subsection, I tackle this issue by homogenizing bids across auctions with distinct features. The working assumptions are: *(i)* firm estimates of bond values are independent from publicly known bond features conditional on the number of participating syndicates; *(ii)* value functions are additively separable in private signals and bond features. Under these assumptions, the marginal effects of bond characteristics on equilibrium bids are identified. (I discuss a specification test of these restrictions below.) Thus bids in distinct auctions can be homogenized by removing differences due to variations in bond features as in Section 5.

In competitive sales with n bidding syndicates, ex ante bond values for a potential bidder is :

$$V_{il} = \mathbf{Z}_l'\boldsymbol{\gamma} + \theta_n(X_{il}, \mathbf{X}_{-il})$$

where $i = 1, .., n$ indexes the bidding syndicates, $l = 1, 2, ..L_n$ indexes auctions with n syndicates, $\mathbf{Z}_l$ is a vector of publicly known features, and $\mathbf{X}_l = (X_{il}, \mathbf{X}_{-il})$ is a $\mathbb{R}^n$-valued random vector of idiosyncratic signals. This specification reflects the intuition that marginal effects of idiosyncratic information (signals $\mathbf{X}_l$) may not interact with those of public information (bond features $\mathbf{Z}_l$).

Syndicates in an auction may differ in two aspects : the number of member firms, and local presence of firms' branch offices in the issuer's state. Recent empirical works suggest there is no conclusive evidence that they can lead to informational asymmetries.[41] Hence I maintain symmetry restrictions of θ_n and $F_{\mathbf{X}}$ as in the benchmark model in Section 2.

The equilibrium strategy is:

$$b_{il}(x_{il}, \mathbf{z}_l; n) = \mathbf{z}_l'\boldsymbol{\gamma} + \lambda_l(x_{il}, n) \tag{1.3}$$

where $\lambda(x,n) \equiv \int_{x_L}^{x} \phi_n(s) dL_n(s|x)$, $L_n(s|x) \equiv \exp\{-\int_s^x \frac{f_{Y|X,n}(u|u)}{F_{Y|X,n}(u|u)} du\}$, $\phi_n(s) \equiv E[\theta_N(X_i, \mathbf{X}_{-i})| X_i = \max_{j\neq i} X_j = s, N = n]$. Thus strategic bids can be decomposed into two additive components. The first term shows marginal effects of bond features are invariant to potential competitions, and the second term captures effects of potential competition on strategic bids. The signals and competitions interact with each other and their effects can not be separated. Regressing bids on bond features and a vector of dummies for the number of potential bidders will estimate $\boldsymbol{\gamma}$ consistently. That is, in the pooled regression,

$$b_{il}(x_{il}, \mathbf{z}_l) = \mathbf{d}_l'\boldsymbol{\delta} + \mathbf{z}_l'\boldsymbol{\gamma} + u_{il} \tag{1.4}$$

where $\mathbf{d}_l$ is a vector of dummies for n, the error term u_{il} is mean independent conditional on $\mathbf{d}_l$ and $\mathbf{z}_l$.[42]

[41] See Shneyerov (2006).

[42] To see this, fix n, then *Proposition 5* shows equilibrium bids are:

$$b_{il}(n) = \gamma_0(n) + \mathbf{z}_l'\boldsymbol{\gamma} + \varepsilon_{il}(x_{il}, n)$$

where $\gamma_0(n) \equiv E[\lambda_l(X_{il}, N_l)|N_l = n]$ and $\varepsilon_{il}(x_{il}, n) \equiv \lambda_l(x_{il}, n) - \gamma_0(n)$. It follows from the independence of $\mathbf{X}_l$ and $\mathbf{Z}_l$ conditional on number of bidders that $E[\varepsilon_{il}(X_{il}, N_l)|N_l = n, \mathbf{Z}_l = \mathbf{z}_l] = 0$ for all $(n_l, \mathbf{z}_l)$.

1.6.4.1. GLS estimates of index coefficients. When there is an intracluster correlation among error terms within auctions, a simple OLS will be inefficient. This can happen when syndicates' signals $\mathbf{X}_l$ are strictly affiliated. One explanation for affiliated signals in finance literature is the "herding" effect among research and sales staff across syndicates. For example, researchers in different syndicates tend to have similar professional backgrounds or trainings and hence are inclined to make similar decisions on the choice and weights of value-related factors in their analyses. Strict affiliation among signals could also happen when syndicates' estimates consist of idiosyncratic noisy measurements of a common, underlying random variable.

Table 2 : Pooled Random Effect Estimates

	Est	Std Err	t-stat	p-value
wacr	1.520	0.122	12.49	0.00
wapn	-1.037	0.061	-17.11	0.00
sectype	2.476	0.458	5.41	0.00
BQ	-0.837	0.056	-15.00	0.00
totpar	1.764	0.108	16.32	0.00
type_cr	-0.680	0.121	-5.64	0.00
HR	0.221	0.175	1.26	0.21
HR_pn	-0.002	0.067	-0.03	0.98
NE	0.428	0.142	3.00	0.00
SW	-0.188	0.188	-1.00	0.32
South	0.226	0.121	1.87	0.06
West	-0.323	0.149	-2.17	0.03
NE_rating	0.017	0.177	0.10	0.92
SW_rating	-0.038	0.231	-0.16	0.87
South_rating	0.309	0.175	1.77	0.08
West_rating	0.389	0.215	1.81	0.07
d3	94.920	0.450	210.97	0.00
d4	94.968	0.445	213.26	0.00
d5	95.323	0.438	217.61	0.00
d6	95.579	0.443	215.91	0.00
d7	95.738	0.438	218.44	0.00
d8	96.128	0.443	216.92	0.00

Number of cluster	5123.00
F(21, 5122)	86.66
Prob > F	0.00
R-squared	0.43
Root MSE	1.98

Table 2 reports the GLS estimates and t-statistics of γ for equation (1.4). The dependent variable is the dollar price bid. The regressors include publicly known bond features : weighted average coupon rate (*wacr*), weighted average maturity (*wapn*), total par value of the series (*totpar*), a dummy for whether the series is supported by full municipal credit (*sectype*), a dummy for whether the series is bank-qualified (*BQ*), a dummy for whether the series is rated with investment grade (*HR*) and two interaction terms *type_ cr* and *HR_ pn* respectively.[43] Butler (2007) suggests local presence of syndicates in the geographical area of the issuer could also influence their private information about the credibility of the issuer and hence their estimates of the value of the series. Therefore I also include in the regressors some dummies for the regions, *MW* (Midwest), *NE* (New England), *SW* (Southwest), *South* and *West*, to test the impact of geographic location on bids.

The weighted average coupon rates and maturity are both highly significant at 1% level, with positive and negative marginal effects respectively. These estimates confirm the intuition that bond values increase with cashflows from coupons and decrease as maturity increases because of higher interest rate and inflation risks. Municipality support has a significant positive effect on the bids. Controlling for other features, the average dollar price is $2.47 higher for bonds supported by municipalities' full credit. Bond ratings by *S&P* and *Moody's* have no significant impact on bids ceteris paribus. A possible explanation is that the syndicates' research forces do not consider ratings informative conditional on their own research on bond values. The dollar prices tendered for bank-qualified series are on average about 84 cents lower than non bank-qualified ones. The effect is statistically significant at 1% level. Besides, an increase of $1m in total par leads to a slight increase of 1.76 cents in

[43]The unit for *wapn* is 10 semin-annual coupon payments and the unit for *totpar* is $100 million.

the dollar price. This can be explained by the fact that average participation costs for a syndicate (e.g. time and effort on research) per \$100 in par is lower for issuance with larger par amount. The interaction of *sectype* and *wacr* are also highly significant at 1% level, suggesting marginal effects of coupon rates are lower for series with full municipal credit supports. There is no conclusive evidence for regional effects on bids except that dollar prices for series issued in New England area are higher on average than those issued in Midwestern states.

1.6.4.2. Specification tests. Two identifying restrictions in the regression equation (1.4) are additive separability and conditional independence of bond features and signals in value functions. A testable implication of these two restrictions is that marginal effects are constant and invariant to the number of potential bidders. That is, for each n, the following regression equation holds:

$$b_{il}(n) = \gamma_0(n) + \mathbf{z}_l' \boldsymbol{\gamma} + \varepsilon_{il}(x_{il}, n)$$

where $\gamma_0(n) \equiv E[\lambda_l(X_{il}, N_l)|N_l = n]$ and $\varepsilon_{il}(x_{il}, n) \equiv \lambda_l(x_{il}, n) - \gamma_0(n)$ is mean-independent conditional on $\mathbf{Z}_l$ and n. On the other hand, if either restriction is not satisfied, bidding strategies are nonseparable in $\mathbf{Z}_l$, X_{il} and n. Consequently, marginal effects of bond features on bids change with the number of potential bidders. Therefore we can test the two restrictions jointly by comparing estimates for auctions with different number of bidding syndicates.

Table 3(a) : GLS estimates for fixed number of potential bidders

number of bidders	4	5	6	7	8
Intercept	98.221 169.540	95.524 91.620	92.578 76.000	93.563 79.570	95.500 64.370
WA coupon rate	0.611 3.790	1.495 5.020	2.205 6.970	2.021 6.280	1.797 4.710
WA maturity	-0.896 -8.690	-1.082 -7.740	-1.025 -8.430	-0.944 -5.090	-1.278 -5.730
Security Type	0.264 0.390	3.297 3.110	4.514 3.960	3.945 3.370	2.180 1.400
Bank Qualified	-0.529 -5.510	-0.723 -6.580	-0.694 -5.330	-0.848 -5.820	-1.277 -6.110
Ratings	0.061 0.170	-0.115 -0.310	0.732 1.750	0.435 1.010	0.174 0.320
*Type*WACR*	-0.051 -0.290	-0.935 -3.330	-1.249 -4.250	-1.036 -3.380	-0.571 -1.430
*Ratings*WAPN*	0.113 0.900	0.219 1.500	-0.158 -1.210	-0.311 -1.530	0.080 0.330
Par amount	1.423 6.220	1.770 8.390	1.449 8.750	2.200 8.570	2.182 6.430
N.E.	0.065 0.280	0.836 2.810	0.905 2.670	0.149 0.280	0.380 0.860
South	-0.225 -1.130	-0.052 -0.220	0.747 2.260	0.382 0.830	0.369 1.060
S.W.	-0.638 -2.360	-0.443 -1.290	0.116 0.290	-0.209 -0.380	0.109 0.140
West	-0.807 -3.690	-0.066 -0.260	0.178 0.460	-0.600 -1.350	-0.364 -0.590
*NE*ratings*	0.104 0.350	-0.572 -1.590	-0.107 -0.260	0.578 1.010	-0.120 -0.210
*South*ratings*	0.506 1.570	0.794 2.290	0.048 0.110	0.349 0.680	-0.169 -0.320
*West*ratings*	0.943 2.700	0.298 0.660	0.016 0.030	1.037 1.760	0.010 0.010
*SW*ratings*	0.195 0.520	0.103 0.240	-0.181 -0.350	0.308 0.490	-0.446 -0.520
number of auctions	1075	1007	852	687	531
number of bids	4300	5035	5112	4809	4248
'R-square'	0.29	0.48	0.44	0.43	0.35
'F-statistic'	17.24	30.24	26.26	25.47	15.73
'p-value'	0.00	0.00	0.00	0.00	0.00

Table 3(a) reports GLS estimates in regressions for n between 4 and 8. The choice of regressors $\mathbf{z}$ is the same as that in (1.4). The estimates are consistent across n in signs and significance. For each significant characteristic of the series, *Table 3(b)* reports test statistics for pair-wise hypotheses that coefficients are the same in two regressions with different n. The statistics are constructed as the ratio of differences between GLS estimates and the standard error of the difference.[44] Under null hypotheses, the test statistics are asymptotically standard normal.

Table 3(b) : Test of equal indices

WA Coupon Rate

	4	5	6	7
4	-			
5	-1.924	-		
6	-3.336	-1.157	-	
7	-2.917	-0.849	0.289	-
8	-2.184	-0.445	0.585	0.318

WA Maturity

	4	5	6	7
4	-			
5	0.767	-		
6	0.575	-0.219	-	
7	0.167	-0.424	-0.263	-
8	1.171	0.538	0.733	0.816

Type

	4	5	6	7
4	-			
5	-1.746	-		
6	-2.339	-0.553	-	
7	-1.992	-0.290	0.246	-
8	-0.858	0.427	0.866	0.647

Par amount

	4	5	6	7
4	-			
5	-0.788	-		
6	-0.064	0.853	-	
7	-1.600	-0.920	0.030	-
8	-1.336	-0.750	-1.453	0.030

Bank Qualified

	4	5	6	7
4	-			
5	0.945	-		
6	0.732	0.853	-	
7	-1.600	-0.920	-1.779	-
8	-1.336	-0.750	-1.453	0.030

The results show differences between sizes of estimates are insignificant. With the exception of weighted average coupon rates for $n = 4$, all other estimates are not significantly different from their counterparts under a different n. There is no statistically significant evidence against the hypotheses that the value function is additively separable and bond

[44]Note GLS estimators for different n are independent, for (Z_l, N_l, X_l) are i.i.d. draws from the same joint distribution. Hence the standard deviation of the difference in two estimators can be consistently estimated by adding up their standard errors.

features have no bearing on the distribution of idiosyncratic signals conditional on the number of participating syndicates.

1.6.5. Results

1.6.5.1. Point and interval estimates for $\hat{F}^k_{R^I(r)}$ and $\hat{F}^k_{R^{II}(r)}$. This section reports bound estimates on counterfactual revenue distributions for a reference bond series in auctions with $n = 4$ bidding syndicates. The reference series is issued in the Midwest, bank-qualified, backed by full municipal credit, and has an investment grade from S&P and the Moody's. The reference series has a weighted average coupon rate of 4% and maturity of 5 years, as well as a total par of \$4.84 million.[45]

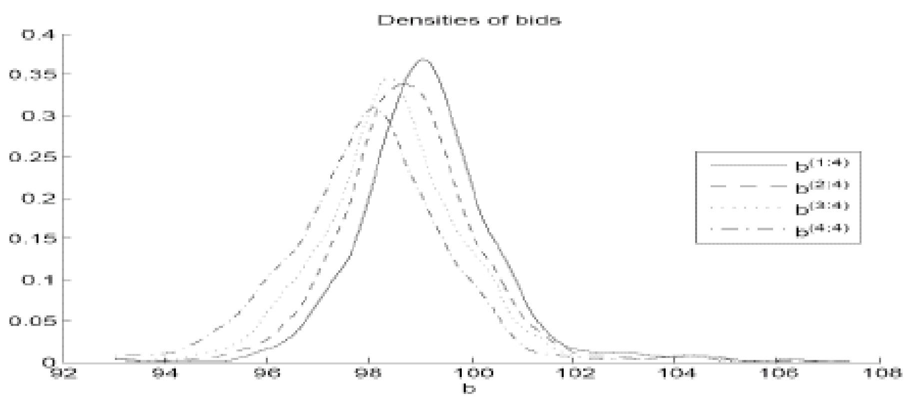


Figure 5(a)

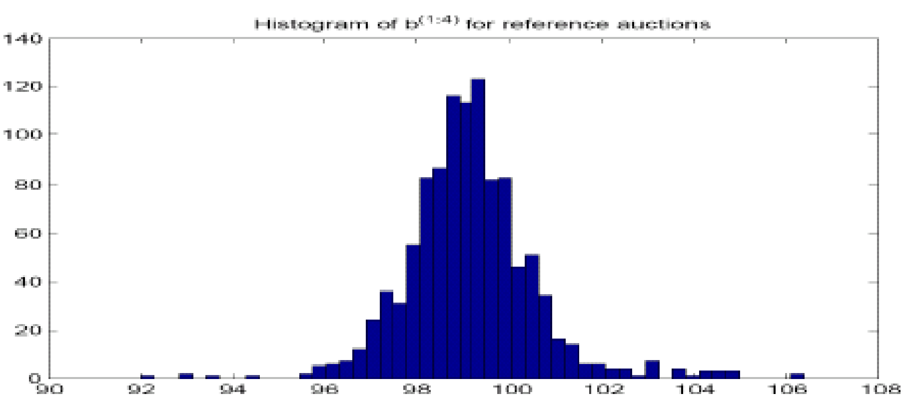


Figure 5(b)

[45]These are median values for series features among auctions with 4 bidding syndicates.

Figure 5(a) plots kernel density estimates of the ordered bids "homogenized" at the reference level, which are calculated using *GLS* estimates in regressions with 4 bidders. Distributions of the ordered bids are approximately normally distributed with similar standard deviations and the differences between the median of adjacent ordered bids are between \$0.25 and \$0.35 per \$100 in par amount. I use the product of tri-weight kernels for estimating $G_{M,B}$ and $g_{M,B}$. The choice of bandwidths follows "rule of thumb" discussed in Monte Carlo section.[46] The data is parse close to the both boundaries even after trimming bids that are within one bandwidth from the minimum and maximum bids reported. To avoid poor performances of the kernel estimates of $\hat{\xi}_l$ for lower dollar values, I trim the bids at the 0.5-th and 99.5-th percentile.[47] In the data, bids from the same auction are almost always trimmed together.

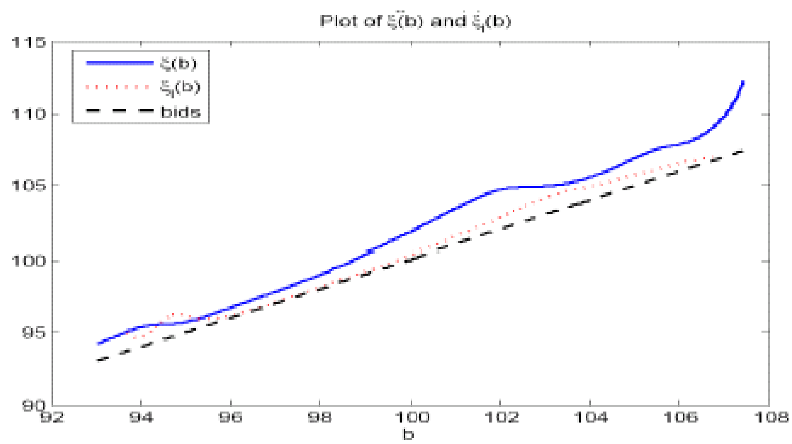


Figure 6

Figure 6 plots estimates $\hat{\xi}$ and $\hat{\xi}_l$ and suggests the estimated bounds on $b_0(x^*(r))$ only widen slowly as r increases. That $\hat{\xi}_l$ stays mostly above the 45-degree line is evidence for strict affiliations between private estimates within each auction. *Table 4* below summarizes

[46] The bandwidths h_G and h_g are respectively $2.98{*}1.06\hat{\sigma}_b{*}(4L_4)^{-\frac{1}{4n-5}} = 2.43$ and $2.98{*}1.06\hat{\sigma}_b{*}(4L_4)^{-\frac{1}{4n-4}} = 2.57$.

[47] The distance between the minimum bid and the 0.5-th percentile is about \$5. The number is greater than the smoothing parameter $h_g = 2.57$ used in the estimation.

estimated bounds on $b_0(x^*(r))$ and the probability that no one bids above r (hereafter referred to as the all-screening probability) for different reserve prices.

Table 4 : Estimated bounds on the all-screening probability

r	$\hat{b}_0(x_l(r))$	$\hat{b}_0(x_h(r))$	*b.w. of* $b_0(x^*(r))$	$\hat{F}^l_{R^I(r)}(r_-)$	$\hat{F}^u_{R^I(r)}(r_-)$
98	97.17	97.89	0.72	0.0540	0.1256
99	98.00	98.83	0.83	0.1488	0.3860
100	98.76	99.73	0.97	0.3609	0.6865
101	99.45	100.60	1.15	0.5935	0.8837
102	100.13	101.39	1.26	0.8074	0.9516
103	100.74	102.14	1.40	0.9042	0.9702

Table 4 suggests marginal bidders under r are estimated to bid lower than r in the scenario with no binding reserve price. It is consistent with the theoretical predictions in Section 2 that $F_{R^I(r)}(r)$ is smaller than $F_{R^I(0)}(r)$. The difference between the boundwidths of the all-screening probability for $r = 98$ and $r = 100$ is mostly due to the distribution of winning bids with no binding reserve prices. *Figure 5(b)* shows the distribution of $b_0^{(1:4)}$ has a larger mass in $[b_0(x_l(100))\ \ b_0(x_h(100))] = [98.75\ \ 99.73]$ than in $[b_0(x_l(98))\ \ b_0(x_h(98))] = [97.17\ \ 97.89]$. Therefore the bounds on the all-screening probability is much wider for $r = 100$ even though bounds on $b_0(x^*(100))$ is only slightly wider than those of $b_0(x^*(98))$.

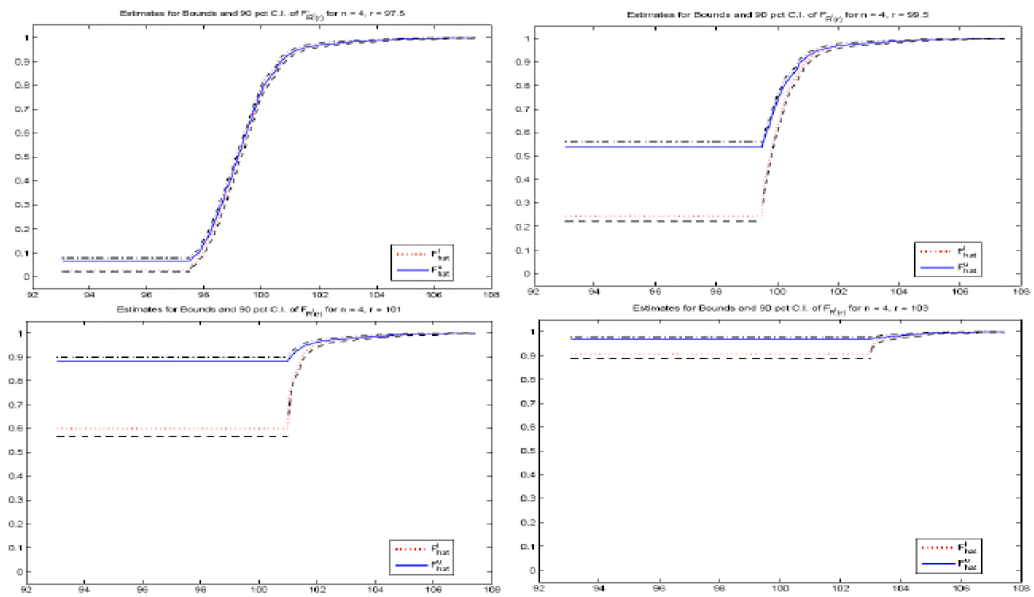


Figure 7

For reserve prices between \$98 and \$103, the solid and dotted lines in the panels of *Figure 7* depict point estimates $\hat{F}^u_{R^I(r)}$ and $\hat{F}^l_{R^I(r)}$ respectively. In addition, I construct 100 bootstrap samples, each containing 1075 auctions drawn with replacement from the estimating data. For all levels of revenue, I record the 5-th percentile of $\hat{F}^l_{R^I(r)}$ and 95-th percentile of $\hat{F}^u_{R^I(r)}$. They form a conservative, pointwise 90% confidence interval of $[F^l_{R^I(r)}, F^u_{R^I(r)}]$, and are plotted in *Figure 7* as broken lines. In addition, the table below reports the bounds on major percentiles according to the estimates of bounds on $\hat{F}^l_{R^I(r)}$ and $\hat{F}^u_{R^I(r)}$.

Table 5 : Estimated bounds on quartiles of $F_{R^I(r)}$

r	$l.b.$ 1st	$u.b.$ 1st	$l.b.$ 2nd	$u.b.$ 2nd	$l.b.$ 3rd	$u.b.$ 3rd
98	98.47	98.66	99.18	99.28	99.93	99.98
99	99.08	99.30	99.29	99.54	100.01	100.14
100	v_0	v_0	v_0	100.07	100.14	100.48
101	v_0	v_0	v_0	v_0	v_0	101.09

Revenue distribution above the reserve price depends on the distribution of $b_0(x)$ and the δ_r functional mapping $b_0(x)$ and $G^0_{\mathbf{B}}$ into $b_r(x)$. The densities plotted in *Figure 5 (a)* illustrate homogenized winning bids are approximately normally distributed. Besides, our estimates of bounds on δ_r are approximately linear. Therefore bound estimates $\hat{F}^k_{R^I(r)}(t)$ for $t > r$ increase at decreasing rates, a pattern similar to normal distributions. By construction, estimates of bounds on the all-screening probabilities are monotone in reserve prices (i.e. $\hat{F}^k_{R^I(r)}(r_-)$ is increasing in r for $k = l, u$). In addition, our estimates suggest that for any pair of reserve prices $r < r'$, $\hat{F}^k_{R^I(r')}(t) < \hat{F}^k_{R^I(r)}(t)$ for $t \geq r'$. This is consistent with the theoretical prediction that for a given signal above the screening level, firms bid less aggressively when the reserve price is lowered.

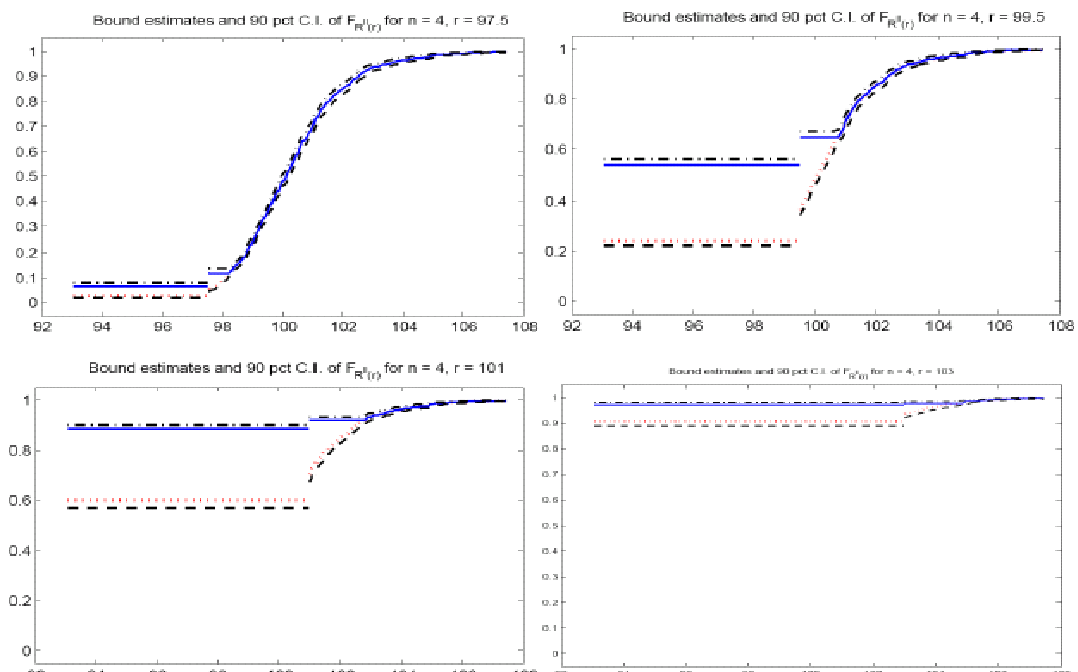


Figure 8

Likewise *Figure 8* plots point estimates for revenue distribution in second-price auctions as well as the 90% confidence intervals for $[F^l_{R^{II}(r)}(t),\ F^u_{R^{II}(r)}(t)]$.

1.6.5.2. Choice of optimal reserve prices. Knowledge of revenue distributions in counterfactual auctions enables the use of other distribution-based criteria for comparing auction revenues, instead of just expectations.[48] This is especially useful when the seller is known to be risk-averse and expected utilities are used as criteria.

A natural consequence of our partial approach is that only bounds on these criteria functions can be calculated. Such bounds on criteria functions are also tight and exhaust all information possible from equilibrium bids without further restrictions on value functions and signal distributions. As a result, answers to policy questions above involves comparing bound estimates rather than point estimates. Bounds on criteria functions can also be used to bound optimal reserve prices.[49]

A value for v_0 is needed for calculating both upper and lower bounds on $E(R^I(r))$ and $E(R^{II}(r))$. This should be measured by the amount of money that a municipality would be able to raise if it had borrowed through an alternative, next-cheapest channel (i.e. a creditor that requires the next lowest interests than syndicates participating in auctions). The proxy for v_0 used in this paper is \$95.71, and it is calculated as the present value per \$100 in par of cash flows from the coupon and principal payments of a reference bond, with the discount rate being the 99-th percentile of total interest rates reported in the data.

Figure 9(a) plots estimated upper and lower bounds on $E(R^I(r))$ (denoted $\hat{E}_h(R^I(r))$ and $\hat{E}_l(R^I(r))$ respectively), which are calculated from $\hat{F}^l_{R^I(r)}$ and $\hat{F}^u_{R^I(r)}$ through discretization and numerical integration using midpoint approximations. The solid lines plot $\hat{E}_k(R^I(r))$ and the dotted lines plot $\hat{E}_k(R^{II}(r))$. The upper bounds of expected revenue correspond to the case of PV auctions. Note estimates for $\hat{E}_h(R^{II}(r))$ are higher than $\hat{E}_h(R^I(r))$ for almost

[48]Within first-price auctions, each $r > v_0$ can be justified as optimal under the criterion of maximizing $\Pr(R^I(\tilde{r}) \geq r)$. That is $r = \arg\max_{\tilde{r} > v_0} \Pr(R^I(\tilde{r}) \geq r)$ for all $r > v_0$.
[49]For example, see Haile and Tamer (2003).

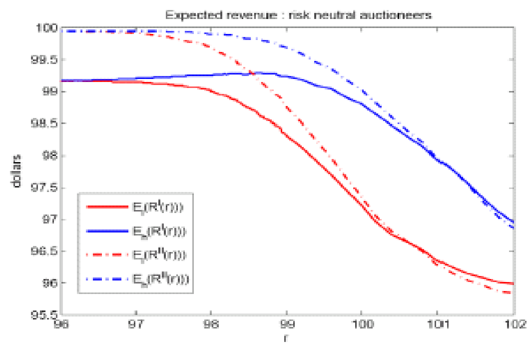


Figure 9(a)

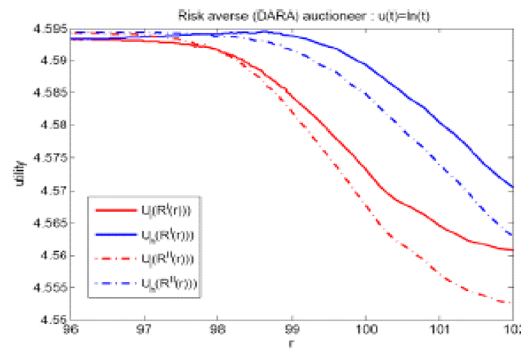


Figure 9(b)

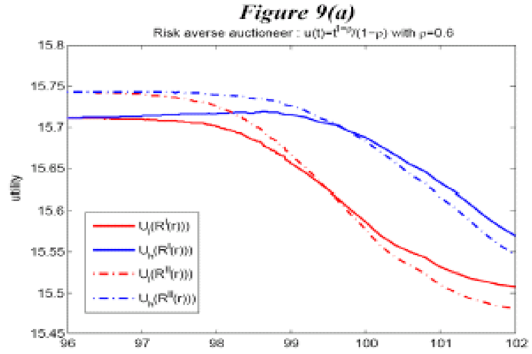


Figure 9 (c)

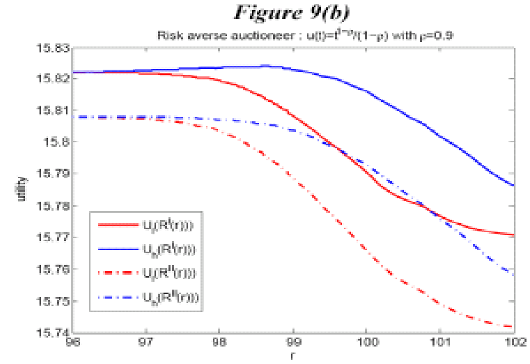


Figure 9(d)

all r in the range. This is consistent with the implication of Revenue Ranking Principle : for a fixed level of r, the expected revenue is higher for second-price auctions when signals are affiliated. For first-price auctions, $\hat{E}_h(R^I(r))$ is maximized at $r = \$98.68$ to be \$99.29, and $\hat{E}_l(R^I(r))$ is maximized at $r = \$96.26$ to be \$99.16. An argument similar to Haile and Tamer (2003) suggests the optimal reserve price that maximizes $E(R^I(r))$ must be in the range $[\$96.12, \$99.21]$. For second-price auctions, $\hat{E}_l(R^I(r))$ and $\hat{E}_h(R^I(r))$ are both maximized at $r = \$96.57$ with the maximum \$99.94, thus providing a point estimate for $E(R^I(r))$-maximizing reserve price. Instead of calculating a range of r that maximizes the expected revenue, an alternative is to pick r that maximizes either the lower or upper bound on $E(R^I(r))$. In the case of risk-neutral bidders, estimates for $\hat{E}_l(R^{II}(r))$, $\hat{E}_h(R^{II}(r))$ and $\hat{E}_l(R^I(r))$ are all close to being monotone, and their maximizers are all close to the boundary \$96.

A major motivation for focusing on revenue distribution in counterfactual analyses is the risk aversion of the seller. Given any specification of the seller's utility function (denoted $u(t)$), $\{\hat{F}^k_{R^j}(r)\}^{k=l,u}_{j=I,II}$ can be used to estimate bounds on the seller's expected utility (denoted $\{U_k(F_{R^j(r)})\}^{k=l,u}_{j=I,II}$). Like the case with a risk-neutral seller, these bounds can be used to put a range on an optimal reserve price that maximizes $U(F_{R^j(r)})$, or be used as criteria themselves for choosing reserve prices.

I consider three specifications of utilities: $u^{DARA}(t) = \ln(t)$ $(DARA)$ and $u^{CRRA}(t) = \frac{t^{1-\rho}}{1-\rho}$ with $\rho = 0.6$ and 0.9 $(CRRA)$. *Figure 9(b), (c)* and *(d)* plot estimated bounds on the expected utilities in first- and second-price auctions (i.e. $\{U_k(F_{R^j(r)})\}^{k=l,u}_{j=I,II}$) for *DARA*, *CRRA*($\rho = 0.6$) and *CRRA*($\rho = 0.9$) utility functions respectively.

Table 6 below summarizes reserve prices that maximize estimated bounds of expected utilities in first-price auctions, as well as estimated bounds on optimal r^* maximizing expected utilities.

Table 6 : Optimal reserve prices for first-price auctions

	$\hat{U}_l^{DARA}$	$\hat{U}_h^{DARA}$	$\hat{U}_l^{\rho=0.6}$	$\hat{U}_h^{\rho=0.6}$	$\hat{U}_l^{\rho=0.9}$	$\hat{U}_h^{\rho=0.9}$
$r_{(maximizer)}$	96.19	98.65	96.23	98.68	96.25	98.66
maximum	4.593	4.594	15.711	15.719	15.822	15.824
bounds on r^*	$[96.24, 99.20]$		$[96.17, 99.32]$		$[96.21, 99.20]$	

In second-price auctions, estimates of bounds on expected utilities under different specifications are all maximized at \$96.26, with the maxima being 4.594, 15.743 and 15.808 respectively. As a result, we get a point estimate of the optimal reserve price r^* at \$96.26 for all three specifications.

Both maximizers across different specifications of utility functions are close to each other and so are the interval estimates. This is because u^{DARA}, $u^{\rho=0.6}$ and $u^{\rho=0.9}$ are all approximately linear for the range of revenues considered in this application. As a result, estimated bounds on $\{U(F_{R^j(r)})\}_{j=I,II}$ as functions of r are close to being linear transformations of each other.

On the other hand, estimates for different $u(.)$ yield different implications regarding the choice of format between first- and second-price auctions. For $DARA$ utility functions, the point estimate for the optimal reserve price in second-price auctions is \$96.26, with a maximum $\hat{U}^{DARA}(F_{R^{II}(96.26)}) = 4.594$. This is equal to the maximized value for $\hat{U}_h^{DARA}(F_{R^I(98.65)})$. Hence estimates suggests a seller with decreasing absolute risk aversion should prefer second-price auctions in general, and may be indifferent between the two formats if the auction is known to belong to the PV paradigm. For $CRRA$ utilities with $\rho = 0.6$, the implication is the same as in the case with risk-neutral sellers. However, for $CRRA$ utilities with $\rho = 0.9$, estimates suggest first-price auctions should be preferred over second-price ones. The pattern is due to the fact that $F_{R^I(r)}$ always crosses $F_{R^{II}(r)}$ from below for any given r, and $u^{\rho=0.6}$ increases faster than $u^{\rho=0.9}$.

Finally a technical note is in order. Except for $\hat{E}_h(R^I(r))$ and $\hat{U}_h(R^I(r))$, other estimates of bounds on $\{E(R^j(r))\}_{j=I,II}$ and $\{U(R^j(r))\}_{j=I,II}$ are almost monotonically decreasing in r. In general this need not be the case in estimation. To see this, note that none of the estimates $\{\hat{F}^k(R^j(r))\}_{j=I,II}^{k=l,h}$ reported in *Figure 7* and *Figure 8* are stochastically ordered in r. In this incidence, the monotonicity is explained by the fact that our measure of v_0 is low at \$95.71 and that estimates $\hat{b}_r(x_h(r'))$ are close to r' for all (r, r').

1.7. Conclusion

In structural models of first-price auctions, interdependence of bidders' values leads to non-identification of model primitives. That is, distributions of equilibrium bids observed in a given auction format can be rationalized by more than one possible specifications of signal and value distributions. While this negative identification result rules out policy analyses that rely on exact knowledge of primitives, the distribution of bids observed in equilibria should still convey useful information about primitives that can be extracted for counterfactual revenue analyses. Following this line of reasoning, this paper derives bounds on revenue distributions in counterfactual auctions with binding reserve prices by using equilibrium conditions. The bounds are the tightest possible under restrictions of interdependent values and affiliated signals, and can be used to compare auction formats or bounds on optimal reserve prices. This approach also addresses the empirical difficulty of differentiating PV and CV paradigms in policy analyses. The bounds can be nonparametrically consistently estimated, and Monte Carlo evidence suggests these estimators also have reasonable finite sample performances. Observed heterogeneity in auction characteristics can be controlled for by conditioning counterfactual analyses on these auction features. Under the restriction of additive separability of signals and auction characteristics in value functions, the marginal effects of auction features can be identified if signals are independent from auction features conditional on the number of bidders. By removing variations due to observable auction heterogeneity, the bids across various auctions can be "homogenized" to bids in auctions with given specific features. The issue of endogenous participation also does not pose major challenges to the construction of bounds, provided the data report the number of potential bidders or good proxies of this number.

Applying this methodology to U.S. municipal bond auctions on the primary market yields informative bound estimates of revenue distributions in counterfactual auctions with binding reserve prices. These estimates are then used to bound the reserve prices that maximize expected revenues for risk-neutral sellers. For risk-averse sellers, bounds on revenue distributions are also used to bound optimal reserve prices which maximize their expected utility under different specifications of utility functions.

Directions for future research include extensions of partial-identification methods for more complicated cases such as asymmetric information among bidders and unobserved auction heterogeneity.

CHAPTER 2

Semiparametric Estimation of Binary Response Models with Inequality Quantile Restrictions

2.1. Introduction

Binary choice models have been used widely in empirical research in fields such as industrial organization and labor economics. In such models, the decision-maker chooses an action out of two alternatives if and only if its payoff is higher than the other. The payoffs is determined by observable state variables (or regressors) and disturbances (or errors) unobservable to researchers. Researchers are interested in using choice data to make inference about structural parameters in payoff the function as well as error distributions.

For the past three decades, econometricians have studied the estimation of binary choice models under various restrictions on payoff functions and distribution of the errors. Among them, a most popular identifying assumption is the statistical independence between errors and regressors. Matzkin (1992) showed with the independence assumption that the payoff function $\mathbf{u}$ and the error distribution F_{ε} can be uniquely recovered from choice probabilities under fairly general form restrictions on $\mathbf{u}$ (such as monotonicity, concavity and homogeneity). Other authors studied the estimation of binary response models under statistical independence but with different form restrictions on the payoff functions (see Cosslett (1983), Han (1987), Klein and Spady (1993), and Ichimura (1998)). Another strand of literature studies binary response models under a weaker assumption that the median of errors is independent from regressors. This restriction allows for endogenous regressors, which are a concern in lots of empirical work. Manski (1985) showed the linear coefficients can be identified up to scale under median independence and fairly weak assumptions on regressors, and proposed a consistency maximum score estimator. Other authors have studied the asymptotic distribution and the refinement of maximum score estimators (see Sherman (1988) and Horowitz (1992)). Furthermore, Manski (1988) suggested median independence has the most identifying power among all stochastic restrictions that allow for the correlation between regressors and unobserved disturbances.[1]

In this paper, I study a class of binary response models where the conditional median of errors is bounded between known functions of the regressors. This generalization is meaningful because it encompasses several interesting micro-econometric sub-models widely applied in empirical work. As shown in Section 2, our specification of the binary response model with bounded conditional medians is general enough to incorporate binary response models with interval data on regressors, simultaneous discrete games with incomplete information, and Markovian binary choice processes. I characterize the identification region of linear coefficients in payoff functions using choice probabilities observed, and derive fairly general restrictions on the distribution of regressors that are sufficient for point identification. I discuss how these conditions can be satisfied by more primitive conditions in the motivating sub-models mentioned above. I then use the sample analog principle to define a two-step extreme estimator based on the form of the identification region, with the first-step being a kernel regression that estimates the conditional choice probabilities. I also show that this two-step extreme estimator is consistent regardless of whether the coefficients are point identified, and it converges in distribution to a normal random variable at the rate of $\sqrt{n}$ under

[1]Manski (1988) showed (1) mean independence has no identifying power in the binary response model; (2) conditional symmetry has no additional identifying power than median independence; (3) distributional index sufficiency can only identify the slope coefficients up to scale and sign.

point identification. Finally, I give Monte Carlo evidence on the estimator's performance in finite samples when the model is partially identified.

The rest of the paper is organized as follows. Section 2 specifies the binary response model with bounded conditional medians and give examples of motivating sub-models. Section 3 and 4 studies the set and point identification of the index coefficients in payoff functions respectively. Section 5 defines the two-step extreme estimator and proves its asymptotic properties, including consistency and asymptotic normality under point identification. Section 6 show Monte Carlo performance of the estimator in finite samples. Section 7 concludes.

2.2. The Model

Consider a binary choice model:[2]

$$Y = 1(\mathbf{X}'\boldsymbol{\beta} + \varepsilon \geq 0),\ \boldsymbol{\beta} \in \mathbb{R}^K,\ \boldsymbol{\beta} \neq 0 \tag{2.1}$$

where the conditional median of ε is defined as:

$$Med(\varepsilon|\mathbf{X}) = \{\eta \in \mathbb{R} : \Pr(\varepsilon \geq \eta|\mathbf{X}) \geq \frac{1}{2} \wedge \Pr(\varepsilon \leq \eta|\mathbf{X}) \geq \frac{1}{2}\}$$

Let $S(\mathbf{X})$ denote the support of $\mathbf{X}$ and $F_{\mathbf{X}}$ denote a probability measure on $S(\mathbf{X})$. The error distribution satisfies the following stochastic restriction.

BCQ (Bounded Conditional Median): The error ε is has continuous density conditional on all $\mathbf{x}$ with $L(\mathbf{x}) \leq \sup Med(\varepsilon|\mathbf{x})$ and $\inf Med(\varepsilon|\mathbf{x}) \leq U(\mathbf{x})$ a.e.$F_{\mathbf{X}}$, where $L(.), U(.)$

[2]Throughout the paper I use bold letters for vectors and non-bold letters for scalars, upper cases for random variables and lower cases for their realizations.

are known functions such that $L(\mathbf{x}) \leq U(\mathbf{x})$ *a.e.*$F_{\mathbf{X}}$ *and* $L \equiv \inf_{\mathbf{x}\in S(\mathbf{X})} L(\mathbf{x}) \geq -\infty$, $U \equiv \sup_{\mathbf{x}\in S(\mathbf{X})} U(\mathbf{x}) \leq +\infty$.[3]

The inequality restriction is violated if and only if there is no median in the interval $[L(\mathbf{x}), U(\mathbf{x})]$. This restriction can be rewritten as: $Med(\varepsilon|\mathbf{x}) \cap [L(\mathbf{x}), U(\mathbf{x})] \neq \varnothing, a.e.F_{\mathbf{X}}$. Obviously the model is general enough to allow for error distributions that are not strictly monotone and that may have an set-valued median. In addition, this setup is general enough to include several interesting micro-econometric models as special cases.

Model 1 *(Partially linear binary choice)* Let $Y = 1(\mathbf{X}'\boldsymbol{\beta} + g(\mathbf{X}) + \varepsilon \geq 0)$ where $Med(\varepsilon|\mathbf{X} = \mathbf{x}) = 0$ and $L(\mathbf{x}) \leq g(\mathbf{x}) \leq U(\mathbf{x})$ for all $\mathbf{x}$ on the support of $\mathbf{X}$ for some known functions $L(.)$ and $U(.)$. This model implies: $Y = 1(\mathbf{X}'\boldsymbol{\beta} + \tilde{\varepsilon} \geq 0)$, where $L(\mathbf{x}) \leq \inf Med(\tilde{\varepsilon}|\mathbf{X} = \mathbf{x}) = \sup Med(\tilde{\varepsilon}|\mathbf{X} = \mathbf{x}) \leq U(\mathbf{x})$ $\forall \mathbf{x} \in S(\mathbf{X})$. An empirical example of this binary choice model with a partially linear latent variable is individual decisions for labor participation. Suppose each individual works if and only if his monthly salary is greater than his unemployment benefits, and both are solely determined by demographic characteristics $\mathbf{X}$ (including gender, education, experience, etc). The median of monthly salary conditional on $\mathbf{X}$ is $\mathbf{X}'\boldsymbol{\beta}$, while the unemployment benefits is given by $g(\mathbf{X})$. Researchers observe individuals' decision to participate in the labor force and are interested in recovering $\boldsymbol{\beta}$, but only knows that unemployment benefits $g(.)$ is bounded between $L(.)$ and $U(.)$.

Another special case of this model is a binary choice model with interval data on a regressor studied in Manski and Tamer (2002). Let $Y_i = 1(\mathbf{X}'\boldsymbol{\beta} + V + \varepsilon \geq 0)$, where $\mathbf{X} \in \mathbb{R}^k, V \in \mathbb{R}$, and $Med(\varepsilon|\mathbf{x}, v) = 0$ $\forall(\mathbf{x}, v)$ on support. Researchers observe a random sample of $(Y, \mathbf{X}, V_0, V_1)$ and (i) $\Pr(V_0 \leq V \leq V_1) = 1$ and both V_0 and V_1 are bounded; (ii) $Med(\varepsilon|\mathbf{x}, v_0, v_1) = 0$ $\forall(\mathbf{x}, v_0, v_1)$. Then $Y = 1(\mathbf{X}'\boldsymbol{\beta} + \tilde{\varepsilon} \geq 0)$ where $\tilde{\varepsilon} = V + \varepsilon$. It follows from

[3] The continuity of the distribution of ε is a technical convenience that can be weakened to $\sup Med(\varepsilon|\mathbf{x}) \in Med(\varepsilon|\mathbf{x})$.

(i) and (ii) that $v_0 \leq \inf Med(\tilde{\varepsilon}|\mathbf{x}, v_0, v_1) \leq \sup Med(\tilde{\varepsilon}|\mathbf{x}, v_0, v_1) \leq v_1 \; \forall(\mathbf{x}, v_0, v_1)$. Denote the $(k+2)$-vectors $[\mathbf{X} \; V_0 \; V_1]$ by $\mathbf{Z}$ and $[\boldsymbol{\beta} \; 0 \; 0]$ by α. Then the model is reformulated as $Y = 1(\mathbf{Z}'\boldsymbol{\alpha} + \tilde{\varepsilon} \geq 0)$, where $L(\mathbf{Z}) \leq \inf Med(\tilde{\varepsilon}|\mathbf{Z}) \leq \sup Med(\tilde{\varepsilon}|\mathbf{Z}) \leq U(\mathbf{Z}) \; a.e.F_{\mathbf{Z}}$ with $L(\mathbf{Z}) = V_0$ and $U(\mathbf{Z}) = V_1$. The parameter space now considered is $\Theta = \{\mathbf{b} \in \mathbb{R}^{k+2} : b_{k+1} = b_{k+2} = 0\}$. Thus the model fits in with our framework of binary regressions with bounded conditional medians.[4] ■

Model 2 *(Simultaneous Discrete games with incomplete information)* Consider a simple, simultaneous 2-by-2 discrete game with the same space of pure strategies $S_i = \{1, 0\}$ for players $i = 1, 2$. The payoff structure is :

	0	1
0	$0, 0$	$0, \mathbf{x}'\boldsymbol{\beta}_2 - \varepsilon_2$
1	$\mathbf{x}'\boldsymbol{\beta}_1 - \varepsilon_1, 0$	$\mathbf{x}'\boldsymbol{\beta}_1 + \delta_1 - \varepsilon_1, \mathbf{x}'\boldsymbol{\beta}_2 + \delta_2 - \varepsilon_2$

where $\mathbf{x} \in \mathbb{R}^K$ is a vector of payoff-related exogenous variables observed by both players, $\varepsilon \equiv (\varepsilon_1, \varepsilon_2)$ are private signals only observable to player i with jointly distribution $F_{\boldsymbol{\varepsilon}}$, which is common knowledge among the players. Furthermore ε_1 is independent from ε_2 conditional on x. The structural parameters of the model is $\Theta \equiv (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \delta_1, \delta_2)$, where $\delta_i < 0$ for $i = 1, 2$. A Bayesian Nash Equilibrium (BNE) of this game of incomplete information is defined by $p(x) \equiv [p_1(x) \; p_2(x)]$ such that

$$\begin{bmatrix} p_1(\mathbf{x}) \\ p_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} F_{\varepsilon_1|\mathbf{X}=\mathbf{x}}(\mathbf{x}'\boldsymbol{\beta}_1 + p_2(\mathbf{x})\delta_1) \\ F_{\varepsilon_2|\mathbf{X}=\mathbf{x}}(\mathbf{x}'\boldsymbol{\beta}_2 + p_1(\mathbf{x})\delta_2) \end{bmatrix} \tag{2.2}$$

[4]As discussed in the section below, Manski and Tamer (2002) provides sufficient conditions on the support of $[V_0, V_1]$ for $\boldsymbol{\beta}$ to be point identified, and proposes a modified maximum score estimator for the identification region that is consistent under the Hausdorff metric.

where $p_i(x)$ is player i's probability of choosing 1 conditional on $X = x$.[5] The existence of BNE follows from Brouwer's Fixed Point Theorem, and Aradillas-Lopez (2007) gives fairly general sufficient and necessary conditions for the equilibrium to be unique.[6] It can be shown that a generic parameters Θ will generate $p(\mathbf{x})$ if and only if it generates $p_i(\mathbf{x})$ in the single-agent binary choice model $Y_i = 1(x'\beta_i + p_{-i}(x)\delta_i - \varepsilon_i \geq 0)$. Suppose $Med(\varepsilon_i|\mathbf{X}) = 0$ for $i = 1, 2$. The δ_i's need to be normalized to -1 for identification. Then the binary choice models fall under our general framework with $L_i(\mathbf{x}) = p_{-i}(\mathbf{x}) = U_i(\mathbf{x})$.

Several recent literature have discussed the estimation of such static games with incomplete information. These include Aradillas-Lopez (2007) and Bajari, Hong, Krainer and Nekipelov (2007). The latter shows the mean utility functions can be identified nonparametrically if the error distributions are i.i.d. across players for any given $\mathbf{x}$, and if the conditional error distribution $F_{\varepsilon_1,\varepsilon_2|\mathbf{X}}$ is completely known to the researcher. Aradillas-Lopez focuses on a case where $(\varepsilon_1, \varepsilon_2)$ are allowed to be correlated with each other but have to be independent from $\mathbf{X}$. He extends the semiparametric likelihood estimator in Klein and Spady (1993) to this game theoretic setup. In contrast, our approach estimates this game of incomplete information through the framework of a system of binary choice models with bounded conditional medians, and does not require the independence between $(\varepsilon_1, \varepsilon_2)$ and exogenous variables $\mathbf{X}$ that shift the payoff structures. ■

Model 3 *(Markovian binary choice process)* Consider a single-agent, Markovian binary choice process in infinite horizon. Time is discrete and indexed by t. In each t, the agent observes states $\mathbf{s}_t = (\mathbf{x}_t, \varepsilon_t) \in S(\mathbf{X}) \otimes S(\epsilon) \subseteq \mathbb{R}^K \otimes \mathbb{R}^2$, and chooses an action d_t from a

[5] This definition of Bayesian Nash Equilibrium is the same as that of Quantal Response Equilibrium in McKinley and Palfrey (1995). The latter is a special case of BNE where the error distributions are independent across actions.

[6] If researchers know a priori these conditions are not satisfied, then it is convenient to discuss identification by maintaining that players stick to an equilibrium selection mechanism that is solely determined given $\mathbf{x}$.

pair of possible actions $D = \{0,1\}$. The state space $S(\mathbf{S}) = S(\mathbf{X}) \otimes S(\epsilon)$ is time-invariant. In each t, researchers observe $\mathbf{x}_t$, but not ε_t. The time-invariant single-period return is $U(s_t, d_t) : S(\mathbf{S}) \otimes D \rightarrow \mathbb{R}^1 \ \forall \ t$. Conditional on the current state s and action d, the distribution of the state next period $\mathbf{s}'$ is given by a time-invariant transitional probability $p_d(.|\mathbf{s}) : S(\mathbf{S}) \rightarrow [0,1]$. The agent has a constant discount factor $\beta \in (0,1)$ for all periods.[7] The agent chooses a deterministic, Markovian decision rule $d(\mathbf{s})$ that maximizes expected present value of future utilities: $E[\sum_{j=0}^{\infty} \beta^j U(\mathbf{s}_{t+j}, d_{t+j})|\mathbf{s}_t, d_t]$.[8] The structure satisfies two restrictions (i) the single-period return is additively separable, i.e. $U(\mathbf{s}, d) = \mathbf{x}'\boldsymbol{\alpha}_d + \varepsilon_d$, $\varepsilon \in \mathbb{R}^2$, $E(\varepsilon_d|x) = 0 \ \forall(\mathbf{x}, d)$; and (ii) the transitional probabilities has conditional independence, i.e. $p_d(\mathbf{s}'|\mathbf{s}) = f(\varepsilon'|\mathbf{x}')g_d(\mathbf{x}'|\mathbf{x}) \ \forall \mathbf{s}, \mathbf{s}' \in S(\mathbf{S})$ and $d = 0, 1$. These and some other regularity conditions implies the Markovian binary decision process has a static representation:[9]

$$d(\mathbf{s}) = \arg\max_{d \in \{0,1\}} \delta_d(\mathbf{x}; \mathbf{u}, \mathbf{g}, f) + \varepsilon_d$$

where $\delta(\mathbf{x}) = [\delta_0(\mathbf{x}) \ \delta_1(\mathbf{x})]^T$ is the fixed point of the operator

$$T \circ \begin{pmatrix} \delta_0(\mathbf{x}) \\ \delta_1(\mathbf{x}) \end{pmatrix} \equiv \begin{pmatrix} \mathbf{x}'\boldsymbol{\beta}_0 + \beta \int \max_{d' \in \{0,1\}} \{\delta_{d'}(\mathbf{x}') + \varepsilon'_{d'}\} p_0(d\mathbf{s}'|\mathbf{x}) \\ \mathbf{x}'\boldsymbol{\beta}_1 + \beta \int \max_{d' \in \{0,1\}} \{\delta_{d'}(\mathbf{x}') + \varepsilon'_{d'}\} p_1(d\mathbf{s}'|\mathbf{x}) \end{pmatrix}. \tag{2.3}$$

Aguirregabiria (2007) showed through recursive substitution, $\delta_d(\mathbf{x}) = \omega_d(\mathbf{x}) + \xi_d(\mathbf{x})$ where $\omega_d(\mathbf{x}) = \mathbf{x}'\alpha_d + \beta \int \omega_d(\mathbf{x}')p_d(d\mathbf{x}'|\mathbf{x})$ and $\xi_d(\mathbf{x}) = \beta \int \kappa_d(\mathbf{x}') + \xi_d(\mathbf{x}')p_d(d\mathbf{x}'|\mathbf{x})$, and $\kappa_d(.)$ are

[7]For notational ease, I will drop time subscripts for the rest of the paper due to time-invariance of period return, transitional probabilities, and the state and action spaces.

[8]In general, the optimal policies should be a function of past histories $\mathbf{H}_t = \{\mathbf{s}_j\}_{j=0}^{t}$. However, Strauch (1966) showed for any history-dependent poilicy and starting state, there always exists a deterministic, Markovian policy (a policy that depends on the current state only) with the same expeced total discounted payoff. The implication is that for analysis of optimal policies, it suffices to focus on Markovian stationary policies. Throughout the paper we focus on the case where the agent only considers deterministic Markovian policies.

[9]The regularity conditions include continuity and boundedness of $u_d(\mathbf{x})$, finite expectation of $\max\{\varepsilon_1, \varepsilon_0\}$ conditional on $(\mathbf{x}, d)$, and that g_d satisfies the Feller Property for $d = 0, 1$.

defined as

$$\begin{aligned}\kappa_1(\mathbf{x};F_{\Delta\varepsilon|\mathbf{X}},h) &= h(\mathbf{x})F^{-1}_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}}(h(\mathbf{x})) - \int_{-\infty}^{F^{-1}_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}}(h(\mathbf{x}))} \Delta\varepsilon dF_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}} \\ \kappa_0(\mathbf{x};F_{\Delta\varepsilon|\mathbf{X}},h) &= \int_{F^{-1}_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}}(h(\mathbf{x}))}^{+\infty} \Delta\varepsilon dF_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}} - (1-h(\mathbf{x}))F^{-1}_{\Delta\varepsilon|\mathbf{X}=\mathbf{x}}(h(\mathbf{x}))\end{aligned}$$

where $h(\mathbf{x})$ is the observed choice probability. Let Γ denotes a set of $F_{\Delta\varepsilon|\mathbf{X}}$ that satisfies some extraneous restrictions (such as bounded support and symmetry), and define:

$$L(\mathbf{x};h,p) \equiv \min_{F_{\Delta\varepsilon|\mathbf{X}}\in\Gamma} \Delta\xi(\mathbf{x};h,p,F_{\Delta\varepsilon|\mathbf{X}}),\ U(\mathbf{x};h,p) \equiv \max_{F_{\Delta\varepsilon|\mathbf{X}}\in\Gamma} \Delta\xi(\mathbf{x};h,p,F_{\Delta\varepsilon|\mathbf{X}})$$

where $\Delta\xi = \xi_1 - \xi_0$ and the existence and the form of extrema is delivered by the nature of the restricted set Γ. The Markovian binary choice process can then be represented by a static analog:

$$Y = 1\{\tilde{\mathbf{x}}'\tilde{\boldsymbol{\gamma}} + \Delta\xi(\mathbf{x}) - \Delta\varepsilon \geq 0\} \tag{2.4}$$

where $\tilde{\mathbf{x}}_t \equiv [\sum_{s=0}^{+\infty}\beta^s E_0(\mathbf{x}_{t+s}|\mathbf{x}_t), \sum_{s=0}^{+\infty}\beta^s E_1(\mathbf{x}_{t+s}|\mathbf{x}_t)]$, and $\tilde{\gamma} \equiv [\boldsymbol{\alpha}_0, -\boldsymbol{\alpha}_1]$. Under the median independence restriction $Med(\Delta\varepsilon|\mathbf{X})$ =0, the static representation in (2.4) fits within the framework of binary choice models with inequality conditional medians. ■

2.3. Partial Identification of $\boldsymbol{\beta}$

Let Γ denote the set of conditional distributions $F_{\varepsilon|\mathbf{X}}$ that satisfy BCQ, and $P^*_{1|\mathbf{x}}$ denote observed conditional choice probability $\Pr(d=1|\mathbf{x})$. In this section I characterize the set of coefficients $\mathbf{b} \in \mathbb{R}^K$ which, for some choice of $F_{\varepsilon|\mathbf{X}} \in \Gamma$, can generate the observed choice probabilities $P^*_{1|\mathbf{x}}$ almost everywhere on the support of $\mathbf{X}$ (denoted $S(\mathbf{X})$). This reveals the limit of what can be learned about the true parameter $\boldsymbol{\beta}$ from observables under BCQ, and leads to the definition of our two-step extreme estimator. For any generic pair of coefficient $\mathbf{b}$ and conditional error distribution $G_{\varepsilon|\mathbf{X}}$, let $P_{1|\mathbf{x}}(\mathbf{b}, G_{\varepsilon|\mathbf{X}})$ denote the probability of choosing $d = 1$ given $\mathbf{x}$, $\mathbf{b}$ and $G_{\varepsilon|\mathbf{X}}$ (i.e.$P_{1|\mathbf{x}}(b, G_{\varepsilon|\mathbf{X}}) \equiv \int 1(\mathbf{x}'\mathbf{b} + \varepsilon \geq 0) dG_{\varepsilon|\mathbf{X}=\mathbf{x}})$, and let $X(b, G_{\varepsilon|\mathbf{X}})$ denote the set $\{x \in S(\mathbf{X}) : P_{1|\mathbf{x}}(\mathbf{b}, G_{\varepsilon|\mathbf{X}}) \neq P^*_{1|\mathbf{x}}\}$.

Definition 1 *The true coefficient* $\boldsymbol{\beta}$ *is identified relative to* $\mathbf{b}$ *if* $\forall\ F_{\varepsilon|\mathbf{X}} \in \Gamma$ $\Pr(x \in X(\mathbf{b}, F_{\varepsilon|\mathbf{X}})) > 0$*. Furthermore,* $\boldsymbol{\beta}$ *is observationally equivalent to* $\mathbf{b}$ *if it is not identified relative to* $\mathbf{b}$*. The identification region of* $\boldsymbol{\beta}$ *is the set of* $\mathbf{b}$ *in* $\mathbb{R}^K$ *that is observationally equivalent to* $\boldsymbol{\beta}$*.*

Lemma 1 *In Model (2.1) under BCQ,* $\mathbf{b}$ *is observationally equivalent to* $\boldsymbol{\beta}$ *if and only if* $\Pr(x \in \xi'_{\mathbf{b}}) = 0$*, where* $\xi'_{\mathbf{b}} \equiv \{\mathbf{x} \in S(\mathbf{X}) : (-\mathbf{x}'\mathbf{b} \leq L(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} < \frac{1}{2}) \vee (-\mathbf{x}'\mathbf{b} \geq U(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} > \frac{1}{2})\}$*.*

An immediate implication of *Lemma 1* is that the identification region under *BCQ* is $\Theta'_I \equiv \{\mathbf{b} \in \mathbb{R}^K : \Pr(\mathbf{x} \in \xi'_{\mathbf{b}}) = 0\}$. Note that Θ'_I is characterized by the distribution of observable regressors and conditional choice probabilities only, and can be used for finding a non-stochastic function $Q(\mathbf{b})$ that is minimized if and only if $\mathbf{b} \in \Theta'_I$. The function $Q(\mathbf{b})$ can be approximated by its sample analog and preliminary kernel estimates of $P^*_{1|\mathbf{x}}$, and will be used to define our extreme estimator below. In general this set of observationally equivalent coefficients Θ'_I will not be a singleton. But as additional restrictions are imposed on error distributions, the size of the identification region of $\boldsymbol{\beta}$ will be reduced. For instance, under a slightly stronger version of *BCQ*, the new identification region will be a subset of Θ'_I.

BCQ-2: The error ε *has continuous density conditional on all* $\mathbf{x}$ *and* $L(\mathbf{x}) \leq \inf Med(\varepsilon|\mathbf{x}) \leq \sup Med(\varepsilon|\mathbf{x}) \leq U(\mathbf{x})$ *a.e.* $F_{\mathbf{X}}$*, where* $L(.), U(.)$ *are known functions such that* $L(\mathbf{x}) \leq U(\mathbf{x})$ *a.e.* $F_{\mathbf{X}}$*,* $L \equiv \inf_{\mathbf{x} \in S(\mathbf{X})} L(\mathbf{x}) > -\infty$*, and* $U \equiv \sup_{\mathbf{x} \in S(\mathbf{X})} U(\mathbf{x}) < +\infty$*.*

Corollary 1 (Lemma 1) *In Model (2.1) under BCQ-2, the identification region of* $\boldsymbol{\beta}$ *is* $\Theta_I \equiv \{\mathbf{b} \in \mathbb{R}^K : \Pr(\mathbf{X} \in \xi'_{\mathbf{b}}) = 0\}$, *where* $\xi_{\mathbf{b}} \equiv \xi'_{\mathbf{b}} \cup \{\mathbf{x} \in S(\mathbf{X}) : (-\mathbf{x}^T\mathbf{b} < L(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} = \frac{1}{2}) \vee (-\mathbf{x}^T\mathbf{b} > U(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} = \frac{1}{2})\}$.

The classical restriction of median independence is a special case of *BCQ-2* with $L(\mathbf{x}) = \inf Med(\varepsilon|\mathbf{x}) = \sup Med(\varepsilon|\mathbf{x}) = U(\mathbf{x}) = 0 \; a.e. F_{\mathbf{X}}$. Under the classical median independence, the identification region is $\Theta^0_I \equiv \{\mathbf{b} \in \mathbb{R}^K : \Pr(\mathbf{x} \in \xi^0_{\mathbf{b}}) = 0\}$, where $\xi^0_{\mathbf{b}} \equiv \{\mathbf{x} \in S(\mathbf{X}) : (-\mathbf{x}'\mathbf{b} \leq 0 \wedge P^*_{1|\mathbf{x}} < \frac{1}{2}) \vee (-\mathbf{x}'\mathbf{b} \geq 0 \wedge P^*_{1|\mathbf{x}} > \frac{1}{2}) \vee (-\mathbf{x}'\mathbf{b} \neq 0 \wedge P^*_{1|\mathbf{x}} = \frac{1}{2})\} = \{x \in S(\mathbf{X}) : (-\mathbf{x}^T\mathbf{b} < 0 \wedge P^*_{1|\mathbf{x}} \leq \frac{1}{2}) \vee (-\mathbf{x}^T\mathbf{b} > 0 \wedge P^*_{1|\mathbf{x}} \geq \frac{1}{2}) \vee (-\mathbf{x}^T\mathbf{b} = 0 \wedge P^*_{1|\mathbf{x}} \neq \frac{1}{2})\}$. Note $\xi'_{\mathbf{b}} \subseteq \xi_{\mathbf{b}} \subseteq \xi^0_{\mathbf{b}}$ when $L(\mathbf{x}) \leq 0 \leq U(\mathbf{x})$ almost everywhere on $S(\mathbf{X})$. Hence $\Theta^0_I \subseteq \Theta_I \subseteq \Theta'_I$. The exact size of differences between these sets will be determined by the distribution of $\mathbf{X}$. These characterizations of the identification regions reveal little information about their analytical properties, as $P^*_{1|\mathbf{x}}$ depends on unknown parameters. However, it can be shown Θ'_I and Θ_I both satisfy the nice property of convexity.

Corollary 2 (Lemma 1) *Under BCQ, the identification region* Θ'_I *is convex. Under BCQ-2, the identification region* Θ_I *is convex.*

2.4. Point identification of β

Point identification is the special case where the identification region is reduced to a singleton. Despite the generality in the characterization of Θ_I, point identification of β is possible under fairly weak conditions on the parameter space, the support of regressors, and the form of bounding functions..

PAR (Parameter space) The true parameter $\boldsymbol{\beta}$ *is in the interior of* Θ, *where* Θ *is a convex, compact subset of* $\mathbb{R}^K$.[10]

[10] The compactness of Θ may be given by extraneous restrictions on the model implied by economic theories, such as signs and bounds on the sizes of coefficients.

SX-1 (Support of $\mathbf{X}$*) (a)* $\exists J \subset \{1, 2, ..., K\}$ *such that for all* $\mathbf{b} \in \Theta$, $b_j = 0\ \forall j \in J$ *and there exists no nonzero vector* $\boldsymbol{\lambda} \in \mathbb{R}^{K-\#\{J\}}$ *such that* $\Pr(\mathbf{X}'_{-J}\boldsymbol{\lambda} = 0) = 1$ *where* $\mathbf{X}_{-J} \equiv (X_j)_{j\in\{1,..,K\}\backslash J}$*; (b) For all* $\mathbf{b}, \tilde{\mathbf{b}} \in \Theta$ *and* $\mathbf{b}_{-J} \neq \tilde{\mathbf{b}}_{-J}$, $\Pr\{\mathbf{X}_{-J} \in T(\mathbf{b}_{-J}, \tilde{\mathbf{b}}_{-J})\} > 0$ *where* $T(\mathbf{b}_{-J}, \tilde{\mathbf{b}}_{-J}) \equiv \{\mathbf{x}_{-J} : (L, U) \cap R(\mathbf{x}_{-J}; \mathbf{b}_{-J}, \tilde{\mathbf{b}}_{-J}) \neq \varnothing \wedge \mathbf{x}'_{-J}(\mathbf{b}_{-J} - \tilde{\mathbf{b}}_{-J}) \neq 0\}$ *and* $R(\mathbf{X}_{-J}; \mathbf{b}_{-J}, \tilde{\mathbf{b}}_{-J})$ *is the random interval between* $\mathbf{X}'_{-J}\mathbf{b}_{-J}$ *and* $\mathbf{X}'_{-J}\tilde{\mathbf{b}}_{-J}$*; (c)* $\Pr(a_0 < L(\mathbf{X}) \wedge U(\mathbf{X}) < a_1 | \mathbf{X}_{-J} = \mathbf{x}_{-J}) > 0$ *for all open interval* $(a_0, a_1) \subset [L, U]$ *and almost everywhere* $\mathbf{x}_{-J}$*.*

Proposition 1 *Under BCQ-2, PAR and SX-1,* $\boldsymbol{\beta}$ *is identified relative to all other* $\mathbf{b} \in \Theta$*.*

The support conditions in *SX-1* are quite general. In particular, they allows for both discrete coordinates and bounded support of $\mathbf{X}$. Below I show how they can be satisfied by more primitive conditions on the support of regressors in some of the motivating models.

Model 2 *(Revisited)* Consider a simple 2-by-2 discrete game with incomplete information. Below I give primitive conditions sufficient for the point identification of $\boldsymbol{\beta}_1$.

(REG) (i) $\exists l \subset \{1, 2, ..., K\}$ *such that* $\beta_{1l} = 0$, $\beta_{2l} \neq 0$ *and* $\forall$ *nonzero vector* $\boldsymbol{\lambda} \in \mathbb{R}^{K-1}$, $\Pr(\mathbf{X}'_{-l}\boldsymbol{\lambda} \neq 0) > 0$*; (ii)* $\exists$ *an unknown constant* $C < \infty$ *such that* $P(|\mathbf{X}'_{-l}\mathbf{b}_{1,-l}| \leq C) = 1\ \forall \mathbf{b}_1 \in \Theta_1$ *where* Θ_1 *is the parameter space for* $\mathbf{b}_1$*; (iii) for all* $\bar{\mathbf{x}}_{-l} \in S(\mathbf{X}_{-l})$, X_l *is continuously distributed on the compact support* $S(X_l|\bar{\mathbf{x}}_{-l})$ *with the conditional density bounded below from zero; (iv) for all* $\mathbf{b}_2 \in \boldsymbol{\Theta}_2$, $\exists x_l^u, x_l^l \in S(X_l|\bar{\mathbf{x}}_{-l})$ *such that* $x_l^u b_{2l} + \bar{\mathbf{x}}_{-l}\mathbf{b}_{2,-l} = \varepsilon_2^u + 1$ *and* $x_l^l \beta_{2l} + \bar{\mathbf{x}}_{-l}\beta_{2,-l} = \varepsilon_2^l - 1$*;*[11] *(v) Let* $\mathbf{X}^c_{-l}$ *and* $\mathbf{X}^d_{-l}$ *denote respectively subvectors of continuous and discrete coordinates of* $\mathbf{X}_{-l}$*. For all* $\mathbf{S}$ *such that* $P(\mathbf{X}^c_{-l} \in \mathbf{S}) > 0$, $P(\mathbf{X}^d_{-l} = \mathbf{0} \wedge \mathbf{X}^c_{-l} \in \alpha\mathbf{S}) > 0\ \forall \alpha \in (-1, 1)$ *where* $\alpha\mathbf{S} \equiv \{\tilde{\mathbf{x}} : \tilde{\mathbf{x}} = \alpha\mathbf{x}\ for\ \ some\ \ \mathbf{x} \in \mathbf{S}\}$*.*

(ERR) For $i = 1, 2$*, (i) for all* $\mathbf{x} \in S(\mathbf{X})$*, the conditional disturbance distributions* $F_{\varepsilon_i|\mathbf{X}=\mathbf{x}}$ *are continuously differentiable for all* ε_i *in the interior of the compact support* $S_{\varepsilon_i} \equiv [\varepsilon^i_L, \varepsilon^i_U]$

[11] The proof below can be adjusted to allow the support S_{ϵ_j} to change with $\mathbf{x}$.

with the conditional median being zero, and is Lipschitz continuous on S_{ε_i} *with an unknown constant* $C_{i,\varepsilon} > 0$*; (ii) there exists an unknown constant* $C_{i,x} > 0$ *s.t.* $\sup_{t\in[\varepsilon_L^i,\varepsilon_U^i]} |F_{\varepsilon_i|\mathbf{x}_{-l},x_l'}(t) - F_{\varepsilon^i|\mathbf{x}_{-l},x_l}(t)| \leq C_{i,x}|x_l' - x_l|$ *for all* $\mathbf{x}_{-l} \in S(\mathbf{X}_{-l})$ *and* $x_l \in S(X_l|\mathbf{x}_{-l})$.

Corollary 1 (Proposition 1) *In Model 2, suppose* $\boldsymbol{\beta}_1 \neq 0$ *belongs to a compact parameter space* Θ_1*, and (REG), (ERR) are satisfied. Then* $\boldsymbol{\beta}_1$ *is identified relative to all other* $\mathbf{b}_1 \in \Theta_1$.

The support conditions are quite general, and allow the regressor to have bounded support. This is an important technical nicety as the compactness of regressor supports will come in hand in the proof of asymptotic properties of the estimator proposed in the sections below. ■

Model 1 *(Revisited)* Consider the binary choice model with interval data on one of the regressors. The augmented vector of regressors is $\mathbf{Z} \equiv [\mathbf{X}\ V_0\ V_1] \in \mathbb{R}^{K+2}$. Note by construction, $\mathbf{Z}_J = [V_0\ V_1]$, and $L(\mathbf{Z}) = V_0$, $U(\mathbf{Z}) = V_1$, and $\boldsymbol{\beta}_J = [0\ 0]$. Let V_0 and V_1 have unbounded support and the support of $\mathbf{X}$ not to be contained in a linear subspace of $\mathbb{R}^K$. Then conditions *SX1-(a)* and *(b)* are satisfied. And $\boldsymbol{\beta}$ is point identified if $\Pr(a_0 < L(\mathbf{X}) \wedge U(\mathbf{X}) < a_1 | \mathbf{X} = \mathbf{x}) > 0$ for all open interval $(a_0, a_1) \subset \mathbb{R}^1$ and all $\mathbf{x} \in S(\mathbf{X})$. This is exactly the conditions specified in Manski and Tamer (2002). ■

The identifying restrictions in *SX-1* is essentially an exclusion restriction in that it requires a regressor that affects $L(.)$ or $U(.)$ but does not enter the linear index. Below I give a different set of exclusion restrictions which requires regressors that do not affect $L(.)$ or $U(.)$ but enter the linear index.

SX-1' (Support of $\mathbf{X}$*) (a)* $\exists k \in \{1, 2, ..., K\}$ *such that* $\beta_k \neq 0$, $L(\mathbf{x}_{-k}) = L(\mathbf{x})$ *and* $U(\mathbf{x}_{-k}) = U(\mathbf{x})\ \forall \mathbf{x}$*, and for almost every value of* $\mathbf{x}_{-k} = (\mathbf{x}_1, ., \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, ., \mathbf{x}_K)$, $\Pr(X_k \in$

$(a_1, a_2)|\mathbf{x}_{-k}) > 0$ *for all open interval* $(a_0, a_1) \subset \mathbb{R}^1$; *; (b)* $\forall \mathbf{b}_{-k} \neq \boldsymbol{\beta}_{-k}$, $\Pr\{L(\mathbf{X}) = U(\mathbf{X}) \wedge \mathbf{X}'_{-k}(\mathbf{b}_{-k} - \boldsymbol{\beta}_{-k}) \neq 0\} > 0$.

Proposition 2 *Under BCQ-2, PAR and SX-1', $\boldsymbol{\beta}$ is identified relative to all other* $\mathbf{b} \in \Theta$.

The support conditions *SX-1'* allows for discrete regressors. However, it does not satisfy the property of bounded support for regressors.

2.5. A Two-step Extreme Estimator

I construct a two-step extreme estimator following standard steps. First I define a non-stochastic function $Q(\mathbf{b})$ which is minimized if and only if $\mathbf{b} \in \Theta_I$, where Θ_I is the identification region of $\boldsymbol{\beta}$ under BCQ-2. Then I construct sample analogs $\hat{Q}_n(\mathbf{b})$ of $Q(\mathbf{b})$ using the empirical distribution and a first-step kernel estimator. The two-step extreme estimator is then defined as the minimizer of the stochastic objective function $\hat{Q}_n$.

SX-2 $\Pr(-\mathbf{X}'\mathbf{b} = U(\mathbf{X}) \vee -\mathbf{X}'\mathbf{b} = L(\mathbf{X})) = \mathbf{0}$ *for all* $\mathbf{b} \in \Theta$.

Lemma 2 *(Identification) Define the nonstochastic function*

$$Q(\mathbf{b}) \equiv E[1(P^*_{1|\mathbf{X}} \geq 1/2)(-U(\mathbf{X}) - \mathbf{X}'\mathbf{b})^2_+ + 1(P^*_{1|\mathbf{X}} \leq 1/2)(-L(\mathbf{X}) - \mathbf{X}'\mathbf{b})^2_-]$$

where $a_+ \equiv \max(0, a)$ *and* $a_- \equiv \max(0, -a)$. *Under BCQ-2 and SX-2,* $Q(\mathbf{b}) \geq 0$ $\forall \mathbf{b} \in \Theta_I$ *and* $Q(\mathbf{b}) = 0$ *if and only if* $\mathbf{b} \in \Theta_I$.

For simplicity in exposition, below I will construct the two-step extreme estimator for the case where all regressors are continuous. The extension to the case where some regressors are discrete does not cause any conceptual or technical difficulty, and will be omitted. The first step estimates the choice probabilities using kernel regressions. Define the kernel estimates for density $f_0(\mathbf{x}_i)$ and $h_0(\mathbf{x}_i) \equiv E(Y_i|\mathbf{X}_i = \mathbf{x}_i)f_0(\mathbf{x}_i)$ as

$$\hat{f}(\mathbf{x}_i) \equiv (n\sigma_n^K)^{-1} \textstyle\sum_{j=1, j\neq i}^n K[(\mathbf{x}_j - \mathbf{x}_i)/\sigma_n], \ \hat{h}(\mathbf{x}_i) \equiv (n\sigma_n^K)^{-1} \textstyle\sum_{j=1, j\neq i}^n y_j K[(\mathbf{x}_j - \mathbf{x}_i)/\sigma_n]$$

where $K(.)$ is a kernel function and σ_n is the chosen bandwidth. The nonparametric estimates for $p(\mathbf{x}_i)$ is $\hat{p}(\mathbf{x}_i) \equiv \hat{h}(\mathbf{x}_i)/\hat{f}(\mathbf{x}_i)$. Now construct the sample analog of $Q(\mathbf{b})$:

$$\hat{Q}_n(\mathbf{b}) = \frac{1}{n} \textstyle\sum_{i=1}^n 1\{\hat{p}(\mathbf{x}_i) \geq \frac{1}{2}\}[-\mathbf{x}_i'\mathbf{b} - U(\mathbf{x}_i)]_+^2 + 1\{\hat{p}(\mathbf{x}_i) \leq \frac{1}{2}\}[-\mathbf{x}_i'\mathbf{b} - L(\mathbf{x}_i)]_-^2$$

The two-step extreme estimator is defined as:

$$\hat{\Theta}_n = \arg\min_{\mathbf{b}\in\Theta} \hat{Q}_n(\mathbf{b})$$

2.5.1. Consistency under set-identification

In general, conditions for point identifying $\boldsymbol{\beta}$ may not be satisfied. Therefore the concept of a consistent estimator when the parameter is point identified needs to be extended to the case of set-identification. The Hausdorff distance between two compact sets A and B in $\mathbb{R}^K$ is defined as

$$\rho(A, B) \equiv \max\{\sup_{a\in A} \inf_{b\in B} ||a - b||, \sup_{b\in B} \inf_{a\in A} ||a - b||\}$$

where $||.||$ is the Euclidean norm. The metric is asymmetric in the sense that $\rho(A, B) \neq \rho(B, A)$. *Proposition 3* below proves the two-step extreme estimator is a consistent estimator of the identification region Θ_I in the Hausdorff metric. For technical reasons, I replace the indicator functions in the definition of $Q(\mathbf{b})$ with smooth functions $\Lambda(p(\mathbf{x}_i) - \frac{1}{2})$ and $\Lambda(p(\mathbf{x}_i) - \frac{1}{2})$ respectively. Regularity conditions for set consistency are collected below.

RD-1 (Regressors and disturbance) (i) the $(K+1)$-dimensional random vector $(\mathbf{X}_i', \varepsilon_i)$ is independently and identically distributed; (ii) The support of $\mathbf{X}$ (denoted $S(\mathbf{X})$) is bounded, *and its continuous coordinates have bounded joint density* $f_0(\mathbf{x})$*; (iii) the density function* f_0 *is* $k+1$ *times continuously differentiable on the interior of the support* $S(\mathbf{X})$*.*

K (Kernel estimator) (i) $K(.)$ *is continuous and zero outside a bounded set; (ii)* $\int K(u)du = 1$ *and for all* $l_1 + .. + l_k < k+1$, $\int u_1^{l_1}...u_k^{l_k} K(u)du = 0$*; (iii)* $n\sigma_n^{2k}/(\log n)^2 \rightarrow \infty$ *and* $n\sigma^{2(k+1)} \rightarrow 0$.

TF (Trimming functions) (i) $\Lambda : \mathbb{R} \rightarrow [0,1]$ *is bounded with continuous and bounded first and second derivatives; (ii)* $\Lambda(t) \in (0,1]$ *for* $t > 0$*, and* $\Lambda(t) = 0$ *otherwise.*

That the trimming function carries positive weights if and only if the argument is positive is essential for identification. The proof of *Lemma 2* still applies with the indicator function replaced by Λ. Conditions on regressors and kernels deliver the uniform convergence in probability of $\hat{p}$ to p. The smoothness of Λ and the boundedness of regressor supports are convenient technicalities for proving the convergence in probability of the stochastic criterion function to $\hat{Q}_n$ to Q, and the root-n asymptotic normality.

Theorem 1 *Suppose BCQ-2, SX-2, PAR, TF, RD-1 and K are satisfied. Then* $\Pr(\rho(\hat{\Theta}_n, \Theta_I) > \varepsilon) \rightarrow 0$ *as* $n \rightarrow \infty$ *for all* $\varepsilon > 0$.

This result of set-consistency incorporates the special case where Θ_I is a singleton ($\boldsymbol{\beta}_0$ is point identified). In general, $\hat{Q}_n$ may not be uniquely minimized even when $\boldsymbol{\beta}_0$ is known to be exactly identified. In this case, $\hat{\boldsymbol{\beta}}_n$ may be chosen randomly from the set of minimizers of $\hat{Q}_n$, and it converges in probability to $\boldsymbol{\beta}_0$. Chernozhukov, Tamer and Hong (2008) studied the inference of extreme estimators in a very general class of partially identified models which includes our model here. Their approach of inference is based on approximating the distributions of criterion functions (maximized over the identification region) through a subsampling procedure.

2.5.2. Asymptotic properties under point-identification

Obviously, consistency under point identification is a special case of Theorem 1. For the rest of this section, I discuss the root-n asymptotic normality of the two-step extreme estimator when $\boldsymbol{\beta}$ is point identified in Θ. For various technical reasons, the estimator is

$$\hat{Q}_n(\mathbf{b}) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \hat{\Lambda}_i^u[-\mathbf{x}_i'\mathbf{b} - U(\mathbf{x}_i)]_+^2 + \hat{\Lambda}_i^l[-\mathbf{x}_i'\mathbf{b} - L(\mathbf{x}_i)]_-^2$$

where $\hat{\Lambda}_i^u \equiv \Lambda(\hat{p}(\mathbf{x}_i) - \frac{1}{2})$, $\hat{\Lambda}_i^l \equiv \Lambda(\frac{1}{2} - \hat{p}(\mathbf{x}_i))$, and Λ is a smooth function that satisfies the regularity conditions above.

RD-2 (i) $f_0(\mathbf{x}) \geq c\ \forall \mathbf{x} \in S(\mathbf{X})$ *for some small constant* $c > 0$*; (ii)* $\forall\ \mathbf{b} \in \Theta$*, the Lebesgue measure of* $S_j(\mathbf{b}, \varepsilon) \equiv \{\mathbf{x} \in S(\mathbf{X}) : sgn(L(\mathbf{x}) - \mathbf{x}'\mathbf{b}) \neq sgn(L(\mathbf{x}) - \mathbf{x}'\mathbf{b} - x_j\varepsilon)\} \to 0$ *as* $\varepsilon \to 0$*; (iii)*$\exists B(.) : S(\mathbf{X}) \to \mathbb{R}^1$ *such that* $|\max\{(v_i^l)_-, (v_i^u)_+\}| \leq B(\mathbf{x}_i)$ *with* $E[B(\mathbf{X})] < \infty$*, and* $E[B(\mathbf{X})||\mathbf{X}||] < \infty$*; (iv)* $L(.) < \infty$ *and* $U(.) < \infty$ *on* $S(\mathbf{X})$*; (v)*$E[||[\Lambda^l(p_i)(V_i^l)_- + \Lambda^u(p_i)(V_i^u)_+]\mathbf{X}_i||^2] < \infty$*; (vi)* $E[||\mathbf{X}_i||^4(V_i^l)_-^4] < \infty$, $E[||\mathbf{X}_i||^4 \cdot (V_i^u)_+^4] < \infty$ *and* $E[||\mathbf{X}_i||^4] \leq \infty$.

Theorem 2 *(Asymptotic Normality) Suppose BCQ-2, PAR, SX-1,2, RD-1,2, TF and K are satisfied, and matrices* J *and* Σ *defined below are both non-singular. Then* $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d}$

$N(0, J^{-1}\Sigma J^{-1})$, *where*

$$\begin{aligned}
\Sigma &\equiv Var[\xi(\mathbf{X}; h_0, f_0) + \Delta(\mathbf{X}, Y; h_0, f_0)] \\
\xi(\mathbf{x}, h_0, f_0) &\equiv 2\left[\Lambda\left(\frac{1}{2} - \frac{h_0(\mathbf{x})}{f_0(\mathbf{x})}\right)(V_i^l)_- + \Lambda\left(\frac{h_0(\mathbf{x})}{f_0(\mathbf{x})} - \frac{1}{2}\right)(V_i^u)_+\right]\mathbf{x} \\
\Delta(\mathbf{x}, y; h_0, f_0) &\equiv \left[a_1(\mathbf{x}; h_0, f_0)y + \frac{h_0(\mathbf{x})}{f_0(\mathbf{x})}a_1(\mathbf{x}; h_0, f_0)\right]\mathbf{x} \\
a_1(\mathbf{x}; h, f) &\equiv 2\left\{\max(0, v_i^u)\Lambda'\left(\frac{h(\mathbf{x})}{f(\mathbf{x})} - \frac{1}{2}\right) - \max(0, -v_i^l)\Lambda'\left(\frac{1}{2} - \frac{h(\mathbf{x})}{f(\mathbf{x})}\right)\right\} \\
J &\equiv 2n^{-1}E\{[\Lambda^l(p_i)1(V_i^l < 0) + \Lambda^u(p_i)1(V_i^u > 0)]\mathbf{X}_i\mathbf{X}_i'\}
\end{aligned}$$

The asymptotic normality proof follows similar steps in Buchinsky and Hahn (1998). First, I let the criterion function in the second step be approximated by a version of 2nd-order Taylor expansion of the limiting function around the true parameter but with the 1st-order term (the "score") replaced by its sample analog that depends on first-step preliminary kernel estimates. Then I showed the approximation error is small enough to be omitted in discussing asymptotic distributions. Next I follow the standard steps in Theorem 8.1 in Newey and McFadden (1994) to show the sample score term converges in distribution to a normal distribution. These arguments combine to prove the asymptotic normality of our two-step extreme estimator. Consistent estimators of the asymptotic covariance matrices can be constructed using the sample analog principle.

Recall the general model described in Section 2 encompasses classical binary regression with median independence as a special submodel, for which Chamberlain (1986) showed there exists no root-n consistent semiparametric estimator when the parameter is point identified. Therefore the result that root-n asymptotic normality is attained in *Theorem 2* appears to be counter-intuitive. However, note the assumptions *SX-1* and *RD-1* in Theorem 2 require that the model is point identified while the support of regressors is bounded.[12] Manski (1986) showed this can never be the case for binary regression model under median independence. As a result, *Theorem 2* does not apply to the class of model studied by Chamberlain (1986).

2.6. Monte Carlo Experiments

In this section, I study the finite sample performance of the two-step extreme estimator in the more general context where $\boldsymbol{\beta}_0$ is set-identified. I experiment with two designs of binary response models with interval data on one of the regressors. Specifically, $Y = 1\{\beta_0 + \beta_1 X + V + \varepsilon \geq 0\}$. In the first design, $V \tilde{} N(0,2)$ and $X \tilde{} N(0,4)$, and in the second design $V \tilde{} Uniform(-2,3)$ and $X \tilde{} Uniform(0,5)$. In both designs, $\varepsilon \tilde{} N(0,1)$, (V, X, ε) are statistically independent, and V is not observed by the researcher. Instead, only $V_0 = int(V)$ and $V_1 = int(V) + 1$ are observed. These are exactly the same designs as considered in Manski and Tamer (2002). The sufficient conditions for point identification in Section 4 are not satisfied and there is no reason to believe the coefficients β_0 and β_1 are point identified.

I do not derive the closed form of the identification region. Instead I simulate a large data set with 10^5 observations, and treat it as the population for our Monte Carlo studies. I apply the two-step extreme estimators to this data set and use the estimates to approximate the real identification region. (See *Figure 1* and *Figure 2*.) For both designs, I reported the performance of the estimator in samples with $N = 500$, 1000 and 3000 respectively. For each sample size N considered, I simulate 100 different samples by making random draws from the population with replacement. I calculate the two-step extreme estimates for each of the 100 samples. I use Naradaya-Watson kernel regressions to estimate conditional choice

[12] Such a submodel exists within our general model. I have proven this through an example of simple discrete games with incomplete information in Section 3.

probabilities in the first step. The bandwidths are chosen through cross-validations and Gaussian kernels are used. The maximization procedure in the second step is done by a two dimensional grid-search.

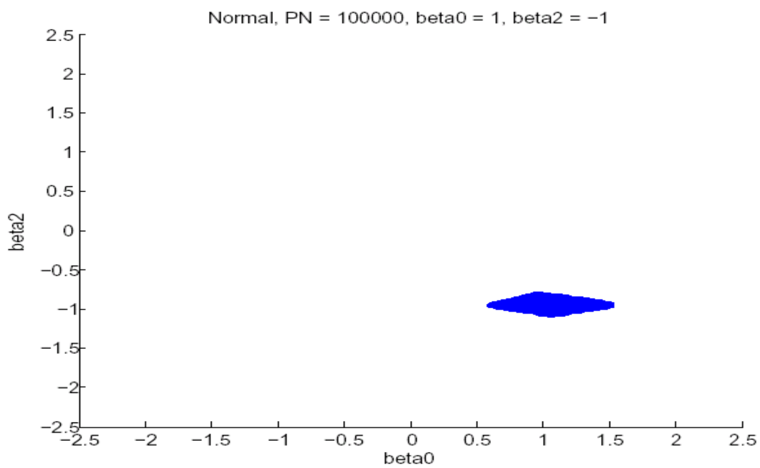


Figure 1: Identification Region (Normal)

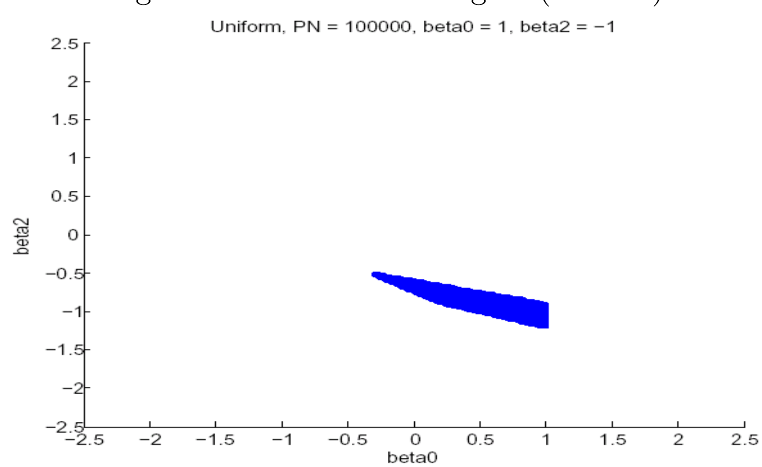


Figure 2: Identification Region (Uniform)

Figure 1 and Figure 2 confirm the earlier proposition about the convexity of the identification region in Section 3. The size of the identification is also small relative to the variance of regressors in the designs. There has been lots of recent contributions in the literature on the inference of non-singleton identification region. For instance, Chernozhukov, Hong and Tamer (2003) proposed a general criterion function approach of set inference for extreme estimators that can be applied to our model here. In this section, I do not take this approach to report any confidence regions. Rather, for each of the 100 estimated sets, I record the percentage of the identification region it covers (denoted $P1$), as well as the proportion of the estimated set contained in the identification region (denoted $P2$). I use these two proportions as measures of discrepancies between the two-step estimates and the real identification region. *Table 1* below reports different percentiles of these two measures among the 100 simulations.

Table 1 (a): Normal Design

	$P1$			$P2$		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0	0	0.054	0	0	0.344
25%	0.017	0.097	0.345	0.228	0.296	0.487
50%	0.370	0.444	0.571	0.409	0.520	0.594
75%	0.653	0.724	0.787	0.597	0.686	0.701
90%	0.841	0.860	0.934	0.808	0.839	0.853

In the normal design, *Table 1(a)* suggests the discrepancies between the worst estimates and the identification region is quite noticeable for small samples. In particular, the first quartile of $P1$ (the percentage of identification region covered by an estimated set) is smaller than 10% for $n = 500$ and $n = 1000$. And the medians for $P1$ are both lower than 50%. The performance is remarkable enhanced when the sample size is increased. In particular, the first quartile for $P1$ with $n = 3000$ reports a much higher proportion. In comparison, the estimators have higher first quartile for $P2$ (the percentage of an estimated set covered by the identification region). The difference between $P1$ and $P2$ for higher quartiles are less pronounced.

Table 1 (b): Uniform Design

	$P1$			$P2$		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0.581	0.627	0.814	0.509	0.635	0.786
25%	0.664	0.755	0.859	0.600	0.719	0.848
50%	0.782	0.854	0.911	0.685	0.807	0.908
75%	0.925	0.949	0.968	0.843	0.895	0.954
90%	0.989	0.984	0.994	0.923	0.965	0.982

Table 1(b) suggests the performance of the estimator under the uniform design is much better than under the normal design. This is best illustrated by lower percentiles for smaller sample sizes. The median for $P1$ and $P2$ are remarkably high for all sample sizes.

Table 1 (c): Min{P1,P2}

	$Normal$			$Uniform$		
percentile	$n = 500$	$n = 1000$	$n = 3000$	$n = 500$	$n = 1000$	$n = 3000$
10%	0	0	0.054	0.503	0.583	0.758
25%	0.013	0.086	0.341	0.582	0.673	0.817
50%	0.294	0.367	0.477	0.659	0.748	0.858
75%	0.453	0.541	0.618	0.756	0.818	0.902
90%	0.546	0.633	0.688	0.854	0.863	0.940

A more comprehensive measure the discrepancies between the estimates and the identification region is $\min\{P1, P2\}$ reported in *Table 1(c)*. By this criterion, the estimator also performs obviously better under the uniform design than under the normal design.

2.7. Conclusion

In this paper I have studied the identification and estimation of a class of binary response models where the conditional median of the error term is bounded between known functions of the regressors. I focus on the case where the payoff functions satisfy a linear index specification. Though the index coefficients may not be exactly identified, a two-step extreme estimator can estimate the identification region consistently regardless of point identification. Furthermore, when point identification is achieved with bounded support of regressors, the estimator is converges in distribution to a normal random variable at a rate of $\sqrt{n}$. Monte Carlo evidence suggests the estimator has good finite sample behavior.

Directions for future research includes the search for point identification conditions when the payoff functions have more general forms than linear indices. Another interesting issue is the estimation of the model when the bounding functions L and U are only known up to finite dimensional parameters. In particular, it will be interesting to look at what can be identified when the payoff functions, as well as L and U, are known only to satisfy certain shape restrictions.

CHAPTER 3

Identification of Dynamic Binary Choice Processes

3.1. Introduction

In a typical dynamic binary choice process, the decision-maker's payoffs in each period depend on contemporary states and his choice of action, which in turn impact the distribution of states in the future. The agent is forward-looking and makes a sequence of choices in each period to maximize the sum of contemporary and future expected returns. The structural parameters of the model are single-period returns, and the transitions between current and future states. Such dynamic binary choice models have found wide applications in the literature of empirical industrial organizations and labor economics. Recent applications include replacement of bus engines in Rust (1994), analysis of unemployment insurance in Ferrall (1997), the inventories of retailing firms in Aguirregabiria (1999), evaluation of welfare policies in Keane and Wolpin (2000), and consumer stockpiling in Hendel and Nevo (2005). Aguirregabiria (2007) gives an updated survey of estimation and inference of dynamic discrete choice processes.

In this paper we study the identification of structural parameters in a class of dynamic binary choice models where transitions to future states are independent from disturbances (i.e. states unobservable to econometricians) conditional on current actions and observable states. We address the question of what can be learned about the decision-maker's single-period payoffs under various restrictions on disturbance distributions. This question of identification is important for several reasons. First, the exact values of structural parameters per se are interesting to researchers; Second, estimates for parameters are often needed in empirical research for policy analyses beyond the simple prediction of choice probabilities on the support of states observed. For example, researchers may wish to study policy implications of counterfactual changes in the structural parameters, or to extrapolate choice probabilities conditional on states out of the support observed. For these questions, structural parameters that are not identified relative to the truth may have different implications, and it is important to find out what features of the parameters can be uniquely recovered from observables in the model.

There has been some recent development in the literature on identification of dynamic binary choice processes under the conditional independence restriction. Rust (1994) argued through an example that single-period payoffs are not identified even in the absence of disturbances. Magnac and Thesmar (2002) noted with this knowledge the differences between expected payoffs from two sequences of choices are identified: one is to choose 1 today, 0 tomorrow and behave optimally afterwards, and the other is to choose 0 for both today and tomorrow and behave optimally afterwards. Aguirregabiria (2005) studied counterfactual choice probabilities instead of focusing on recovering structural parameters. He showed counterfactual choice probabilities under policy changes of single-period payoffs can be fully nonparametrically recovered from choice probabilities observed, provided the form of policy change is known to the researcher. Berry and Tamer (2006) considered an optimal stopping problem where the decision to stop brings an end to the choice process, and showed the single-period return from not stopping is uniquely recovered when the disturbance distribution is known.

This paper contributes to this growing literature in several ways: First, we give a full characterization of the set of structural parameters (single-period payoffs and disturbance distributions) that can generate the same choice probabilities as observed in a given dynamic binary choice process. This introduces a convenient framework for studying the identification of single-period payoffs under parametric and stochastic restrictions on disturbance distributions. Second, we show that with knowledge of disturbance distributions, the differences between payoffs from two trivial policies of choosing one of the actions forever can be uniquely recovered from choice probabilities observed. Third, we analyzed the identification of single-period payoffs when the distribution of unobservable states is statistically independent from, or symmetric conditional on observable states. For the case of finite space of observable states, the set of observationally equivalent structural parameters is characterized by a system of *linear equations*. Then by definition, the identification region of single-period payoffs under these stochastic restrictions is the set of vector values that guarantee the existence of distributions which satisfy the linear equations subject to systems of *linear inequality* constraints implied by these restrictions. Hence the identification region of single-period payoffs under these restrictions is characterized by checking feasibility of the augmented system of linear equations in the nuisance (distributional) parameters with inequality constraints. Though proposed in the context where no form restrictions is imposed on payoff functions, this approach of identification using linear programming can be readily extended to cases where single-period payoffs are known to satisfy any form of restrictions.

The rest of the paper proceeds as follows. Section 2 specifies the model of dynamic binary choice models and characterize the joint identification region of the structural parameters in the absence of any parametric or stochastic restrictions. Section 3 discusses the benchmark situation where the distribution of disturbances is completely known. Section 4 examines the identifying power of various parametric and stochastic restrictions on disturbance distributions. Section 5 concludes.

3.2. The Model

We consider a single-agent, dynamic binary choice process in an infinite horizon. The time is discrete and indexed by t. In each period t, the decision maker observes a state vector $\mathbf{s}_t = (\mathbf{x}_t, \boldsymbol{\varepsilon}_t)$ from the support $S(\mathbf{S}) = S(\mathbf{X}) \otimes S(\boldsymbol{\epsilon}) \subseteq \mathbb{R}^{D+2}$, and chooses an action j_t from a pair of possible actions $\mathbf{J} = \{0, 1\}$.[1] The state space $S(\mathbf{S}) \subseteq \mathbb{R}^{D+2}$ is fixed over time. For each period, researchers can observe $\mathbf{x}_t$, but not $\boldsymbol{\varepsilon}_t$. The latter is only observed by the decision-maker. The single-period return for the decision-maker is $U(\mathbf{s}_t, j_t) : S(\mathbf{S}) \otimes \mathbf{J} \to \mathbb{R}^1$ for all t. Conditional on the current state $\mathbf{s}$ and action j, the distribution of states in the next period $\mathbf{s}'$ is given by the transition function $H_j(\mathbf{s}'|\mathbf{s}) : S^2(\mathbf{S}) \to [0, 1]$. The agent has the same discount factor $\beta \in (0, 1)$ forever. Both the single-period return and the transition probability are fixed over time.[2] The decision-maker chooses a deterministic, Markovian decision rule $j(\mathbf{s})$ that maximizes the sum of expected present and future payoffs: $E[\sum_{s=0}^{\infty} \beta^j U(\mathbf{s}_{t+s}, j_{t+s})|\mathbf{s}_t, j_t]$.[3] The following restrictions are maintained throughout the paper unless noted otherwise.

AS (Additive Separability) $U(\mathbf{s}, j) = u_j(\mathbf{x}) + \varepsilon_j$, $\boldsymbol{\varepsilon} \equiv [\varepsilon_0, \varepsilon_1] \in \mathbb{R}^2$, $E(\varepsilon_j|\mathbf{x}) = 0\ \forall(\mathbf{x}, j)$;

CI (Conditional Independence) $H_j(\mathbf{s}'|\mathbf{s}) = F_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')G_j(\mathbf{x}'|\mathbf{x})\ \forall \mathbf{s}, \mathbf{s}' \in \mathbf{S},\ j \in \{0, 1\}$, where $F_{\boldsymbol{\varepsilon}|\mathbf{X}}(.|\mathbf{x})$ and $G_j(.|\mathbf{x})$ are distributions defined on $S(\boldsymbol{\epsilon})$ and $S(\mathbf{X})$ respectively for all $\mathbf{x} \in S(\mathbf{X})$ and $j \in \{0, 1\}$.

[1]Throughout the paper I use bold letters to denote vectors.

[2]For notational ease, I will drop time subscripts for the rest of the paper due to time-invariance of period return, transitional probabilities, and the state and action spaces.

[3]In general, the optimal policies should be a function of past histories $\boldsymbol{\pi}_t = \{\mathbf{s}_j\}_{j=0}^{t}$. However Strauch (1966) showed for any history-dependent poilicy and starting state, there always exists a deterministic, Markovian policy (a policy that depends on the current state only) with the same expeced total discounted payoff. The implication is that for analysis of optimal policies, it suffices to focus on Markovian stationary policies. Throughout the paper we focus on the case where the agent only considers deterministic Markovian policies.

The transitions $\mathbf{G} \equiv [G_1 \; G_0]$ are identified from data of observed states and actions $\{j_t, \mathbf{x}_t\}_{t=0}^{+\infty}$ directly. Throughout the paper, we maintain that the constant discount factor β is known to econometricians, while structural parameters $\mathbf{u} \equiv [u_0(.) \; u_1(.)]$ and $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ are to be identified. *CI* requires that persistence between current and future states is captured by the persistence between observable states $\mathbf{x}'$ and $\mathbf{x}$, and therefore actions affect future states only through $\{G_j\}_{j=0,1}$. Given our focus on Markovian policies, an important implication of *CI* is that choice probabilities conditional on current states is independent from past states. That is, $\Pr(j_t = 1|\mathbf{x}_t) = \Pr(j_t = 1|\mathbf{x}_t, \mathbf{x}_{t-1})$ for all $\mathbf{x}_t, \mathbf{x}_{t-1}$. This is a testable implication using observable distributions. Therefore, the choice probability function $p(\mathbf{x}) \equiv \Pr(j_t = 1|\mathbf{x}_t)$ will be a sufficient statistic for the purpose of identifying $\mathbf{u}$ and $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$. Lemma 1 below shows under *AS*, *CI* and some regularity conditions, the dynamic binary choice process has a static representation.

REG (Regularity Conditions) (i) For $j \in \{0,1\}$, $u_j \in B(S(\mathbf{X}))$, *where* $B(S(\mathbf{X}))$ *is the set of bounded, continuous, real-valued functions on* $S(\mathbf{X})$*; (ii) For* $j \in \{0,1\}$, G_j *satisfies the Feller Property;*[4] *(iii) For all* $\mathbf{x} \in S(\mathbf{X})$, $j \in \{0,1\}$, $E[\max_{k\in\{0,1\}}\{\varepsilon_{t+1,k}\}|\mathbf{x}_t, j] < \infty$.

Lemma 1 Under *AS, CI* and *REG (i)-(iii)*, the value function of the dynamic binary decision process has a static representation:

$$j(\mathbf{s}) = \arg\max_{j\in\{0,1\}} \delta_j(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) + \varepsilon_j$$

where $\boldsymbol{\delta}(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) \equiv [\delta_0(\mathbf{x}) \; \delta_1(\mathbf{x})]'$ is the fixed point of the following operator $T \circ \boldsymbol{\delta}(\mathbf{x}) \equiv [T_1(\mathbf{x}; \boldsymbol{\delta}) \; T_0(\mathbf{x}; \boldsymbol{\delta})]$, where

$$T_j(\mathbf{x}) \equiv u_j(\mathbf{x}) + \beta \int \max_{k\in\{0,1\}}\{\delta_k(\mathbf{x}') + \varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) \tag{3.1}$$

[4] $G_j(\mathbf{x}'|\mathbf{x})$ satisfies the *Feller Property* if for each bounded, continuous function $f : S(\mathbf{X}) \rightarrow \mathbb{R}^1$, $\int f(\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x})$ is also bounded and continuous in $\mathbf{x}$.

As a result of this lemma, the conditional choice probability has a static representation: $p(\mathbf{x};\mathbf{u},F_{\boldsymbol{\varepsilon}|\mathbf{X}})=F_{\Delta\varepsilon|\mathbf{X}}[\Delta\delta(\mathbf{x})|\mathbf{x}]$, where $\Delta\varepsilon \equiv \varepsilon_0-\varepsilon_1$, $\Delta\delta(\mathbf{x}) \equiv \delta_1(\mathbf{x})-\delta_0(\mathbf{x})$, and the economic interpretation of $\delta_j(\mathbf{x}_t)$ is the expected return from choosing j in the current period conditional on observable states $\mathbf{x}_t$.[5] Throughout this paper, we focus on the question whether the single-period payoff $\mathbf{u}$ can be uniquely recovered from observable distributions of $\{j_t,\mathbf{x}_t\}_{t=0}^{\infty}$, and treat $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ as a nuisance parameter. Several technical notes are necessary before giving a formal definition of identification. First, we adopt the conventional sup norm on the space of $\mathbb{R}^2$-valued functions $||\mathbf{u}||_\infty \equiv \sup_{j\in\{0,1\},\mathbf{x}\in\mathbf{X}} |u_j(\mathbf{x})|$. Second, to fix ideas, we focus on cases with *finite* spaces of observable states, and leave the generalization to infinite spaces for future work. Specifically, we maintain the following support condition for the rest of the paper.

REG-(iv) (Discrete support of observable states) The space of observable states is time-invariant and $S(\mathbf{X})=\{\mathbf{x}_1,\mathbf{x}_2,..,\mathbf{x}_K\}$, with $\mathbf{x}_k\in\mathbb{R}^D$ for all k.

Definition 1 Two sets of structural parameters $\boldsymbol{\theta}\equiv(\mathbf{u},F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ and $\boldsymbol{\theta}'\equiv(\mathbf{u}',F'_{\boldsymbol{\varepsilon}|\mathbf{X}})$ are *observationally equivalent* if $p(\mathbf{x};\boldsymbol{\theta})=p(\mathbf{x};\boldsymbol{\theta}')$ for all $\mathbf{x}\in S(\mathbf{X})$. Let U and $\mathcal{F}$ denote respectively sets of single-period returns and conditional error distributions. We say $\mathbf{u}$ is *identified relative to* $\mathbf{u}'$ *under* $\mathcal{F}$ if $\forall$ $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$, $F'_{\boldsymbol{\varepsilon}|\mathbf{X}}\in\mathcal{F}$, $(\mathbf{u},F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ and $(\mathbf{u}',F'_{\boldsymbol{\varepsilon}|\mathbf{X}})$ are not observationally equivalent; and $\mathbf{u}$ is *identified within* U *under* $\mathcal{F}$ if $\mathbf{u}$ is identified relative to all $\mathbf{u}'\neq\mathbf{u}$ in U under $\mathcal{F}$. Let $p^*(\mathbf{x})$ be the choice probabilities observed. The *joint identification region* is the set of all $\boldsymbol{\theta}$ such that $p(\mathbf{x};\boldsymbol{\theta})=p^*(\mathbf{x})$, and the *identification region under* $\mathcal{F}$ is the set of all $\mathbf{u}$ such that $\exists F_{\boldsymbol{\varepsilon}|\mathbf{X}}\in\mathcal{F}$ with $p(\mathbf{x};\mathbf{u},F_{\boldsymbol{\varepsilon}|\mathbf{X}})=p^*(\mathbf{x})$.

[5] The conditional independence restriction can be weakened to *A2'*: $H_j(.|\mathbf{s})=H_j(.|\mathbf{x}),\forall j,\mathbf{s}$ and the representation result is still valid.

As a starting point for discussing identifications, Proposition 1 below characterizes the joint identification region of $(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ without further identifying restrictions. Let $F^{-1}_{\Delta\varepsilon|\mathbf{X}}(t|\mathbf{x})$ denote the inverse of $F_{\Delta\varepsilon|\mathbf{X}}(.|\mathbf{x})$ at $t \in [0,1]$.

Proposition 1 Suppose *AS*, *CI* and *REG (i)-(iv)* are satisfied. For any observed choice probability $p(\mathbf{x})$, the joint identification region is

$$\Theta_I \equiv \{(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) : \Delta\omega(\mathbf{x};\mathbf{u}) = F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x}) - \Delta\xi(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}}, p) \text{ for all } \mathbf{x} \in S(\mathbf{X})\} \tag{3.2}$$

where $\Delta\omega(\mathbf{x};\mathbf{u}) \equiv \omega_1(\mathbf{x}) - \omega_0(\mathbf{x})$, $\Delta\xi(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}}, p) \equiv \xi_1(\mathbf{x}) - \xi_0(\mathbf{x})$; $\omega_j(\mathbf{x})$ and $\xi_d(\mathbf{x})$ are unique fixed points of following operators:

$$T_\omega \circ (\omega_j(\mathbf{x})) \equiv u_j(\mathbf{x}) + \beta \int \omega_j(\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) \tag{3.3}$$

$$T_\xi \circ (\xi_j(\mathbf{x})) \equiv \beta \int \kappa_j(\mathbf{x}'; p, F_{\Delta\varepsilon|\mathbf{X}}) + \xi_j(\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) \tag{3.4}$$

with κ_d defined as

$$\kappa_0(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}}) \equiv \int_{-\infty}^{q(\mathbf{x})} [q(\mathbf{x}) - s] dF_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x})$$

$$\kappa_1(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}}) \equiv \int_{q(\mathbf{x})}^{+\infty} [s - q(\mathbf{x})] dF_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x})$$

where $q(\mathbf{x}) \equiv F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$.

Proposition 1 gives a formal characterization of the joint identification region of $(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ in the absence of further restrictions. Note the assumption of a finite state space is not essential for the proposition. It shows how changes in different specifications of $\mathbf{u}(\mathbf{x})$ can be offset by varying $F_{\Delta\varepsilon|\mathbf{X}}$ so as to generate the same choice probabilities $p(\mathbf{x})$. In the special case of static binary choice models (where the discount factor is 0 forever), the joint identification region takes the familiar form: $\{(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) : F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x}) = \Delta u(\mathbf{x}) \ \forall \ \mathbf{x} \in S(\mathbf{X})\}$.[6] Furthermore, note $\omega_j(\mathbf{x})$ is the expected payoff from the trivial policy of choosing j forever conditional on current states.[7] Hence Proposition 1 suggests the conditional expectation of differences in payoffs from two trivial policies can be completely recovered from observables with the knowledge of $F_{\Delta\varepsilon|\mathbf{X}}$.

3.3. Identification with Known Transitions

This sections focuses on the identification of $\mathbf{u}$ when $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is known in the dynamic binary choice processes. Since the transition between $\mathbf{x}$ is observed, this implies the transition $H_j(\mathbf{s}'|\mathbf{s}) = F_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')G_j(\mathbf{x}'|\mathbf{x})$ is also known under the conditional independence restriction. Formally, we maintain the following assumption throughout this section:

KD (Known distribution) The true conditional error distribution $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is known to the econometrician (i.e. the set of possible error distributions $\mathcal{F}$ is a known singleton).

Berry and Tamer (2006) studied the identification of an optimal stopping model when the disturbance distribution conditional on observable states is known. An optimal stopping problem is qualitatively different from dynamic binary choice models in that the decision to stop brings an end to the process. More importantly, the expected current and future payoffs from stopping is independent of payoffs from not stopping, and therefore can be normalized to zero for identification. In their paper Berry and Tamer showed when $F_{\varepsilon|\mathbf{X}}$ is known, the single-period payoffs for not stopping can be fully nonparametrically identified. Whether $\mathbf{u}$ can be identified with knowledge of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ in a general dynamic binary choice process is an

[6] Aguirregabiria (2005a) was the first to show that $\Delta\delta(\mathbf{x})$ can be decomposed into $\Delta\omega(\mathbf{x})$ and $\Delta\xi(\mathbf{x})$ when the space of observed states $\mathbf{X}$ is finite.

[7] Recursive substitution in (3.3) implies $\omega_j(\mathbf{x}_t) = \sum_{s=0}^{\infty} \beta^s E_{js}[u_j(\mathbf{x}_{t+s})|\mathbf{x}_t]$, where $E_{js}(.|\mathbf{x}_t)$ is the expectation with respect to the distribution induced by *unconditionally choosing j for s* consecutive periods after the current state $\mathbf{x}_t$.

open question not addressed by the literature so far. Aguirregabiria (2005) showed when $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is known, the differences between the expected payoff from two sequences of actions can be identified. The first is to take action 1 in the current period and 0 forever in the future, and the second is to take action 0 both now and in the future forever.

In this section, we show the knowledge of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ helps identify $\mathbf{u}$ under fairly general rank conditions on observable state transitions and locational normalization $u_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbf{S}(\mathbf{X})$. First note locational normalization of $\mathbf{u}_0$ is necessary for identification. To see this, let $\mathbf{u}_j$ denote a K-vector with its k-th element being $u_j(\mathbf{x}_k)$. Let $\mathbf{G}^j$ denote matrices of transitions with the (m, n)-th component defined as $G^j_{m,n} = \Pr(\mathbf{x}_n|\mathbf{x}_m, j)$.

REG (v) $\mathbf{G}^j_\infty \equiv \lim_{T\to\infty} \sum_{t=1}^{T} \beta^t [\mathbf{G}^j]^t$ exists, where $[\mathbf{G}^j]^t$ is the t-th power of $\mathbf{G}^j$.

By recursive substitution, the left hand side of (3.2) is $(\mathbf{I}+\mathbf{G}^1_\infty)\mathbf{u}_1 - (\mathbf{I}+\mathbf{G}^0_\infty)\mathbf{u}_0$, where $\mathbf{I}$ is the K-by-K identity matrix. On the other hand, the right hand side is known for any fixed $F_{\Delta\varepsilon|\mathbf{X}}$. Therefore (3.2) is a system of K linear equations for $2K$ variables, with infinitely many solutions. The corollary below shows that once $\mathbf{u}_0$ is normalized to be a zero vector, $\mathbf{u}_1$ can be identified under fairly week rank conditions on observable state transitions.

Corollary 1 *(Proposition 1)* Suppose *AS, CI, REG (i)-(v)* and *KD* are satisfied. Then $(\mathbf{I} + \mathbf{G}^1_\infty)\mathbf{u}_1 - (\mathbf{I} + \mathbf{G}^0_\infty)\mathbf{u}_0$ is identified from the conditional choice probabilities observed. If $\mathbf{u}_0$ is normalized to $\mathbf{0}$, then $\mathbf{u}_1$ is uniquely recovered in $\mathbb{R}^K$ under the singleton $\mathcal{F}$ if and only if the matrix $(\mathbf{I} - \beta\mathbf{G}^1)$ has full rank.

To my knowledge, this is the first result in the literature that specifies conditions for non-parametric identification of $\Delta\mathbf{u}$ with the knowledge of $F_{\Delta\varepsilon|\mathbf{X}}$ and the normalization of $\mathbf{u}_0$ to the zero vector. It also reveals precisely the impact of normalizing $\mathbf{u}_0$ on the identification of $\mathbf{u}_1$. We propose the following algorithm for nonparametric estimation of $\mathbf{u}_1$ with the knowledge of $F_{\Delta\varepsilon|\mathbf{X}}$.

Step 1: estimate $p(\mathbf{x})$ nonparametrically;

Step 2: use knowledge of $F_{\Delta\varepsilon|\mathbf{X}}$ and first-step estimates $\hat{p}$ to compute $\Delta\xi(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}}, \hat{p})$;

Step 3: calculate $\Delta\hat{\omega}(\mathbf{x}) \equiv F^{-1}_{\Delta\varepsilon|\mathbf{X}}(\hat{p}(\mathbf{x})|\mathbf{x}) - \Delta\xi(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}}, \hat{p})$;

Step 4: estimate $\mathbf{G}^j$ nonparametrically and check the rank condition;

Step 5: calculate $\Delta\hat{u}(\mathbf{x}) \equiv (\mathbf{I} - \beta\hat{\mathbf{G}}^j)\Delta\hat{\omega}(\mathbf{x})$.

Relative to most of the maximum-likelihood based estimation procedure in the literature, an obvious advantage of this algorithm is that it circumvents the numerically intensive task of solving for fixed points through iterations, and then maximizing the likelihood over the space of payoff parameters. Instead, with knowledge of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$, $\Delta\hat{u}(\mathbf{x})$ is computed by directly plugging in preliminary kernel estimates. A direction for future research is to find regularity conditions on $F_{\Delta\varepsilon|\mathbf{X}}$, $\mathbf{G}$ and the kernels in step 1 and 4 that could deliver desirable asymptotic properties of the estimator.

3.4. Identification with Unknown Transitions

The identification result in the section above reveals what can be learned about $\mathbf{u}$ from the history of actions and observable states, with full knowledge of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$. However, in practice, econometricians do not always have the luxury of knowing $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$. This section studies the identification in the cases where $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is known to belong to a parametric family, or to satisfy certain stochastic restrictions such as median and statistical independence. As before we focus on the case where the space of observable states $S(\mathbf{X})$ is finite with K elements, and leave the generalization to infinite spaces for future research.

3.4.1. Parametric identification

First, we study below the identifying power of the most restrictive assumptions on $\mathcal{F}$. This is the case where structural parameters are known to belong to parametric families.

PAR (Parametric Family) For all $\Delta\epsilon$ *and* $\mathbf{x}$, $F_{\Delta\varepsilon|\mathbf{X}}(\Delta\epsilon|\mathbf{x}) = \bar{F}(\Delta\epsilon, \mathbf{x}; \boldsymbol{\theta}_F)$, $u_j(\mathbf{x}) = \bar{u}_j(\mathbf{x}; \boldsymbol{\theta}_u)$, *where* $\bar{F}$ *and* $\bar{\mathbf{u}}$ *are known up to finite dimensional parameters* $\boldsymbol{\theta}_F$ *and* $\boldsymbol{\theta}_u$, *and are continuously differentiable in* $\boldsymbol{\theta}_F$ *and* $\boldsymbol{\theta}_u$ *respectively for all* $\mathbf{x}$.

Let $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_u, \boldsymbol{\theta}_F)$, and $\boldsymbol{\omega}_{j,l}(\boldsymbol{\theta}_u)$, $\mathbf{u}_{j,l}(\boldsymbol{\theta}_u)$, $\boldsymbol{\kappa}_{j,l}(\boldsymbol{\theta}_F)$ and $\boldsymbol{\xi}_{j,l}(\boldsymbol{\theta}_F)$ denote four K-vectors with the k-th coordinates being $\frac{\partial\omega_j(\mathbf{x}_k)}{\partial\theta_{u,l}}$, $\frac{\partial u_j(\mathbf{x}_k)}{\partial\theta_{u,l}}$, $\frac{\partial\kappa_j(\mathbf{x}_k)}{\partial\theta_{F,l}}$ and $\frac{\partial\xi_j(\mathbf{x}_k)}{\partial\theta_{F,l}}$ respectively and l being the index for coordinates of $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_F$. Then by definition of ω_j and ξ_j in (3.3) and (3.4),

$$\begin{aligned}\boldsymbol{\omega}_{j,l}(\boldsymbol{\theta}_u) &= (\mathbf{I} - \beta\mathbf{G}^j)^{-1}\mathbf{u}_{j,l}(\boldsymbol{\theta}_u) \\ \boldsymbol{\xi}_{j,l}(\boldsymbol{\theta}_F) &= (\mathbf{I} - \beta\mathbf{G}^j)^{-1}\beta\mathbf{G}^j\boldsymbol{\kappa}_{j,l}(\boldsymbol{\theta}_F)\end{aligned}$$

By definition, the choice probabilities $p(\mathbf{x}_k; \boldsymbol{\theta}) = \bar{F}[\Delta\omega(\mathbf{x}_k; \boldsymbol{\theta}_u) + \Delta\xi(\mathbf{x}_k; \boldsymbol{\theta}_F), \mathbf{x}_k; \boldsymbol{\theta}_F]$. Then the gradient with respect to the parameters is $\nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k; \boldsymbol{\theta}) = [\nabla_{\boldsymbol{\theta}_u} p(\mathbf{x}_k; \boldsymbol{\theta}) \ \nabla_{\boldsymbol{\theta}_F} p(\mathbf{x}_k; \boldsymbol{\theta})]'$, where

$$\begin{aligned}\nabla_{\boldsymbol{\theta}_u} p(\mathbf{x}_k; \boldsymbol{\theta}) &= \bar{f}(\mathbf{x}_k; \boldsymbol{\theta})[\nabla_{\theta_u}\omega_1(\mathbf{x}_k; \boldsymbol{\theta}_u) - \nabla_{\theta_u}\omega_0(\mathbf{x}_k; \boldsymbol{\theta}_u)] \\ \nabla_{\boldsymbol{\theta}_F} p(\mathbf{x}_k; \boldsymbol{\theta}) &= \bar{f}(\mathbf{x}_k; \boldsymbol{\theta})[\nabla_{\theta_F}\xi_1(\mathbf{x}_k; \boldsymbol{\theta}_F) - \nabla_{\theta_F}\xi_0(\mathbf{x}_k; \boldsymbol{\theta}_F)] + \bar{F}_{\boldsymbol{\theta}_F}(\mathbf{x}_k; \boldsymbol{\theta})\end{aligned}$$

with

$$\begin{aligned}\bar{f}(\mathbf{x}; \boldsymbol{\theta}) &\equiv \nabla_\tau \bar{F}(\tau, \mathbf{x}; \boldsymbol{\theta}_F)|_{\tau=\Delta\omega(\mathbf{x};\boldsymbol{\theta}_u)+\Delta\xi(\mathbf{x};\boldsymbol{\theta}_F)} \\ \bar{F}_{\boldsymbol{\theta}_F}(\mathbf{x}; \boldsymbol{\theta}) &\equiv \nabla_{\theta_F}\bar{F}[\tau, \mathbf{x}; \boldsymbol{\theta}_F]|_{\tau=\Delta\omega(\mathbf{x};\boldsymbol{\theta}_u)+\Delta\xi(\mathbf{x};\boldsymbol{\theta}_F)}\end{aligned}$$

The following proposition gives sufficient conditions for the local identification of $\boldsymbol{\theta}$.

Proposition 2 Let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_u^0, \boldsymbol{\theta}_F^0)$ denote the vector of true structural parameters. Suppose *AS*, *CI*, *REG* and *PAR* are satisfied, and $\nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k; \boldsymbol{\theta}_0)$ exists and is continuous at $\boldsymbol{\theta}_0$ for all $\mathbf{x}_k$. If $\nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k; \boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k; \boldsymbol{\theta}_0)'$ has full rank for some $\mathbf{x}_k \in S(\mathbf{X})$, then there exists an open neighborhood around $\boldsymbol{\theta}_0$ that contains no other $\boldsymbol{\theta}$ observationally equivalent to $\boldsymbol{\theta}_0$.

Next we show an example of how *Proposition 2* can be applied to check local identification when single-period returns take a linear index specification, and $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is uniform and independent of $\mathbf{X}$.

Example 1 Let $u_j(\mathbf{x}) = \mathbf{x}'\boldsymbol{\gamma}_j$ for all $\mathbf{x}$, and let $\Delta\varepsilon$ be independent from $\mathbf{x}$ and distributed as uniform on $[0, a]$. Then $\theta_u^0 = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_0)$, $\theta_F^0 = a$. Let $\mathbf{X}$ denote a K-by-d matrix with the k-th row being $\mathbf{x}_k$. Let $\mathbf{P}_0$ and $\mathbf{P}_1$ denote K-vectors with k-th coordinates being $p(\mathbf{x}_k)^2$ and $[1 - p(\mathbf{x}_k)]^2$ respectively. Then $\boldsymbol{\kappa}_j(a) = \frac{a}{2}\mathbf{P}_j$, $\boldsymbol{\xi}_j(a) = (\mathbf{I} - \beta\mathbf{G}^j)^{-1}\beta\mathbf{G}^j\boldsymbol{\kappa}_j(a)$, and $\boldsymbol{\omega}_j(\boldsymbol{\gamma}_j) = (\mathbf{I} - \beta\mathbf{G}^j)^{-1}\mathbf{X}\boldsymbol{\gamma}_j$. Let $\Delta\boldsymbol{\omega}(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) = \boldsymbol{\omega}_1(\boldsymbol{\gamma}_1) - \boldsymbol{\omega}_0(\boldsymbol{\gamma}_0)$, $\Delta\boldsymbol{\xi}(a) = \boldsymbol{\xi}_1(a) - \boldsymbol{\xi}_0(a)$, and $\mathbf{1}_k$ denote a unit column vector with the k-th element being 1. It follows from some algebra:

$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}_u} p(\mathbf{x}_k; \boldsymbol{\theta}) &= \frac{1}{a}[\nabla_{\boldsymbol{\theta}_u}\boldsymbol{\omega}_1(\boldsymbol{\gamma}_1) - \nabla_{\boldsymbol{\theta}_u}\boldsymbol{\omega}_0(\boldsymbol{\gamma}_0)]'\mathbf{1}_k \\
\nabla_{\boldsymbol{\theta}_F} p(\mathbf{x}_k; \boldsymbol{\theta}) &= \{\frac{1}{a}[\nabla_a\boldsymbol{\xi}_1(a) - \nabla_a\boldsymbol{\xi}_0(a)] + \bar{F}_{\boldsymbol{\theta}_F}(\boldsymbol{\theta})\}'\mathbf{1}_k
\end{aligned}
$$

where

$$
\begin{aligned}
\nabla_a\boldsymbol{\xi}_j(a) &= \frac{1}{2}(\mathbf{I} - \beta\mathbf{G}^j)^{-1}\beta\mathbf{G}^j\mathbf{P}_j \\
\nabla_{\boldsymbol{\theta}_u}\boldsymbol{\omega}_j(\boldsymbol{\gamma}_j) &= (\mathbf{I} - \beta\mathbf{G}^j)^{-1}\mathbf{X} \\
\bar{F}_{\boldsymbol{\theta}_F}(\boldsymbol{\theta}) &= -a^{-2}[\Delta\boldsymbol{\omega}(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) + \Delta\boldsymbol{\xi}(a)]
\end{aligned}
$$

The rank condition in Proposition 2 can then be checked for all $\mathbf{x}_k$ using knowledge of the primitives $\mathbf{G}^j$, β, and the observable $\mathbf{P}_j$. ■

Directions for future research along the line of parametric identification includes checking local identifications for different parametric families, as well as search for primitive conditions for global identification within certain parametric families.

3.4.2. Semiparametric identification

In practice, it is not always justifiable to restrict the set of possible error distributions $\mathcal{F}$ to a parametric family. On the other hand, on most occasions it is plausible to introduce stochastic restrictions such as the conditional symmetry or statistical independence of $F_{\Delta\varepsilon|\mathbf{X}}$. In this section, we study the identifying power of these restrictions. The definition of the joint identification region (3.2) suggests any stochastic restrictions on $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ will have to interact with observable transitions $\mathbf{G}^j$ to give identifying power. However, $\mathbf{G}^j$ is observable and researchers do not have the freedom to put restrictions on them. In this subsection, we first look at what can be learned about $\mathbf{u}$ under the conditional symmetry and statistical independence of $\Delta\varepsilon$ from $\mathbf{x}$. We will also characterize the identification region of $\mathbf{u}$ under these restrictions, and discuss the impact of particular properties of $\mathbf{G}^j$ on identification.

3.4.2.1. Review of static binary choice models. Manski (1988) gave a thorough treatment of the identification of the threshold-crossing model of binary response, where the binary outcome y is determined by an observable random vector $\mathbf{x}$ and by an unobservable scalar ε through a model $y = 1\{\mathbf{x}\boldsymbol{\beta} + \varepsilon \geq 0\}$. The conditional distribution $F_{\varepsilon|\mathbf{X}}$ is continuous and strictly increasing. Given these maintained restrictions, he investigated the identifiability of $\boldsymbol{\beta}$ under different restrictions on $F_{\varepsilon|\mathbf{X}}$. These include mean independence, quantile independence, index sufficiency, statistical independence and the case where $F_{\varepsilon|\mathbf{X}}$

is completely known. He found that (1) mean independence has no identifying power; (2) quantile independence implies that $\boldsymbol{\beta}$ is identified up to scale, provided that the distribution of $\mathbf{x}$ has sufficiently rich support; (3) index sufficiency can identify the slope components of β up to scale and sign, also under certain rich support condition on $\mathbf{x}$; (4) statistical independence subsumes both quantile and statistical independence and therefore implies all positive findings above; and (5) when ε is independent from $\mathbf{x}$ with a known distribution, identification of $\boldsymbol{\beta}$ only requires the distribution of $\mathbf{x}$ to have full rank.

Matzkin (1992) discussed nonparametric identification of static binary choice models, where $\Pr(d = 1|\mathbf{x}) = \Pr(\Delta\varepsilon \leq u_1(\mathbf{x}) - u_0(\mathbf{x}))$ and $\Delta\varepsilon \equiv \varepsilon_0 - \varepsilon_1$. She maintained the assumption of statistical independence between $\Delta\varepsilon$ and $\mathbf{x}$, and normalized $u_0(\mathbf{x}) = 0$ for all $\mathbf{x}$. She showed that, under regularity conditions such as continuous support of $\mathbf{x}$ and strict monotonicity of $F_{\Delta\boldsymbol{\varepsilon}}$, the true parameters $\mathbf{u}$ and $F_{\Delta\boldsymbol{\varepsilon}}$ can be uniquely recovered from the choice probabilities $\Pr(d = 1|\mathbf{x})$ within a set of utility functions such that $\nexists\, \mathbf{u}, \tilde{\mathbf{u}}\ s.t.\ u_1(\mathbf{x}) - u_0(\mathbf{x})$ is a monotone transformation of $\tilde{u}_1(\mathbf{x}) - \tilde{u}_0(\mathbf{x})$. More interestingly, she constructed such sets of utility functions using restrictions with economic substances such as monotonicity, concavity and homogeneity. She also generalized this positive identification result under statistical independence from binary to polychotomous choice models in Matzkin (1992).

In Section 3 above, we have already shown that when $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is known, $\Delta\mathbf{u}$ can be nonparametrically identified under fairly weak rank conditions of $\mathbf{G}$ if $\mathbf{u}_0$ is normalized to $\mathbf{0}$. This is the dynamic analog of the positive result in the static case in Manski (1988). On the other hand, as it turns out below, there is no direct analog of the static case in the identification of dynamic binary choice processes under the stochastic restrictions such as conditional symmetry and statistical independence. This is not a surprising result as the definition of the joint identification region suggests any restriction on $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ has to interact with $\mathbf{G}^j$ to give identifying power.

3.4.2.2. Statistical independence. Statistical independence is a strong stochastic restriction that implies several popular weaker forms in the literature of semiparametric identifications. These include both distributional and mean index sufficiency, as well as distributional and mean exclusion restrictions. By construction, the identification region of $\mathbf{u}$ under statistical independence is a strict subset of that under any of these weaker stochastic restrictions.

SI (Statistical Independence) $\Delta\varepsilon$ is statistically independent from observable states $\mathbf{x}$*, and is continuously distributed with positive density on a closed interval* $[\varepsilon_L, \varepsilon_U]$.

Without further restrictions on $\mathbf{u}$, the difference between single-period returns is not unidentified under *SI*. To see this, note that given observable choice probabilities, the joint identification region under SI is

$$\Theta_I \equiv \{(\mathbf{u},F_{\Delta\varepsilon}) : \Delta\omega(\mathbf{x};\mathbf{u}) = F_{\Delta\varepsilon}^{-1}(p(\mathbf{x})) - \Delta\xi(\mathbf{x}; F_{\Delta\varepsilon}, p) \ a.e.\ \mathbf{x}\}$$

where $\Delta\omega$ and $\Delta\xi$ are defined as in *Proposition 1* above, with κ_d simplified as

$$\kappa_0(\mathbf{x}; p, F_{\Delta\varepsilon}) \equiv \int_{\varepsilon_L}^{q(\mathbf{x})} F_{\Delta\varepsilon}(s)ds \tag{3.5}$$

$$\kappa_1(\mathbf{x}; p, F_{\Delta\varepsilon}) \equiv -\int_{\varepsilon_L}^{q(\mathbf{x})} 1 - F_{\Delta\varepsilon}(s)ds \tag{3.6}$$

where $q(\mathbf{x}) \equiv F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}))$. Then it can be shown that for a given pair of true parameter $(\mathbf{u}^*, F_{\boldsymbol{\Delta\varepsilon}}^*)$, it is always possible to perturb $F_{\boldsymbol{\Delta\varepsilon}}^*$ slightly to $F'_{\Delta\varepsilon}$ and find a corresponding $\mathbf{u}'$ that is close to $\mathbf{u}^*$ such that $(\mathbf{u}', F'_{\boldsymbol{\Delta\varepsilon}})$ is observationally equivalent to $(\mathbf{u}^*, F_{\boldsymbol{\Delta\varepsilon}}^*)$. Hence $\mathbf{u}^*$ is not locally identified.

Let $\mathcal{F}_{SI}$ be the set of $F_{\Delta\varepsilon|\mathbf{X}}$ that satisfies the *SI* restriction. The next question is how to characterize the identification region of $\mathbf{u}$ under $\mathcal{F}_{SI}$. Intuitively this is the set of all $\mathbf{u}$ for which there exists certain $F_{\Delta\varepsilon|\mathbf{X}}$ in $\mathcal{F}_{SI}$ such that $p(\mathbf{x};\mathbf{u},F_{\Delta\varepsilon|\mathbf{X}}) = p^*(\mathbf{x})$. The next corollary of *Proposition 1* formalizes this idea. Without loss of generality, order $(\mathbf{x}_1,..,\mathbf{x}_K)$ such that $p(\mathbf{x}_m) \leq p(\mathbf{x}_n)$ for $m \leq n$. Let $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$ are K-by-1 vectors, with $Q_k \equiv F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_k))$ and $\kappa_k^0 \equiv \kappa_0(\mathbf{x}_k;F_{\Delta\varepsilon})$.

Corollary 2 *(Proposition 1)* Suppose *AS*, *CI, SI* and *REG* are satisfied, and normalize $\mathbf{u}_0 = \mathbf{0}$. Then the identification region of $\mathbf{u}_1$ under $\mathcal{F}_{SI}$ is $\Theta_{u_1} \equiv \{\mathbf{u}_1 \in \mathbb{R}^K : \Psi_{SI}$ has solutions in $\mathbf{Q}$ and $\boldsymbol{\kappa}^0\}$, where Ψ_{SI} is a system of linear inequalities defined as:

$$\begin{aligned}
(3.7)\qquad (\mathbf{I}-\beta\mathbf{G}^1)^{-1}(\mathbf{u}_1-\mathbf{Q}) &= [(\mathbf{I}-\beta\mathbf{G}^0)^{-1}\beta\mathbf{G}^0-(\mathbf{I}-\beta\mathbf{G}^1)^{-1}\beta\mathbf{G}^1]\boldsymbol{\kappa}^0\\
\varepsilon_L &\leq Q_1 \leq Q_2 \leq .. \leq Q_K \leq \varepsilon_U\\
0 &\leq \kappa_1^0 \leq p(\mathbf{x}_1)(Q_1-\varepsilon_L)\\
p(\mathbf{x}_{k-1})(Q_k-Q_{k-1}) &\leq \kappa_k^0-\kappa_{k-1}^0 \leq p(\mathbf{x}_k)(Q_k-Q_{k-1}),\ for\ k=2,.,K
\end{aligned}$$

Example 2 For simplicity in algebra, consider the special case of optimal stopping problem where one of two actions $j=0$ is irreversible and leads to zero expected payoffs both in the current and future periods. Let the single-period return for $j=1$ be $U_1(\mathbf{x},\varepsilon) = u(\mathbf{x})-\varepsilon$, where the unobserved state ε is continuously distributed on a closed interval $[\varepsilon_L,\varepsilon_U]$ in $\mathbb{R}^1$. Let the transition when $j=1$ be $G(\mathbf{x}'|\mathbf{x})$. Then by similar arguments, the joint identification region is $\Theta_I \equiv \{(u,F_\varepsilon) : F_\varepsilon^{-1}(p(\mathbf{x})) = u(\mathbf{x}) + \beta\int\kappa_0(\mathbf{x}';F_\varepsilon,p)dG(\mathbf{x}'|\mathbf{x})\ a.e.\ \mathbf{x}\}$, where ω and ξ are fixed points defined as before. Now consider the case for $K=2$. Without loss of generality, order $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $p(\mathbf{x}_1) \leq p(\mathbf{x}_2)$. Let $\mathbf{u}$ be a 2-by-1 vector with the k-th coordinate being $u(\mathbf{x}_k)$. Denote $\mathbf{G} = [1-a\ a; 1-b\ b]$. The joint identification region is given by the following linear system $\mathbf{u} = \mathbf{Q} - \beta \mathbf{G} \boldsymbol{\kappa}$, where $\mathbf{Q}$ and $\boldsymbol{\kappa}$ are 2-by-1 vectors, with $q_k \equiv F_\varepsilon^{-1}(p(\mathbf{x}_k))$ and $\kappa_k \equiv \kappa_0(\mathbf{x}_k; F_\varepsilon)$ for $k = 1, 2$. Let p_k denote $p(\mathbf{x}_k)$. By definition, coordinates of $\mathbf{Q}$ and $\boldsymbol{\kappa}$ have to satisfy the following restrictions:

$$q_1 \geq \varepsilon_L; q_2 - q_1 \geq 0; -q_2 + \varepsilon_U \geq 0;$$

$$\kappa_1 \geq 0; -\kappa_1 + p_1(q_1 - \varepsilon_L) \geq 0;$$

$$\kappa_2 - \kappa_1 - p_1(q_2 - q_1) \geq 0; \kappa_1 - \kappa_2 + p_2(q_2 - q_1) \geq 0$$

Applying the Fourier-Motzkin procedure of iterated eliminations suggests the system has solutions $(q_1, q_2, \kappa_1, \kappa_2)$ as long as the following conditions are satisfied:

$$u_2 \geq \varepsilon_L; \quad u_1 - \varepsilon_L + \beta a p_2(\varepsilon_U - \varepsilon_L) \geq 0$$

$$(1 - \beta b p_2)(u_1 - \varepsilon_L) + \beta a p_2(u_2 - \varepsilon_L) \geq 0$$

This gives a characterization of the identification region of (u_1, u_2) under the statistical independence of ε from $\mathbf{x}$. ■

Of course, a general dynamic binary choice model is qualitatively different from an optimal stopping problem. But the example above shares the same basic idea in the general case in that the identification regions are defined by checking the consistency of a system of linear inequalities. Though the corollary is given with no restriction on $\mathbf{u}$, the methodology extends immediately to semiparametric cases where $\mathbf{u} = [u_1(\mathbf{x}; \boldsymbol{\theta}_u) \ u_2(\mathbf{x}; \boldsymbol{\theta}_u)]$ is known up to finite dimensional parameters. The identification region of θ_u under $\mathcal{F}_{SI}$ is simply the set of values for which the system Ψ_{SI} has solutions in $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$. In the example given above, the identification region is rather wide. However, the size of the set itself is informative, as it reveals what can be learned about $\mathbf{u}$ under the assumption of statistical independence between $\Delta\varepsilon$ and $\mathbf{x}$.

3.4.2.3. Conditional symmetry. Conditional symmetry is the strongest locational stochastic restriction. It implies both mean and median independence.

CS (Conditional Symmetry) $F_{\Delta\varepsilon|\mathbf{X}}$ *is symmetric around* ε_M *conditional on all observable states* $\mathbf{x}$*, and is continuously distributed with positive density on a known closed interval* $[\varepsilon_M - C, \varepsilon_M + C]$.[8]

Let $\mathcal{F}_{CS}$ be the set of $F_{\Delta\varepsilon|\mathbf{X}}$ that satisfies the *CS* restriction. The identification region of $\mathbf{u}$ under $\mathcal{F}_{CS}$ can be characterized by similar arguments as in Corollary 2, except that the inequality constraints will take a different form. Without loss of generality, let $p(\mathbf{x}_1) \leq p(\mathbf{x}_2) \leq .. \leq p(\mathbf{x}_M) \leq \frac{1}{2}$, and $\frac{1}{2} \leq p(\mathbf{x}_{M+1}) \leq .. \leq p(\mathbf{x}_K)$. Let $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$ denote K-by-1 vectors with the k-th coordinate being $Q_k \equiv F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x}_k)|\mathbf{x}_k)$ and $\kappa_0(\mathbf{x}_k; F_{\Delta\varepsilon|\mathbf{X}}) = \int_{\varepsilon_L}^{Q_k} F_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x}_k)ds$ respectively.

Corollary 3 *(Proposition 1)* Suppose *AS, CI, CS* and *REG* are satisfied, and normalize $\mathbf{u}_0 = \mathbf{0}$. Then the identification region of $\mathbf{u}_1$ under $\mathcal{F}_{SI}$ is $\Theta_{u_1} \equiv \{\mathbf{u}_1 \in \mathbb{R}^K : \Psi_{CS}$ has solutions in $\mathbf{Q}$ and $\boldsymbol{\kappa}^0\}$, where Ψ_{CS} is a system of linear inequalities defined as:

$$\begin{aligned}
(\mathbf{I} - \beta\mathbf{G}^1)^{-1}(\mathbf{u}_1 - \mathbf{Q}) &= [(\mathbf{I} - \beta\mathbf{G}^0)^{-1}\beta\mathbf{G}^0 - (\mathbf{I} - \beta\mathbf{G}^1)^{-1}\beta\mathbf{G}^1]\boldsymbol{\kappa}^0 \\
\varepsilon_M - C &\leq Q_k \leq \varepsilon_M,\ k \leq M \\
\varepsilon_M &\leq Q_k \leq \varepsilon_M + C,\ k > M \\
0 &\leq \kappa_k^0 \leq p(\mathbf{x}_k)(Q_k - \varepsilon_M + C),\ \forall k \leq M \\
Q_k - \varepsilon_M &\leq \kappa_k^0 \leq C - p(\mathbf{x}_k)[\varepsilon_M + C - Q_k],\ \forall k > M
\end{aligned}$$

[8] I focus on the case with fixed support of $\Delta\varepsilon$ for the sake of simplicity in explanation. The methodology proposed below can be generalized to allow supports to vary with $\mathbf{x}$.

Example 3 Again consider the special case of optimal stopping problem with $K = 2$. Suppose $p(\mathbf{x}_1) < \frac{1}{2} < p(\mathbf{x}_2)$. As before, the joint identification region is given by the following linear system $\mathbf{u} = \mathbf{Q} - \beta\mathbf{G}\boldsymbol{\kappa}$, subject to the inequalities above. Applying the procedures of Fourier-Motzkin eliminations shows solutions exists if and only if $\varepsilon_M \geq u_1$ and $\varepsilon_M \geq u_2$. This is the identification region of $\mathbf{u}$ under the conditional symmetry of $F_{\varepsilon|\mathbf{X}}$. ■

3.4.2.4. Restrictions on observable transitions. As discussed earlier in this section, restrictions on $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ can only impact the identification region of $\mathbf{u}$ through interactions with observable transitions $\mathbf{G}^j$. In practice it is possible that $\mathbf{G}^j$ might be invariant within certain subset of observable states. Knowledge of such a property can contribute to the approach of identification by checking the feasibility of systems of linear inequalities.

SUB (Subset Invariance) $\exists\bar{\mathbf{X}} \subset S(\mathbf{X})$ such that $G^j(.|\mathbf{x}) = G^j(.|\tilde{\mathbf{x}})$ for $j = 0, 1$ and all $\mathbf{x}, \tilde{\mathbf{x}} \in \bar{\mathbf{X}}$.

To incorporate this into the framework of linear system, consider the case under SI with a finite observable state space of K elements. Let $\bar{K} \subset K$ denote the set of states with invariant observable transitions. Note the linear equalities in the systems Ψ_{SI} can be written as

$$\mathbf{u}_1 - \mathbf{Q} = -\left[\sum_{t=1}^{\infty} \beta^t(\mathbf{G}^1)^t\right]\mathbf{u}_1 - \left[\sum_{t=1}^{\infty} \beta^t(\mathbf{G}^1)^t\right]\boldsymbol{\kappa}^1 + \left[\sum_{t=1}^{\infty} \beta^t(\mathbf{G}^0)^t\right]\boldsymbol{\kappa}^0$$

The right-hand side is a K-by-1 vector, whose coordinates in $\bar{K}$ are identical under *SUB*. It follows that for all $k_1, k_2 \in \bar{K}$, $u_1(\mathbf{x}_{k_1}) - u_1(\mathbf{x}_{k_2}) = F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_{k_1})) - F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_{k_2}))$. Therefore, a necessary condition for feasibility of the system is that $u_1(\mathbf{x}_k)$ has to be ranked in the same order as $F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_k))$ for $k \in \bar{K}$. Similar necessary conditions also exists for the case under CS restrictions, except that $u_1(\mathbf{x}_k)$ in that case need to be ranked in the same order as $F_{\Delta\varepsilon|\mathbf{X}}^{-1}(p(\mathbf{x}_k)|\mathbf{x}_k)$ instead of $F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_k))$ on $\bar{K}$.

3.5. Conclusions

In this paper, we have introduced a new approach for studying identification of structural parameters in dynamic binary choice processes. The approach is based on characterizing the joint identification region Θ_I of single-period payoffs $\mathbf{u}$ and disturbance distributions $F_{\varepsilon|\mathbf{X}}$ through a system of linear equations in these parameters. Using this framework, we show that with knowledge of the distribution of disturbances, the differences between two trivial policies of choosing one of the two actions forever can be uniquely recovered. Furthermore the identification region of $\mathbf{u}$ under various stochastic restrictions on the nuisance parameter $F_{\varepsilon|\mathbf{X}}$ can be defined as the set of $\mathbf{u}$ for which there exist nuisance parameters which satisfy the linear equations characterizing Θ_I subject to inequality constraints implied by these restrictions. Under this framework, we show through examples that both the conditional symmetry and the statistical independence of unobservable states have limited identifying power on single-period payoffs. This approach of identification through linear programming can be extended immediately to study the identifying power of any parametric or shape restrictions on single-period payoffs. Directions of future research include the search for restrictions on $\mathbf{u}$ and $F_{\varepsilon|\mathbf{X}}$ that can deliver greater identifying power, as well as the definition and statistical properties of new estimators that make use of the positive identification results.

References

[1] Aguirregabiria, V., "Another Look at the Identification of Dynamic Discrete Decision Processes", working paper, Boston University, 2005

[2] Aguirregabiria, V. and P. Mira, "Dynamic Discrete Choice Structural Models: A Survey", working paper, University of Toronto July 2007

[3] Andrews, D.W.K.(1994): "Empirical Process Methods in Econometrics", Handbook of Econometrics, Volume IV, Chapter 37, Elsevier Science

[4] Aradillas-López, A. (2007): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information", Working Paper, Princeton University

[5] Athey, S. and P. Haile (2002), "Identification of Standard Auction Models," Econometrica 70:2107-2140

[6] Athey, S. and P. Haile (2005), "Nonparametric Approaches to Auctions," Handbook of Econometrics, Vol.6, forthcoming

[7] Bajari, P., Hong, H., Krainer, J. and D. Nekipelov (2007): "Estimating Static Models of Strategic Interactions", Working Paper, University of Minnesota

[8] Bajari, P. and A. Hortacsu (2003), "The Winner's Curse, Reserve Prices, and Endogenous Entry" Empirical Insights from eBay Auctions," RAND Journal of Economics 34:329-355

[9] Berry, S. and E. Tamer, "Identification in Models of Oligopoly Entry," working paper, April 2006.

[10] Buchinsky, M. and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model", Econometrica 66, p653-672.

[11] Butler, Alexander W. (2007), "Distance Still Matters: Evidence from Municipal Bond Underwriting"

[12] Chamberlain, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models with Censoring", Journal of Econometrics, Volume 32, Issue 2, July 1986, Pages 189-218

[13] Chernozhukov, V., Tamer, E. and H. Hong, "Identification and Inference on Identified Parameter Sets", Econometrica 2008

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

[15] Haile, P., H. Hong and M. Shum (2004) "Nonparametric Tests for Common Values in First-Price Sealed Bid Auctions," mimeo, Yale University

[16] Haile, P. and E. Tamer (2003), "Inference with an Incomplete Model of English Auctions," Journal of Political Economy, vol.111, no.1

[17] Hardle, W. (1991), Smoothing Techniques with Implementation in S. New York : Springer-Verlag, 1991

[18] Hendel, I. and A. Nevo, "Measuring the Implications of Sales and Consumer Stockpiling Behavior", forthcoming Econometrica 2006

[19] Hendricks, K., J. Pinkse and R. Porter (2003), "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions," Review of Economic Studies 70:115-145

[20] Hendricks, K. and R. Porter (2007), "An Empirical Perspective on Auctions," Handbook of Industrial Organization, Vol 3, Elsevier

[21] Hong, H and M. Shum (2002), "Increasing Competition and the Winner's Curse: Evidence from Procurement"

[22] Horowitz, J. (1992):"A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, Econometric Society, 60(3), 505-31

[23] Ichimura, H. (1991), "Semiparametric Least Squares Estimation of Single Index Models," Journal of Econometrics

[24] Imbens and Manski (2004), "Confidence Intervals for Partially Identified Parameters", Econometrica, Vol 72, No 6.

[25] Keane, M. and K. Wolpin, "Estimating Welfare Effects Consistent with Forward-Looking Behavior", PIER Working Paper 01-019, University of Pennsylvania, 2000

[26] Klemperer, P. (1999) "Auction Theory: A Guide to the Literature," Journal of Economic Surveys 12:227-286

[27] Krishna, Vijay (2002), "Auction Theory", Academic Press

[28] Laffont, J. and Q. Vuong (1996), "Structural Analysis of Auction Data," American Economic Review, Papers and Proceedings 86:414-420

[29] Levin, D. and J. Smith (1994), "Optimal Reservation Prices in Auctions," Economic Journal 106:1271-1283

[30] Li, T., I. Perrigne, and Q. Vuong (2002), "Conditionally Independent Private Information in OCS Wildcat Auctions", Journal of Econometrics 98:129-161

[31] Li, T., I. Perrigne, and Q. Vuong (2003), "Estimation of Optimal Reservation Prices in First-price Auctions", Review of Economics and Business Statistics

[32] Magnac, T. and D. Thesmar, "Identifying dynamic discrete decision processes", Econometrica, 70, 2002, p801-816

[33] Manski, C. (1985): "Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data," Econometrica, 55(2), 357-362

[34] Manski, C. (1988): "Identification of Binary Response Models ", Journal of American Statistical Association, Vol 83, No. 403, p729-738

[35] Manski, C. and Tamer, E.(2002): "Inference on Regressions with Interval Data on a Regressor or Outcome", Econometrica, 70, 519-547

[36] Matzkin, R.L., "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold Crossing Models", Econometrica: Vol. 60, No. 2, 1992, p. 239.

[37] Matzkin, R.L., "Nonparametric Identification and Estimation of Polychotomous Choice Models," Journal of Econometrics, Vol. 58, 1993

[38] McAdams, D. (2006), "Uniqueness in First Price Auctions with Affiliation," Working Paper

[39] McAfee, R and J. McMillan (1987), "Auctions and Bidding," Journal of Economic Literature 25:669-738

[40] Milgrom, P. and R. Weber (2000), "A Theory of Auctions and Competitive Bidding," Econometrica, 50:1089-1122

[41] Newey, W.K. and McFadden, D (1994): "Large Sample Estimation and Hypothesis Testing", Handbook of Econometrics, Volume IV, Chapter 36, Elsevier Science

[42] Paarsch, H. (1997), "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbian Timber Sales," Journal of Econometrics 78:333-57

[43] Pagan and Ullah (1999): "Nonparametric Econometrics", Cambridge University Press.

[44] Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators", Econometric Theory, 7, 1991, p186-199

[45] Powell, J., J. Stock and T. Stocker (1989) "Semiparametric Estimation of Index Coefficients," Econometrica, 57: 1403-1430

[46] Rust, J., "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher", Econometrica, 55, p999-1033

[47] Rust, J., "Structural estimation of Markov decision processes", Handbook of econometrics Volume 4, North-Holland

[48] Securities Industry & Financial Markets Association Publications, 2005

[49] Shneyerov, A. (2006) "An Empirical Study of Auction Revenue Rankings: the Case of Municipal Bonds," Rand Journal of Economics, Forthcoming

[50] Temel, J. (2001), The Fundamentals of Municipal Bonds, 5th Edition, Wiley Finance

APPENDIX

Appendices for Chapter 1-3

1. Appendix for Chapter 1

1.1. Proofs of identification results

Proof of Proposition 1. To prove necessity, suppose $\{\theta, F_{\mathbf{X}}\} \in \Theta_{CV} \otimes \mathcal{F}$ generates $G^0_{\mathbf{B}}$ in such an equilibrium. Then the support of $\mathbf{B}$ is $S(\mathbf{B}) \equiv [b_L, b_U]^n$, with $b^0_L = v_h(x_L; \theta, F_{\mathbf{X}})$ and $b^0_U = b_0(x_U; \theta, F_{\mathbf{X}})$, where $v_h\,(x; \theta, F_{\mathbf{X}})$ is a shorthand for $v_{h,n}(x, x; \theta, F_{\mathbf{X}})$ (with subscripts for n dropped for notational ease). Note $\forall \mathbf{b} \in [b^0_L, b^0_U]^n$, $G^0_{\mathbf{B}}(\mathbf{b}) \equiv \Pr(\mathbf{b_0}(\mathbf{X};\theta, F_{\mathbf{X}}) \leq \mathbf{b}) = \Pr(\mathbf{X} \leq \mathbf{b_0}^{-1}(\mathbf{b})) \equiv F_{\mathbf{X}}(\mathbf{b_0}^{-1}(\mathbf{b};\theta, F_{\mathbf{X}}))$, where the equality follows from the strict monotonicity of equilibrium strategies. Then symmetry of the equilibrium and exchangeability of $F_{\mathbf{X}}$ implies $G^0_{\mathbf{B}}(\mathbf{b})$ is exchangeable in $\mathbf{b}$ $\forall \mathbf{b} \in S(\mathbf{B})$. The affiliation of $\mathbf{B} = (b_0(X_1; \theta, F_{\mathbf{X}}), .., b_0(X_n; \theta, F_{\mathbf{X}}))$ follows from the monotonicity of $b_0(.)$ and the affiliation of $\mathbf{X}$ (by Theorem 3 in Milgrom and Weber (1982)). The first-order condition (1.2) implies $\xi(b; G^0_{\mathbf{B}}) = v_h(b_0^{-1}(b); \theta, F_{\mathbf{X}})$ $\forall b \in [b^0_L, b^0_U]$, where $v_h\,(x; \theta, F_{\mathbf{X}})$ is increasing on the support of $F_{\mathbf{X}}$ by the definition of $(\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$. Hence the strict monotonicity of $b_0^{-1}(.; \theta, F_{\mathbf{X}})$ implies $\xi(b; G^0_{\mathbf{B}})$ is increasing over $[b_L, b_U]$. The proof of sufficiency makes use of the following claim and an example constructed below.

Claim A1 Suppose a bid distribution $G^0_{\mathbf{B}}$ satisfies the necessary conditions in Proposition 1. Then $\{\theta, F_{\mathbf{X}}\} \in \Theta \otimes \mathcal{F}$ rationalizes $G^0_{\mathbf{B}}$ in a first-price auctions if and only if

$$F_{\mathbf{X}}(\mathbf{x}) = G^0_{\mathbf{B}}(\xi^{-1}(v_h(x_1; \theta, F_{\mathbf{X}}); G^0_{\mathbf{B}}), .., \xi^{-1}(v_h(x_n; \theta, F_{\mathbf{X}}); G^0_{\mathbf{B}})) \tag{.1}$$

for all x on the support of $F_{\mathbf{X}}$.

Proof of Claim A1 Suppose $\{\theta, F_{\mathbf{X}}\} \in \Theta \otimes \mathcal{F}$ rationalizes such a $G^0_{\mathbf{B}}$. Then $F_{\mathbf{X}}(\mathbf{x}) = G^0_{\mathbf{B}}(\mathbf{b}_0(\mathbf{x};\theta, F_{\mathbf{X}}))$ for all $\mathbf{x} \in [x_L, x_U]^N$, where $b_0(.;\theta, F_{\mathbf{X}})$ is the equilibrium strategy characterized by the first-order condition (1.2), which implies $b_0(x;\theta, F_{\mathbf{X}}) = \xi^{-1}(v_h(x;\theta, F_{\mathbf{X}}); G^0_{\mathbf{B}})$ for all x on support by the monotonicity of $\xi(.; G^0_{\mathbf{B}})$. It follows $F_{\mathbf{X}}(\mathbf{x}) = G^0_{\mathbf{B}}(\boldsymbol{\xi}^{-1}(\mathbf{v}_h(\mathbf{x};\theta, F_{\mathbf{X}}); G^0_{\mathbf{B}}))$. To prove sufficiency, suppose $G^0_{\mathbf{B}}$ is symmetric and affiliated with support $[b^0_L, b^0_U]^n$, $\xi(.; G^0_{\mathbf{B}})$ is increasing on the support, and there exists $(\theta, F_{\mathbf{X}}) \in \Theta \otimes \mathcal{F}$ that satisfies (.1). We need to show $G^0_{\mathbf{B}}(\mathbf{b}) = F_{\mathbf{X}}(\mathbf{b}_{\mathbf{0}}^{-1}(\mathbf{b};\theta, F_{\mathbf{X}}))$ $\forall \mathbf{b} \in [b^0_L, b^0_U]^n$, where $b_0(x;\theta, F_{\mathbf{X}})$ is the symmetric, increasing equilibrium strategy characterized by (1.1).[1] By supposition of (.1), $F_{\mathbf{X}}(\mathbf{x}) = G^0_{\mathbf{B}}(\boldsymbol{\xi}^{-1}(\mathbf{v}_h(\mathbf{x};\theta, F_{\mathbf{X}}); G^0_{\mathbf{B}}))$ for all x on support, where the support of $F_{\mathbf{X}}$ is on $[x_L, x_U]^N$, with $x_k = v_h^{-1}(\xi(b_k; G^0_{\mathbf{B}}); \theta, F_{\mathbf{X}})$ for $k = L, U$. Hence the monotonicity of v_h and ξ implies $G^0_{\mathbf{B}}(\mathbf{b}) = F_{\mathbf{X}}(\mathbf{v}_h^{-1}(\boldsymbol{\xi}(\mathbf{b}; G^0_{\mathbf{B}}); \theta, F_{\mathbf{X}}))$ for all $\mathbf{b} \in [b_L, b_U]^n$. Therefore it suffices to show that $\xi^{-1}(v_h(.;\theta; F_{\mathbf{X}}); G^0_{\mathbf{B}})$ satisfies the characterization of equilibrium strategies (i.e., the differential equation (1.1) with the boundary condition $\xi^{-1}(v_h(x_L;\theta; F_{\mathbf{X}}); G^0_{\mathbf{B}}) = v_h(x_L;\theta, F_{\mathbf{X}})$). Note definition of the support of $F_{\mathbf{X}}$ in (.1) implies $v_h(x_L;\theta, F_{\mathbf{X}}) = \xi(b_L; G^0_{\mathbf{B}})$, and $\lim_{b \to b_L} \xi(b; G^0_{\mathbf{B}}) = b_L$. A similar argument to Li et.al (2002) completes the proof. *Q.E.D.*

Suppose $\boldsymbol{\theta}(\mathbf{x}) = (\{\tilde{\theta}(x_i, y_i)\}_{i=1}^n)$ $\forall \mathbf{x} \in [x_L, x_U]^n$, where $y_i \equiv \max_{j \neq i} x_j$. That is, bidders' valuations only depend on his own signal and the highest rival signal, and it is a strictly interdependent value auction provided $\tilde{\theta}$ is not degenerate in the second argument. Then $\mathbf{v}_h(\mathbf{x};\theta, F_{\mathbf{X}}) = (\{\tilde{\theta}(x_i, y_i)\}_{i=1}^n)$. Therefore, a distribution $G^0_{\mathbf{B}}$ that satisfies the necessary conditions is rationalized by any $\theta \in \Theta$ that satisfies "$\max_{j \neq i} X_j$-sufficiency" with the

[1]Existence and uniqueness of symmetric, increasing *PSBNE* is not an issue because the definition of Θ and $\mathcal{F}$ guarantees they exist for all $\{\theta, F_{\mathbf{X}}\} \in \Theta \otimes \mathcal{F}$.

boundary conditions: $\tilde{\theta}(x_k, x_k) = \xi(b_k)$, for $k \in \{L, U\}$, and a signal distribution defined as $F_{\mathbf{X}}(\mathbf{x}) = G^0_{\mathbf{B}}(\xi^{-1}(\tilde{\theta}(x_1, x_1); G^0_{\mathbf{B}}), .., \xi^{-1}(\tilde{\theta}(x_n, x_n); G^0_{\mathbf{B}}))$. □

Proof of Lemma 1. The proof uses the monotonicity and differentiability of $b_0(.)$. By change of variables, $\frac{f_{Y|X}(x|x)}{F_{Y|X}(x|x)} = b_0'(x)\tilde{\Lambda}(b_0(x); G^0_{\mathbf{B}})$ and for all $s \leq x$, $L(s|x; F_{\mathbf{X}}) = \tilde{L}(b_0(s)|b_0(x); G^0_{\mathbf{B}})$. Furthermore, in equilibria $v_h(x, x; \theta, F_{\mathbf{X}}) = \xi(b_0(x); G^0_{\mathbf{B}})$ for all $x \in [x_L, x_U]$. By definition, for $x \geq x^*(r)$,

$$\begin{aligned} b_r(x) &= rL(x^*(r)|x; F_{\mathbf{X}}) + \int_{x^*(r)}^{x} v_h(s, s; \theta, F_{\mathbf{X}})L(s|x; F_{\mathbf{X}})\Lambda(x; F_{\mathbf{X}})dx \\ &= r\tilde{L}(b_0(x^*(r))|b_0(x); G^0_{\mathbf{B}}) + \int_{x^*(r)}^{x} \xi(b_0(s); G^0_{\mathbf{B}})\tilde{L}(b_0(s)|b_0(x); G^0_{\mathbf{B}})\tilde{\Lambda}(b_0(s); G^0_{\mathbf{B}})b_0'(s)ds \\ &= \delta_r(b_0(x); G^0_{\mathbf{B}}) \end{aligned}$$

where the last equality follows from change of variables in the integrand. □

Proof of Lemma 2. *Proof of (i) :* By definition all structures in $\Theta \otimes \mathcal{F}$ satisfy *A1* and *A2*. The affiliation of signals and monotonicity of θ implies that $v_h(x, y)$ is increasing in x and non-decreasing in y. For all $x \geq y$,

$$v_h(x, y) \geq \int_{x_L}^{y} v_h(x, s)\frac{f_{Y|X}(s|x)}{F_{Y|X}(y|x)}ds \equiv v(x, y) \geq \int_{x_L}^{y} v_h(s, s)\frac{f_{Y|X}(s|x)}{F_{Y|X}(y|x)}ds \equiv v_l(x, y)$$

Therefore $v_h(x_L) = v(x_L) = v_l(x_L)$ and $v_h(x) \geq v(x) \geq v_l(x)$ $\forall x \in [x_L, x_U]$. The proof of monotonicity of $v_h(x, x)$ in x is standard and not repeated here. For any $x < x'$ on support, the law of total probability implies

$$\begin{aligned} v_l(x') &= E(v_h(Y)|X_i = x', Y_i \leq x') \\ &= E(v_h(Y)|X_i = x', Y_i \leq x)P(Y_i \leq x|X_i = x', Y_i \leq x') + ... \\ & \quad E(v_h(Y)|X_i = x', x < Y_i \leq x')P(x < Y_i \leq x'|X_i = x', Y_i \leq x') \end{aligned}$$

By monotonicity of v_h and $x' > x$, $E(v_h(Y)|X_i = x', x < Y_i \leq x') > v_l(x)$. By affiliation of X and Y, $E(v_h(Y)|X_i = x', Y_i \leq x) \geq v_l(x)$. Therefore $v_l(x') \geq v_l(x)$.

Proof of (ii) : follows immediately from proof of *(i)*. □

Proof of Lemma 3. For the first claim, note $v_h(x, x; \psi) = v(x, x; \psi)$ $\forall \psi \in \Theta_P \otimes \mathcal{F}$. Hence $x_l(r; \psi) = x^*(r; \psi)$ $\forall \psi \in \Theta_P \otimes \mathcal{F}$. For the second claim, consider $\Theta_S \equiv \{\theta \in \Theta : \theta(x_i, \mathbf{x}_{-i}) = ax_i + \hat{\theta}(\mathbf{x}_{-i})$ *for some* $a > 0$ *and* $\tilde{\theta}$ *exchangeable and non-decreasing in* $\mathbf{x}_{-i}\}$, and $\mathcal{F}_I \equiv \{F_{\mathbf{X}} \in \mathcal{F} : F_{\mathbf{X}}(\mathbf{x}) = \Pi_{i=1}^n F_X(x_i)$ *for some* $F_X\}$. Then $\Theta_S \otimes \mathcal{F}_I$ is a non-empty subset of $\Theta \otimes \mathcal{F}$.[2] By definition, $v(x, x; \psi) - v_l(x, x; \psi) = \int_{x_L}^{x} (v_h(x, s) - v_h(s, s)) \frac{f_{Y|X}(s|x)}{F_{Y|X}(y|x)} ds = \int_{x_L}^{x} a(x - s) \frac{f_{Y|X}(s|x)}{F_{Y|X}(y|x)} ds \leq a(x_U - x_L)$ for all $x \in [x_L, x_U]$ and $\psi \in \Theta_S \otimes \mathcal{F}_I$. Hence $\forall \varepsilon > 0$, $\sup_x |v(x, x; \psi) - v_l(x, x; \psi)| \leq \varepsilon$ $\forall a \in (0, \dfrac{\varepsilon}{x_U - x_L})$. That is, $v_l(x, x)$ converges to $v(x, x)$ uniformly over x as the weight on a bidder's own signal a approaches zero. The rest of the proof shows this uniform convergence of v_l to v implies the uniform convergence of $x_h(r)$ to $x^*(r)$ for r in the nontrivial range as $a \downarrow 0$.

Claim: Suppose $\theta(\mathbf{x}) = ax_i + b\sum_{j \neq i} x_j$ *and private signals are i.i.d. with marginal distribution* $F \in C^1[x_L, x_U]$ *such that* $f_l \leq F'(x) \leq f_u$ $\forall x \in [x_L, x_U]$. *Then* $\forall \varepsilon > 0$ *and* $y \geq x$, $v(y, y) - v(x, x) \leq \varepsilon$ *implies* $y - x \leq \varepsilon / k$, *where* $k \equiv a + (n-1)b\frac{f_l^2}{f_u^2}$.

[2] $\mathcal{F}_I \subset \mathcal{F}$ because independence is a special case of affiliation.

Proof: Denote $\varphi(x) \equiv v(x,x)$. Then $\varphi'(x) = a+(n-1)bf_X(x)[\int_{x_L}^{x}(x-\tilde{x})f_X(\tilde{x})d\tilde{x}]/F_X(x)^2$. The bounds on f_X imply $\varphi'(x) \in [a+(n-1)b\frac{f_l^2}{f_u^2}, a+(n-1)b\frac{f_u^2}{f_l^2}]$. Now suppose $\exists\varepsilon > 0$ and $y \geq x$ on $[x_L, x_U]$ such that $v(y,y) - v(x,x) \leq \varepsilon$ but $y - x > \varepsilon/k$. Then by the Mean Value Theorem, $\exists\tilde{x}_k$ between $x_{\bar{k}}$ and $y_{\bar{k}}$ such that $\varphi'(\tilde{x}) < k$. This contradicts the lower bound of $\varphi'(x)$ on $[x_L, x_U]$. *Q.E.D.*

For all $\varepsilon > 0$, we can pick $a < \dfrac{k\varepsilon}{x_U - x_L}$ and define a structure $\psi = \{a, F\}$ with any $F \in C^1[x_L, x_U]$. Then it follows from the claim that for all $r \geq 0$,

$$\begin{aligned} k\varepsilon &\geq \sup_{x\in[x_L,x_U]} v(x,x;\psi_{a,F}) - v_l(x,x;\psi_{a,F}) \\ &\geq v(x_h(r),x_h(r);\psi_{a,F}) - v_l(x_h(r),x_h(r);\psi_{a,F}) \\ &= v(x_h(r),x_h(r);\psi_{a,F}) - v(x^*(r),x^*(r);\psi_{a,F}). \end{aligned}$$

Hence by the claim above, $x_h(r) - x^*(r) \leq \varepsilon$ for all $r \geq 0$. □

Proof of Lemma 4. For all $x \geq x_l(r)$, that $v_h(x,x) \geq r\ \forall x \in [x_l(r), x_U]$ suggests

$$b_r(x) \leq rL(x_l(r)|x;F_{\mathbf{X}}) + \int_{x_l(r)}^{x} v_h(s,s)dL(s|x;F_{\mathbf{X}})$$

and for all $x \geq x_h(r)$,

$$b_r(x) \geq rL(x_h(r)|x;F_{\mathbf{X}}) + \int_{x_h(r)}^{x} v_h(s,s)dL(s|x;F_{\mathbf{X}})$$

By non-negativity of θ, $x^*(0) = x_L$ and $x_h(r) \geq x^*(r) \geq x_l(r) \geq x^*(0)$ for all $r \geq 0$. Hence equation (1.2) holds for $x_l(r)$ and $x_h(r)$. Substitution and change of variable shows for all

$x \geq x_l(r)$,

$$b_r(x) \leq \delta_{r,l}(b_0(x);G^0_{\mathbf{B}}) \equiv r\tilde{L}(b_0(x_l(r))|b_0(x)) + \int_{b_0(x_l(r))}^{b_0(x)} \xi(\tilde{b})d\tilde{L}(\tilde{b}|b_0(x))$$

and for all $x \geq x_h(r)$,

$$b_r(x) \geq \delta_{r,h}(b_0(x);G^0_{\mathbf{B}}) \equiv r\tilde{L}(b_0(x_h(r))|b_0(x)) + \int_{b_0(x_h(r))}^{b_0(x)} \xi(\tilde{b})d\tilde{L}(\tilde{b}|b_0(x))$$

For all $b \geq b_0(x_k(r))$ and $k \in \{l,h\}$,

$$\delta'_{r,k}(b;G^0_{\mathbf{B}}) = \tilde{\Lambda}(b)\left[\xi(b) - \left(r\tilde{L}(b_0(x_k(r)|b) + \int_{b_0(x_k(r))}^{b} \xi(\tilde{b})d\tilde{L}(\tilde{b}|b)\right)\right] > 0$$

Since $b'_0(x) \geq 0 \ \forall x > x_L$, this implies $\delta_{r,k}(b_0(.);G^0_{\mathbf{B}})$ is increasing for $x \geq x_k(r)$. □

Proof of Proposition 2. It has been shown above that $b_0(x^*(r)) \in [b_0(x_l(r)), b_0(x_h(r))]$. By construction, $\delta_{r,k}(b_0(x_k(r));G^0_{\mathbf{B}}) = r = b_r(x^*(r))$. Hence both $\{\delta_{r,k}(b_0(.);G^0_{\mathbf{B}})\}_{k\in\{l,h\}}$ are invertible at $t \geq r$ over the interval $[x_k(r), x_U]$ for $k \in \{l,h\}$. It follows from the lemma above that $\delta^{-1}_{r,l}(t;G^0_{\mathbf{B}}) \leq b_0(b^{-1}_r(t)) \leq \delta^{-1}_{r,h}(t;G^0_{\mathbf{B}})$ for $t \geq r$. The rest of the proof follows immediately. □

Proof of Lemma 5. To prove *(i)*, note in equilibria $b_r(x^*(r)) = r = v(x^*(r)) \equiv E(V_i|X_i = x^*(r), Y_i \leq x^*(r)) \geq b_0(x^*(r))$, where the last inequality holds by equilibrium bidding conditions with no reserve prices. Besides $\forall x \geq x^*(r)$, $b_r(x) < b_0(x)$ implies $b'_r(x) > b'_0(x)$. It follows from Lemma 2 in Milgrom and Weber (1982) that $b_r(x) \geq b_0(x)$ for all $x \geq x^*(r)$. For *(ii)*, it suffices to note $sgn(b'_r(x) - b'_0(x)) = -sgn(b_r(x) - b_0(x)) \ \forall x \geq x^*(r)$. □

Proof of Proposition 3. By definition of v_0, $\Pr\{R^{II}(r) < v_0\} = 0$. Note $\Pr\{R^{II}(r) = v_0\} = \Pr\{X^{(1)} < x^*(r)\}$, $\Pr\{v_0 < R^{II}(r) < r\} = 0$ and $\Pr\{R^{II}(r) = r\} = \Pr\{X^{(1)} \geq$

$x^*(r) \wedge X^{(2)} < x^*(r)\}$. Because $\beta_r'(x) > 0 \; \forall x \in [x^*(r), x_U]$ and $\beta_r(x^*(r)) = v_h(x^*(r)) \geq r$, it follows $\Pr\{r < R^{II}(r) < v_h(x^*(r))\} = 0$. Hence:

$$\begin{aligned} F_{R^{II}(r)}(t) &= 0 \quad \forall t < v_0 \\ &= \Pr\{X^{(1)} < x^*(r)\} \quad \forall t \in [V_0, r) \\ &= \Pr\{X^{(2)} < x^*(r)\} \quad \forall t \in [r, v_h(x^*(r))) \end{aligned}$$

Next note $\forall t \in [v_h(x^*(r)), +\infty)$, $\Pr\{R^{II}(r) \in [v_h(x^*(r)), t]\} = \Pr\{v_h(X^{(2)}) \in [v_h(x^*(r)), t]\}$. Hence for all t in this range,

$$\begin{aligned} \Pr\{R^{II}(r) \leq t\} = & \quad \Pr\{R^{II}(r) < v_h(x^*(r))\} + \Pr\{R^{II}(r) \in [v_h(x^*(r)),\ t]\} \\ = & \quad \Pr\{X^{(2)} < x^*(r)\} + \Pr\{v_h(X^{(2)}) \in [v_h(x^*(r)),\ t]\} \\ = & \quad \Pr\{v_h(X^{(2)}) \leq t\} \end{aligned}$$

This completely characterizes the counterfactual distribution of $R^{II}(r)$.

For $t < r$, $F^l_{R^{II}(r)}(t) = \Pr\{b_0(X^{(1)}) < b_0(x_l(r))\} \leq \Pr\{X^{(1)} < x^*(r)\} = F_{R^{II}(r)}(t) \leq \Pr\{b_0(X^{(1)}) < b_0(x_h(r))\} = F^u_{R^{II}(r)}(t)$. For $t \in [r, v(x^*(r)))$,

$$\begin{aligned} F^l_{R^{II}(r)}(t) &= \Pr\{v_h(X^{(2)}) \leq t\} \leq \Pr\{v_h(X^{(2)}) < v_h(x^*(r))\} \\ &= \Pr\{X^{(2)} < x^*(r)\} = F_{R^{II}(r)}(t) \leq \Pr\{b_0(X^{(2)}) < b_0(x_h(r))\} = F^u_{R^{II}(r)}(t) \end{aligned}$$

due to the monotonicity of $b_0(.)$. For $t \in [v(x^*(r)), v_h(x_h(r)))$,

$$\begin{aligned} F^l_{R^{II}(r)}(t) &= F_{R^{II}(r)}(t) = \Pr\{v_h(X^{(2)}) \leq t\} \leq \Pr\{v_h(X^{(2)}) < v_h(x_h(r))\} \\ &= \Pr\{b_0(X^{(2)}) < b_0(x_h(r))\} = F^u_{R^{II}(r)}(t) \end{aligned}$$

due to the monotonicity of $b_0(.)$ and $v_h(.)$. For $t \in [v_h(x_h(r)), +\infty)$,

$$F^l_{R^{II}(r)}(t) = F_{R^{II}(r)}(t) = F^u_{R^{II}(r)}(t) = \Pr\{v_h(X^{(2)}) \le t\}$$

and point identification of revenue distribution is achieved on this range. □

Proof of Proposition 5. Auction characteristics are common knowledge among all bidders. Hence the symmetric equilibrium satisfies the following first-order condition: (By symmetry among the bidders, bidder indices are dropped for notational ease.)

$$\tfrac{\partial}{\partial X} b(x,z;n) = [\tilde{v}_h(x,z;n) - b(x,z;n)]\tfrac{f_{Y|X,Z;N}(x|x,z;n)}{F_{Y|X,Z;N}(x|x,z;n)}$$

where $\tilde{v}_h(x,z;n) \equiv E(V_i|X_i = Y_i = x, Z = z; N = n)$, $Y_i \equiv \max_{j\neq i} X_i$, $F_{Y|X,Z;N}(t|x,z;n) \equiv \Pr(\max_{j\neq i} X_j \le t|X_i = x, Z = z; N = n)$ and $f_{Y|X,Z;N}(t|x,z;n)$ is the corresponding conditional density. The equilibrium boundary condition for all (z,n) is $b(x_L,z;n) = \tilde{v}_h(x_L,z;n)$. For every z and n, the differential equation is known to have the following closed form solution :

$$b(x,z;n) = \int_{x_L}^{x} h(z'\gamma) + \phi(s;n)dL(s|x;n)$$

Independence of X_i and Z conditional on N implies both $\phi(x;n)$ and $L(s|x;n)$ are invariant to z for all s and x. Hence under assumption *A1',A2 and A4*, $b(x_L,z;n) = \tilde{v}_h(x_L,z;n) = h(z'\gamma) + \phi(x_L;n)$. For $x > x_L$, $b(x,z;n) = h(z'\gamma) + \int_{x_L}^{x} \phi(s;n)dL(s|x;n)$ for all (x,z,n). □

Proof of Proposition 6. Differentiating $b_r(x)$ for $x \ge x^*(r)$ gives

$$b_r'(x;\theta,F_{\mathbf{X}})/\Lambda(x;F_{\mathbf{X}}) + b_r(x;\theta,F_{\mathbf{X}}) = v_h(x;\theta,F_{\mathbf{X}}) \tag{.2}$$

For all $r \geq 0$ and $x, y \geq x^*(r)$,

$$
\begin{aligned}
(.3)\ F_{Y|X}(y|x) &\equiv \Pr(Y \leq y|x = x) \\
&= \Pr(Y < x^*(r)|X = x) + \Pr(x^*(r) \leq Y \leq y|X = x) \\
&= \Pr(b_r(Y) < r|b_r(X) = b_r(x)) + \Pr(r \leq b_r(Y) \leq b_r(y)|b_r(X) = b_r(x)) \\
&\equiv G^r_{M|B}(b_r(y)|b_r(x))
\end{aligned}
$$

The equality of the two terms follows *respectively* from the facts that $Y < x^*(r)$ if and only if $b_r(Y) < r$ and $b_r(x)$ is increasing for $x \geq x^*(r)$. Taking derivative of both sides w.r.t. y for $y \geq x^*(r)$ gives

$$
f_{Y|X}(y|x) = b'_r(y) g^r_{M|B}(b_r(y)|b_r(x)) \tag{.4}
$$

for all $x, y \geq x^*(r)$. Substitute (.4) and (.3) into (.2) proves the lemma. □

Proof of Proposition 7. Let $X^{(i:n)}$ denote the ith largest signal among n potential bidders. Then $\Pr(X^{(2:n)} < x^*(r)|X^{(1:n)} \geq x^*(r))$ is observed. By the i.i.d. assumption, $\Pr(X^{(2:n)} < x^*(r)|X^{(1:n)} \geq x^*(r)) = \dfrac{nF_r^{n-1}(1-F_r)}{1-F_r^n}$, where $F_r \equiv \Pr(X_i \leq x^*(r))$. The expression is increasing in F_r. Therefore F_r is identified, and $\Pr(X^{(1)} < x^*(r)) = F_r^n$. □

1.2. Proof of the consistency of $\{\hat{F}^k_{R^I(r)}\}_{k=l,u}$

The lemma below extends the Basic Consistency Theorem of extreme estimators to those defined over random compact sets rather than fixed compact sets. It will be applied repeatedly in our proof of consistency of the three-step estimators.

Lemma B1 *Let $Q(.)$ and $\hat{Q}_N(.)$ be nonstochastic and stochastic real-valued functions defined respectively on compact intervals $\Theta \equiv [\theta^l, \theta^u]$ and $\Theta_N \equiv [\theta^l_N, \theta^u_N]$, where $\Pr\{[\theta^l_N, \theta^u_N] \subseteq [\theta^l, \theta^u]\} = 1$ for all N and $\theta^k_N \to \theta^k$ almost surely for $k = l, u$. For every $N = 1, 2, ...,$ let $\hat{\theta}_N \in \Theta_N$ be such that $\hat{Q}_N(\hat{\theta}_N) \leq \inf_{\theta \in \Theta_N} \hat{Q}_N(\theta) + o_p(1)$. If $Q(.)$ is continuous on Θ with a unique maximizer on Θ at $\theta_0 \in [\theta^l, \theta^u]$ and (ii) $\sup_{\theta \in \Theta_N} \left| \hat{Q}_N(\theta) - Q(\theta) \right| \xrightarrow{p} 0$ as $N \longrightarrow +\infty$, then $\hat{\theta}_N \xrightarrow{p} \theta_0$.*

Proof. In the case $\theta_0 \in (\theta^l, \theta^u)$, the proof is an adaptation from that of Theorem 4.1.1 in Amemiya (1985) and is included in *Lemma A2* of Li et.al (2003). In the case $\theta_0 = \theta^k$ for $k = l, u$, the continuity of $Q(\theta)$ at $\theta = \theta^k$ is sufficient for $\lim_{N \to +\infty} \Pr(\theta_N > \theta^k + \varepsilon) = 0$ for all $\varepsilon > 0$. The proof is standard and omitted. □

1.2.1. Regularity properties of $G_{M,B}$ and $g_{M,B}$. Let $f_{Y,X}$ and $F_{Y,X}$ denote the joint density and distribution of Y_i and X_i respectively. Let $\beta(.)$ be the bidding strategy under increasing, pure-strategy perfect Bayesian Nash equilibria. That is, $\beta(x) = \int_{x_L}^{x} v_h(s) dL(s|x)$ where $L(s|x) = \exp\{-\int_s^x \frac{f_{YX}(u,u)}{F_{YX}(u,u)} du\}$. The lemma below gives regularity results about the smoothness of the equilibrium bidding strategy.

Lemma B2 *Under S1 and S2, the equilibrium bidding function $\beta(.)$ admits up to R continuous bounded derivatives on $[x_L, x_U]$, and $\beta'(.)$ is bounded below from zero on $[x_L, x_U]$.*

Proof. The proof is similar to Li et.al (2002) and omitted. □

This leads to following results of further regularity conditions of the joint density of equilibrium bids B and highest rival bid M (denoted $g_{M,B}$) and $G_{M,B}(m, b) \equiv \int_{b_L}^{m} g_{M,B}(\tilde{b}, b) d\tilde{b}$. The relevant support is $[b_L, b_U]^2$ where $b_L = \beta(x_L) = v_h(x_L) = \theta(\mathbf{x}_L)$ and $b_U = \beta(x_U)$.

Proposition B1 *Under S1 and S2, (i) ξ has R continuous bounded derivatives on $[b_L, b_U]$ and $\xi'(.) \geq c > 0$ for some constant on $[b_L, b_U]$; (ii) $G_{M,B}$ and $g_{M,B}$ both have $R-1$ continuous bounded partial derivatives on $[b_L, b_U]^2$.*

Proof. By definition in equilibrium, $\beta(v_h^{-1}(\xi(b))) = b$. Hence $\xi'(b) = \{\beta'[v_h^{-1}(\xi(b))]v_h^{-1\prime}(\xi(b))\}^{-1}$ where both $v_h^{-1\prime}(.)$ and $\beta'(.)$ are bounded below from zero and have $R-1$ continuous derivatives under *S2* and *Lemma B2*. For Part(ii), note $\Pr(M \leq m, B \leq b) = \Pr(Y \leq v_h^{-1}(\xi(m)), X \leq v_h^{-1}(\xi(b)))$ by the monotonicity of $\beta(.)$. Hence $G_{M,B}(m,b) = \frac{\partial}{\partial B}\Pr(M \leq m, B \leq b) = v_h^{-1\prime}(\xi(b))\xi'(b)\Pr[Y \leq v_h^{-1}(\xi(m)),\ X = v_h^{-1}(\xi(b))]$, where the third term has $R+n-1$ continuous derivatives and the first two terms have $R-1$ continuous bounded derivatives on $[b_L, b_U]^2$. And $g_{M,B}(m,b) = \frac{\partial}{\partial M}G_{M,B}(m,b) = v_h^{-1\prime}(\xi(b))\xi'(b)\, v_h^{-1\prime}(\xi(m))\xi'(m)\ f_{Y,X}[v_h^{-1}(\xi(m)), v_h^{-1}(\xi(b))]$, where the last term has $R+n-2$ continuous derivatives on $[b_L, b_U]^2$ and $v_h^{-1\prime}(.)$ has $R-1$ continuous derivatives (see proof of *Lemma B2* below). Hence $g_{M,B}(m,b)$ has $R-1$ continuous derivatives. □

1.2.2. Consistency of $\hat{b}^0_{l,r}$ and $\hat{b}^0_{h,r}$. The following lemma give the rate of uniform convergence of kernel estimates $\hat{G}_{M,B}$ and $\hat{g}_{M,B}$ to $G_{M,B}$ and $g_{M,B}$ over $C^2_\delta(B)$, and $\tilde{G}_{M,B}$ to $G_{M,B}$ over $\hat{C}^2_\delta(B)$. It lays a foundation for our proof of uniform convergence of $\hat{\xi}_l$ and $\hat{\xi}$ as well as $\hat{\delta}_{l,r}$ and $\hat{\delta}_{h,r}$.

Lemma B3 *Let $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ and $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$. Under S1, S2, and S3,*

$$\sup_{C^2_\delta(B)} |\hat{G}_{M,B} - G_{M,B}| = O(h_G^{R-1}), \qquad \sup_{C^2_\delta(B)} |\hat{g}_{M,B} - g_{M,B}| = O(h_g^{R-1})$$

$$\sup_{b \in \hat{C}_\delta(B)} |\tilde{G}_{M,B}(b,b) - G_{M,B}(b,b)| = O_p(h_g^{R-1})$$

Furthermore, if $R > n$,

$$\sup_{\tilde{b}_L \leq b, b \in C_\delta(B)} |\frac{\hat{G}_{M,B}(\tilde{b}_L, b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L, b)}{G_{M,B}(b,b)}| = O_p(h_g^{R-n})$$

Proof. See the last section of Appendix B. □

The next lemma proves the uniform convergence of $\hat{\xi}$ and $\hat{\xi}_l$ over the relevant expanding supports.

Lemma B4 *Let* $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ *and* $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$. *Under S1, S2, and S3,* $\sup_{b \in C_\delta(B)} |\hat{\xi}(b) - \xi(b)| = O_p(h_g^{R-(n-1)})$ *if* $R > n-1$. *Furthermore* $\sup_{b \geq \tilde{b}_L, b \in C_\delta(B)} |\hat{\xi}_l(b) - \xi_l(b)| = O_p(h_g^{R-2(n-1)})$ *if* $R > 2n-2$.

Proof. See the last section of Appendix B. □

The proposition below establishes the consistency of $\hat{b}^0_{k,r}$ using the extended version of the Basic Consistency Theorem. The proof proceeds by verifying assumptions in *Lemma B1*. Note the range of r for nontrivial analyses is the interval $S_r \equiv [v(x_L), v(x_U)]$.

Proposition B2 *Let* $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ *and* $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$. *Under S1, S2, and S3,* $\hat{b}^0_{l,r} \xrightarrow{p} b^0_{l,r}$ *if* $R > n-1$ *and* $\hat{b}^0_{h,r} \xrightarrow{p} b^0_{h,r}$ *if* $R > 2(n-1)$ *for all* $r \in S_r$.

Proof. It suffices to show that for all $r \in S_r$, *(i)* $(\hat{\xi}(b) - r)^2$ and $(\hat{\xi}_l(b) - r)^2$ converge in probability to $(\xi(b) - r)^2$ and $(\xi_l(b) - r)^2$ uniformly over $\hat{C}_\delta(B)$; *(ii)* $(\xi(b) - r)^2$ and $(\xi_l(b) - r)^2$ are continuous on $[b^0_L, b^0_U]$ with unique minimizers $b^0_{l,r}$ and $b^0_{h,r}$ respectively on $[b^0_L, b^0_U]$; and *(iii)* $\tilde{b}_k \to b^0_k$ almost surely for $k = L, U$. First, by *Lemma B4*, $\sup_{b \in C_\delta(B)} |\hat{\xi}(b) - \xi(b)| \xrightarrow{p} 0$ and $\sup_{b \geq \tilde{b}_L, (\tilde{b}_L, b) \in C^2_\delta(B)} |\hat{\xi}_l(b) - \xi_l(b)| \xrightarrow{p} 0$. And $\sup_{b \in C_\delta(B)} \left|(\hat{\xi}(b) - r)^2 - (\xi(b) - r)^2\right| \leq \sup_{b \in C_\delta(B)} \left|\hat{\xi}^2(b) - \xi^2(b)\right| + 2r \sup_{b \in C_\delta(B)} \left|\hat{\xi}(b) - \xi(b)\right|$, where both terms converge to 0 in probability since $\sup_{b \in C_\delta(B)} \xi(b) \leq \xi(b_U) = v_h(x_U, x_U) < \infty$.

Likewise $\sup_{b\geq\tilde{b}_L,(\tilde{b}_L,b)\in C^2_\delta(B)}\left|(\hat{\xi}_l(b)-r)^2-(\xi_l(b)-r)^2\right|\overset{p}{\longrightarrow}0$ by similar arguments. Next, the continuity of $(\xi(b)-r)^2$ and $(\xi_l(b)-r)^2$ follows from the smoothness of ξ shown above. Also both ξ and ξ_l are increasing on $[b^0_L, b^0_U]$ by the monotonicity of $v_h(.)$ and $v_l(.)$ as well as $\beta(.)$ on $[x_L, x_U]$. Thus for all $r\in[b^0_L, v(x_U,x_U)]$, the minimizers of $(\xi(b)-r)^2$ and $(\xi_l(b)-r)^2$ are unique on $[b^0_L, b^0_U]$. Finally, that $\tilde{b}_k\to b^0_k$ almost surely for $k=L,U$ follows from $\delta\longrightarrow 0$ and $\hat{b}_L\overset{a.s.}{\longrightarrow}b^0_L$. □

1.2.3. Uniform convergence of $\hat{\delta}_{k,r}(.;\hat{b}^0_{k,r})$. Recall for $k=l,h$ and $r\in S_r$, $\delta_{r,k}(.;b^0_{k,r})$ are defined as:

$$\begin{aligned}\delta_{r,k}(b;b^0_{k,r}) &\equiv rL(b^0_{k,r}|b)+\int_{b^0_{k,r}}^{b}\xi(t)\Lambda(t)L(t|b)dt \quad \forall b\in(b^0_{k,r},b_U]\\ &\equiv r \quad \forall b\in[b_L,b^0_{k,r}]\end{aligned}$$

where $b^0_{h,r}=\inf\{b\in C(B):\xi_l(b)\geq r\}$ and $b^0_{l,r}=\inf\{b\in C(B):\xi(b)\geq r\}$, and $L(t|b)=\exp\left(-\int_t^b\frac{g(u,u)}{G(u,u)}du\right)$ for $b_L\leq t\leq b\leq b_U$.

Lemma B5 *Let $h_G=c_G(\log L/L)^{1/(2R+2n-5)}$ and $h_g=c_g(\log L/L)^{1/(2R+2n-4)}$. Under S1, S2, and S3 and if $R>2n-1$,*

$$\sup_{b\in C_\delta(B)}\left|\frac{\hat{g}_{M,B}(b,b)}{\hat{G}_{M,B}(b,b)}-\frac{g_{M,B}(b,b)}{G_{M,B}(b,b)}\right|=O_p(h^{R-2n+1})$$

Proof. See the last section of Appendix B. □

The following lemma shows the uniform convergence of $\hat{\delta}_{k,r}(.;\hat{b}^0_{k,r})$ over $\hat{C}_\delta(B)=[\hat{b}_L+\delta,\hat{b}_U-\delta]$ for the relevant range of r. By construction $\hat{b}^0_{k,r}\in\hat{C}_\delta(B)\subseteq C_\delta(B)$ for $r\in S_r$.

Lemma B6 *Let* $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ *and* $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$*. Under S1, S2, and S3 and suppose* $R > 2n-1$*, then* $\sup_{b\in C_\delta(B)}\left|\hat{\delta}_{k,r}(b;\hat{b}^0_{k,r}) - \delta_{k,r}(b;b^0_{k,r})\right| \xrightarrow{p} 0$ *for all* $r \in S_r$*.*

Proof. First consider the case $r \in interior(S_r)$ (or $b^0_{k,r} \in (b^0_L, b^0_U)$). By definition, $\hat{b}^0_{k,r} \in C_\delta(B)$, and for L large enough, $b^0_{k,r}$ is in the interior of $C_\delta(B)$. By triangular inequality,

$$
\begin{aligned}
& \sup_{b\in C_\delta(B)}\left|\hat{\delta}_{k,r}(b;\hat{b}^0_{k,r}) - \delta_{k,r}(b;b^0_{k,r})\right| \\
\leq \quad & \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})1(b > b^0_{k,r})\left|\hat{\delta}_{k,r}(b) - \delta_{k,r}(b)\right| + ... \\
& \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} > b^0_{k,r})1(b > \hat{b}^0_{k,r})\left|\hat{\delta}_{k,r}(b) - \delta_{k,r}(b)\right| + ... \\
& \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})1(b \in (\hat{b}^0_{k,r}, b^0_{k,r}])\left|\hat{\delta}_{k,r}(b) - r\right| + ... \\
& \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} > b^0_{k,r})1(b \in (b^0_{k,r}, \hat{b}^0_{k,r}])\left|\delta_{k,r}(b) - r\right|
\end{aligned}
$$

It suffices to show all four terms (denoted A_1, A_2, A_3 and A_4 respectively) converge in probability to 0 uniformly over $b \in C_\delta(B)$ as sample size increases. For A_1,

$$
\begin{aligned}
& \sup\nolimits_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})1(b > b^0_{k,r})\left|\hat{\delta}_{k,r}(b;\hat{b}^0_{k,r}) - \delta_{k,r}(b;b^0_{k,r})\right| \\
\leq \quad & \sup\nolimits_{b^0_{k,r}\leq b\leq b_U-\delta} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})\{r\left|\hat{L}(\hat{b}^0_{k,r}|b) - L(b^0_{k,r}|b)\right| + ... \\
& \left|\int_{b^0_{k,r}}^{b} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) - \xi(t)\Lambda(t)L(t|b)dt\right| + \left|\int_{\hat{b}^0_{k,r}}^{b^0_{k,r}} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt\right|\}
\end{aligned}
$$

It can be shown $\sup_{t\leq b,(t,b)\in C_\delta^2(B)}\left|\hat{L}(t|b)-L(t|b)\right|\xrightarrow{p}0$ using convergence results from previous lemmae.[3] Note

$$\begin{aligned}&\sup_{b^0_{k,r}<b\leq b_U-\delta}1(\hat{b}^0_{k,r}\leq b^0_{k,r})r\left|\hat{L}(\hat{b}^0_{k,r}|b)-L(b^0_{k,r}|b)\right|\\ \leq\;&\sup_{b^0_{k,r}<b\leq b_U-\delta}1(\hat{b}^0_{k,r}\leq b^0_{k,r})r\left|\hat{L}(\hat{b}^0_{k,r}|b)-L(\hat{b}^0_{k,r}|b)\right|+...\\ &\sup_{b^0_{k,r}<b\leq b_U-\delta}1(\hat{b}^0_{k,r}\leq b^0_{k,r})r\left|L(\hat{b}^0_{k,r}|b)-L(b^0_{k,r}|b)\right|\end{aligned}$$

For sufficiently small δ, $b^0_{k,r}>b_L+\delta$. Since by construction $\hat{b}^0_{k,r}\in C_\delta(B)$, $\sup_{b^0_{k,r}<b\leq b_U-\delta}1(\hat{b}^0_{k,r}\leq b^0_{k,r})r\left|\hat{L}(\hat{b}^0_{k,r}|b)-L(\hat{b}^0_{k,r}|b)\right|\xrightarrow{p}0$. Also by mean value theorem,

$$\left|L(\hat{b}^0_{k,r}|b)-L(b^0_{k,r}|b)\right|=\left|\frac{\partial}{\partial t}L(t|b)|_{t=\tilde{b}^0_{k,r}}(\hat{b}^0_{k,r}-b^0_{k,r})\right|=\left|\Lambda(\tilde{b}^0_{k,r})L(\tilde{b}^0_{k,r}|b)\right|\left|\hat{b}^0_{k,r}-b^0_{k,r}\right|$$

for some $\tilde{b}^0_{k,r}$ between $\hat{b}^0_{k,r}$ and $b^0_{k,r}$. The consistency of $\hat{b}^0_{k,r}$ suggests $\tilde{b}^0_{k,r}$ is bounded away from b^0_L as sample size increases. Thus both $\Lambda(\tilde{b}^0_{k,r})$ and $L(\tilde{b}^0_{k,r}|b)$ converge in probability to some finite constant since $\sup_{C_\delta(B)}|g|<\infty$ and $\inf_{C_\delta(B)}|g'|>c>0$ and hence $\sup_{b^0_{k,r}\leq b\leq b_U-\delta}\left|L(\hat{b}^0_{k,r}|b)-L(b^0_{k,r}|b)\right|$ is $o_p(1)$. Next

$$\begin{aligned}&\sup_{b^0_{k,r}<b\leq b_U-\delta}\left|\int_{b^0_{k,r}}^{b}\hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)-\xi(t)\Lambda(t)L(t|b)dt\right|\\ \leq\;&\sup_{b^0_{k,r}\leq b\leq b_U-\delta}\left|\hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)-\xi(t)\Lambda(t)L(t|b)\right|\left|b-b^0_{k,r}\right|\xrightarrow{p}0\end{aligned}$$

where the right hand side is $o_p(1)$ by the uniform convergence of $\hat{\xi}$, $\hat{\Lambda}$, and $\hat{L}$ over $C_\delta(B)$ under the assumption of the lemma and the boundedness of ξ, Λ and L on the closed interval

[3]For details of the proof, see Li et.al (2003).

$[b^0_{k,r}, b_U - \delta]$. Finally

$$
\begin{aligned}
&\sup_{b^0_{k,r} < b \le b_U - \delta} 1(\hat{b}^0_{k,r} \le t \le b^0_{k,r}) \left| \int_{\hat{b}^0_{k,r}}^{b^0_{k,r}} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt \right| \\
\le &\sup_{b^0_{k,r} < b \le b_U - \delta, (t,b) \in C^2_\delta} 1(\hat{b}^0_{k,r} \le t \le b^0_{k,r}) \left| \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) \right| \left| \hat{b}^0_{k,r} - b^0_{k,r} \right|
\end{aligned}
$$

Since

$$
\sup_{t \le b, (t,b) \in C^2_\delta(B)} \left| \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) - \xi(t)\Lambda(t)L(t|b) \right| \xrightarrow{p} 0
$$

and $\sup_{b^0_{k,r} < b \le b_U - \delta, (t,b) \in C^2_\delta} 1(\hat{b}^0_{k,r} \le t \le b^0_{k,r}) \left| \xi(t)\Lambda(t)L(t|b) \right|$ is bounded $w.p.a.1$ by the consistency of $\hat{b}^0_{k,r}$, then

$$
\sup_{b^0_{k,r} < b \le b_U - \delta} \left| \int_{\hat{b}^0_{k,r}}^{b^0_{k,r}} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt \right| = o_p(1)
$$

and it follows $A_1 = o_p(1)$. For A_2,

$$
\begin{aligned}
&\sup\nolimits_{b \in C_\delta(B)} 1(\hat{b}^0_{k,r} > b^0_{k,r}) 1(b > \hat{b}^0_{k,r}) \left| \hat{\delta}_{k,r}(b) - \delta_{k,r}(b) \right| \\
\le &\sup_{b^0_{k,r} < b \le b_U - \delta} 1(b > \hat{b}^0_{k,r} > b^0_{k,r}) \{ r \left| \hat{L}(\hat{b}^0_{k,r}|b) - L(b^0_{k,r}|b) \right| + ... \\
&\left| \int_{\hat{b}^0_{k,r}}^{b} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) - \xi(t)\Lambda(t)L(t|b)dt \right| + \left| \int_{b^0_{k,r}}^{\hat{b}^0_{k,r}} \xi(t)\Lambda(t)L(t|b)dt \right| \}
\end{aligned}
$$

where sup of the first and last term over $b^0_{k,r} < b \le b_U - \delta$ are $o_p(1)$ by the same argument as above, and

$$
\begin{aligned}
&\sup\nolimits_{b^0_{k,r} \le b \le b_U - \delta} 1(b > \hat{b}^0_{k,r} > b^0_{k,r}) \left| \int_{\hat{b}^0_{k,r}}^{b} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) - \xi(t)\Lambda(t)L(t|b)dt \right| \\
\le &\sup\nolimits_{b^0_{k,r} \le b \le b_U - \delta, (t,b) \in C^2_\delta} \left| \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b) - \xi(t)\Lambda(t)L(t|b) \right| \left| b - \hat{b}^0_{k,r} \right| = o_p(1)
\end{aligned}
$$

For A_3,

$$\begin{aligned}
& \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})1(b \in (\hat{b}^0_{k,r}, b^0_{k,r}]) \left|\hat{\delta}_{k,r}(b) - r\right| \\
= & \sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r})1(b \in (\hat{b}^0_{k,r}, b^0_{k,r}]) \left| r\hat{L}(\hat{b}^0_{k,r}|b) - r + \int_{\hat{b}^0_{k,r}}^{b} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt \right|
\end{aligned}$$

By construction, $\hat{b}^0_{k,r} \in C_\delta(B)$. The uniform convergence of $\hat{L}(t|b)$ for all $t \leq b$ on $C^2_\delta(b)$, the continuity of $L(t|b)$ in both arguments, and $\hat{b}^0_{k,r} \xrightarrow{p} b^0_{k,r}$ suggest $\sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r} \cap b \in (\hat{b}^0_{k,r}, b^0_{k,r}]) \left| r\hat{L}(\hat{b}^0_{k,r}|b) - r \right| = o_p(1)$. Also for large samples, $\hat{b}^0_{k,r}$ is bounded away from b_L $w.p.a.1$ and $\sup_{b\in C_\delta(B)} 1(\hat{b}^0_{k,r} \leq b^0_{k,r} \cap b \in (\hat{b}^0_{k,r}, b^0_{k,r}]) \left| \int_{\hat{b}^0_{k,r}}^{b} \hat{\xi}(t)\hat{\Lambda}(t)\hat{L}(t|b)dt \right| = o_p(1)$. For A_4, note $\delta_{k,r}$ is continuous at $b^0_{k,r}$ with $\delta_{k,r}(b^0_{k,r}) = r$ and is increasing beyond $b^0_{k,r}$ (as proven in the paper). Hence the consistency of $\hat{b}^0_{k,r}$ is sufficient for $A_4 \xrightarrow{p} 0$. In the boundary case where $b^0_{k,r} = b^0_L$, it suffices to show the convergence of terms $A2$ and $A4$. The same argument above applies. □

1.2.4. Final step of the proof. In this subsection, I apply the extended BCT over random compact sets to prove the consistency of $\hat{\delta}^{-1}_{k,r}(t; \hat{b}^0_{k,r})$ for an interesting range of r and t.

Lemma B7 *Let $h_G = c_G(\log L/L)^{1/(2R+2n-5)}$ and $h_g = c_g(\log L/L)^{1/(2R+2n-4)}$. Under S1, S2, and S3 and suppose $R > 2n-1$, for any $r \in S_r$, $\hat{\delta}^{-1}_{k,r}(t; \hat{b}^0_{k,r}) \xrightarrow{p} \delta^{-1}_{k,r}(t; b^0_{k,r})$ for $k = \{l, h\}$ and all $t > r$.*

Proof. For the range of $r \in S_r$ and $t > r$, $\delta^{-1}_{k,r}(t; b^0_{k,r})$ are unique minimizers of $[\delta_{k,r}(b) - t]^2$ over $C(B)$. *Lemma B6* showed that $\sup_{b\in C_\delta(B)} \left|\hat{\delta}_{k,r}(b; \hat{b}^0_{k,r}) - \delta_{k,r}(b; b^0_{k,r})\right| \xrightarrow{p} 0$ and $\delta_{k,r}(.; b^0_{k,r})$ is also continuous on $C(B)$. Also $\tilde{b}_k \to b^0_k$ almost surely for $k = L, U$ as sample size increases. All conditions for *Lemma B1* are satisfied and claim is proven. □

Lemma B8 *Let $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1(Z_i \leq t)$ where $\{Z_i\}_{i=1}^{n}$ is an i.i.d. sample from a population distributed as F_Z. Then $\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F_Z(t)| \overset{a.s}{\to} 0$. If $F_Z(t_0)$ is continuous at t_0 and a sequence of random variable $\hat{t}_n \overset{p}{\to} t_0$ and t_0 is a continuity point of F, then $\hat{F}_n(\hat{t}_n) \overset{p}{\to} F_Z(t_0)$.*

Proof. The first claim follows from Glivenko-Cantelli Lemma and the proof of the second claim is standard (e.g. see Theorem 4.1.5 Amemiya 1985). □

The proof of *Proposition 4* follows directly from results of the lemmae.

Proof of Proposition 4. By the first part of *Lemma B8*, $\frac{1}{L_n}\sum_{l=1}^{L_n} 1(B_l^{\max} \leq b)$converge in probability to $\Pr(B_l^{\max} \leq b)$ uniformly over $C(B)$. By *Lemma B7*, $\hat{\delta}_{r,k}^{-1}(t) \overset{p}{\to} \delta_{r,k}^{-1}(t)$ for all r and t in the stated range of interests. The second part of *Lemma B8* proves $\hat{F}_{R(r)}^{l}(t) \overset{p}{\to} F_{R(r)}^{l}(t)$ and $\hat{F}_{R(r)}^{u}(t) \overset{p}{\to} F_{R(r)}^{u}(t)$ for given r and t. □

1.2.5. Proofs of lemmae in the consistency proof.

Proof of Lemma B3. That $\sup_{C_\delta^2(B)} \left|\hat{G}_{M,B} - G_{M,B}\right| = O(h_G^{R-1})$ and $\sup_{C_\delta^2(B)} |\hat{g}_{M,B} - g_{M,B}| = O(h_g^{R-1})$ follow from Lemma *A5* in Li et.al (2002). By triangular inequality, for all $b \in \hat{C}_\delta(B)$, $\left|\tilde{G}_{M,B}(b,b) - G_{M,B}(b,b)\right| \leq \int_{\tilde{b}_L}^{b} |\hat{g}_{M,B}(t,b) - g_{M,B}(t,b)|\, dt + \left|\hat{G}_{M,B}(\tilde{b}_L,b) - G_{M,B}(\tilde{b}_L,b)\right|$. Thus

$$\begin{aligned}
&\sup_{b\geq \tilde{b}_L, b\in C_\delta(B)} |\tilde{G}_{M,B}(b,b) - G_{M,B}(b,b)| \\
\leq\ & |b_U - b_L| \sup_{t\leq b,(t,b)\in C_\delta^2(B)} |\hat{g}_{M,B}(t,b) - g_{M,B}(t,b)| + O_p(h_G^{R-1}) = O_p(h_g^{R-1})
\end{aligned}$$

since by construction $\hat{C}_\delta(B) \subseteq C_\delta(B)$ with probability 1 and $h_G < h_g$ for L large enough. Furthermore, note:

$$\begin{aligned}
1\{\tilde{b}_L &\leq b\} \left| \tfrac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \tfrac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right| \\
&\leq \tfrac{\mathbf{1}\{\tilde{b}_L \leq b\}}{\hat{G}_{M,B}(b,b)} \left| \hat{G}_{M,B}(\tilde{b}_L,b) - G_{M,B}(\tilde{b}_L,b) + \tfrac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \left(G_{M,B}(b,b) - \hat{G}_{M,B}(b,b) \right) \right| \\
&\leq \tfrac{\mathbf{1}\{\tilde{b}_L \leq b\}}{\inf_{b \geq \tilde{b}_L} \left| \hat{G}_{M,B}(b,b) \right|} \left(\left| \hat{G}_{M,B}(\tilde{b}_L,b) - G_{M,B}(\tilde{b}_L,b) \right| + \left| G_{M,B}(b,b) - \hat{G}_{M,B}(b,b) \right| \right)
\end{aligned}$$

Thus

$$\begin{aligned}
&\sup_{\tilde{b}_L \leq b, b \in C_\delta(B)} \left| \frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right| \\
\leq\ & \tfrac{\mathbf{1}}{\inf_{b \in C_\delta(B)} \left| \hat{G}_{M,B}(b,b) \right|} \left\{ \begin{array}{c} \sup_{\tilde{b}_L \leq b, b \in C_\delta^2(B)} |\hat{G}_{M,B}(\tilde{b}_L,b) - G_{M,B}(\tilde{b}_L,b)| + ... \\ \sup_{\tilde{b}_L \leq b, b \in C_\delta(B)} |G_{M,B}(b,b) - \hat{G}_{M,B}(b,b)| \end{array} \right\}
\end{aligned}$$

where

$$\begin{aligned}
\inf_{b \in C_\delta(B)} |\hat{G}_{M,B}(b,b)| &\geq \inf_{b \in C_\delta(B)} |G_{M,B}(b,b)| - \sup_{b \in C_\delta(B)} |\hat{G}_{M,B}(b,b) - G_{M,B}(b,b)| \\
&= \inf_{b \in C_\delta(B)} |G_{M,B}(b,b)| + O_p(h_g^{R-1})
\end{aligned}$$

Note $G_{M,B}(b,b) = \int_{b_L}^{b} ... \int_{b_L}^{b} g(b, b_2, b_3, .., b_n)\, db_2...db_n$ and $g(b_1, .., b_n) = f(\beta^{-1}(b_1), .., \beta^{-1}(b_n))$ has R continuous derivatives on $[b_L, b_U]^n$. Using a Taylor expansion of $\dot{g}(.)$ around $(b_L, .., b_L)$, $G_{M,B}(b,b) = a(b - b_L)^{n-1} + o(|b - b_L|^{n-1})$ with $a \equiv g(b_L, .., b_L) > 0$, and it can be shown $\inf_{b \in C_\delta(B)} |G_{M,B}(b,b)| \geq \alpha h^{n-1}$ for some $\alpha > 0$ and $h = \max(h_g, h_G)$.[4] Then $R > n$ and $h = h_g$ for L large enough implies $\inf_{b \in C_\delta(B)} \left| \hat{G}_{M,B}(b,b) \right| \geq \alpha h_g^{n-1} + o_p(h_g^{n-1})$. Since

[4]For details, see Lemma *A6* in Li et.al 2002.

$\sup_{\tilde{b}_L \le b, b \in C_\delta(B)} |\hat{G}_{M,B}(\tilde{b}_L, b) - G_{M,B}(\tilde{b}_L, b)|$ and $\sup_{\tilde{b}_L \le b, b \in C_\delta(B)} |\hat{G}_{M,B}(\tilde{b}_L, b) - G_{M,B}(\tilde{b}_L, b)|$ are both bounded by $O_p(h_g^{R-1})$, it follows $\sup_{\tilde{b}_L \le b, b \in C_\delta(B)} |\frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)}| = O_p(h_g^{R-n})$. $\square$

Proof of Lemma B4. Proposition *A2 (ii)* in Li et.al (2002) showed $\sup_{b \in C_\delta(B)} |\hat{\xi}(b) - \xi(b)| = O_p(h_g^{R-(n-1)})$.

By definition, $\tilde{b}_L \in C_\delta(B)$. Note :

$$
\begin{aligned}
& \sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} |\hat{\xi}_l(b) - \xi_l(b)| \\
\le \; & \sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \frac{\int_{\tilde{b}_L}^b \hat{\xi}(t) \hat{g}_{M,B}(t,b) dt}{\hat{G}_{M,B}(b,b)} - \frac{\int_{\tilde{b}_L}^b \xi(t) g_{M,B}(t,b) dt}{G_{M,B}(b,b)} \right| + ... \\
& \sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \hat{\xi}(\tilde{b}_L) \frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \int_{b_L}^{\tilde{b}_L} \xi(t) \frac{g_{M,B}(t,b)}{G_{M,B}(b,b)} dt \right|
\end{aligned}
$$

The proof proceeds by showing the two terms both converge in probability to 0.

By definition, $\tilde{b}_L \ge b_L$, and $\int_{b_L}^{\tilde{b}_L} \xi(t) \frac{g_{M,B}(t,b)}{G_{M,B}(b,b)} dt$ is bounded between $\xi(b_L) \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)}$ and $\xi(\tilde{b}_L) \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)}$. With probability one,

$$
\begin{aligned}
& \sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \hat{\xi}(\tilde{b}_L) \frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \int_{b_L}^{\tilde{b}_L} \xi(t) \frac{g_{M,B}(t,b)}{G_{M,B}(b,b)} dt \right| \\
\le \; & \max\{\sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} T_1(b; \tilde{b}_L), \sup_{\tilde{b}_L \le b, (\tilde{b}_L, b) \in C^2_\delta(B)} T_2(b; \tilde{b}_L)\}
\end{aligned}
$$

where

$$
\begin{aligned}
T_1(b; \tilde{b}_L) & \equiv \left| \hat{\xi}(\tilde{b}_L) \frac{\hat{G}_{M,B}(\tilde{b}_L, b)}{\hat{G}_{M,B}(b,b)} - \xi(b_L) \frac{G_{M,B}(\tilde{b}_L, b)}{G_{M,B}(b,b)} \right| \\
T_2(b; \tilde{b}_L) & \equiv \left| \hat{\xi}(\tilde{b}_L) \frac{\hat{G}_{M,B}(\tilde{b}_L, b)}{\hat{G}_{M,B}(b,b)} - \xi(\tilde{b}_L) \frac{G_{M,B}(\tilde{b}_L, b)}{G_{M,B}(b,b)} \right|
\end{aligned}
$$

With probability one, for all $(b, \tilde{b}_L) \in C^2_\delta(B)$ such that $b \geq \tilde{b}_L$,

$$T_1(b; \tilde{b}_L) \leq \left| \hat{\xi}(\tilde{b}_L) \left(\frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right) \right| + \left| \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \left(\hat{\xi}(\tilde{b}_L) - \xi(b_L) \right) \right|$$

where $\left| \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right| \leq 1$ by construction. Thus

$$\begin{aligned} & \sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} T_1(b; \tilde{b}_L) \\ \leq \; & \sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right| \left| \hat{\xi}(\tilde{b}_L) \right| + \left| \hat{\xi}(\tilde{b}_L) - \xi(b_L) \right| \end{aligned}$$

Note $\left| \hat{\xi}(\tilde{b}_L) \right| \xrightarrow{p} |\xi(b^0_L)| < \infty$ and $\left| \hat{\xi}(\tilde{b}_L) - \xi(b_L) \right| \xrightarrow{p} 0$ by the uniform convergence of $\hat{\xi}$ over $C_\delta(B)$ and $\tilde{b}_L \xrightarrow{p} b^0_L$. By Lemma 3 $\sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \frac{\hat{G}_{M,B}(\tilde{b}_L,b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L,b)}{G_{M,B}(b,b)} \right| \xrightarrow{p} 0$ if $R > n$. Hence $\sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} T_1(b) \xrightarrow{p} 0$ if $R > n$. Then $\sup_{b \in \tilde{C}_T(B)} T_2(b) \xrightarrow{p} 0$ follows from the same arguments.

By the triangular inequality, for all $b \geq \tilde{b}_L$, and $(\tilde{b}_L, b) \in C^2_\delta(B)$

$$\begin{aligned} & \left| \int_{\tilde{b}_L}^{b} \hat{\xi}(t) \frac{\hat{g}_{M,B}(t,b)}{\tilde{G}_{M,B}(b,b)} dt - \int_{\tilde{b}_L}^{b} \xi(t) \frac{g_{M,B}(t,b)}{G_{M,B}(b,b)} dt \right| \\ \leq \; & \frac{1}{\inf_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} \left| \tilde{G}_{M,B}(b,b) \right|} \left\{ \begin{array}{c} \left| \int_{\tilde{b}_L}^{b} \hat{\xi}(t) \hat{g}_{M,B}(t,b) dt - \xi(t) g_{M,B}(t,b) dt \right| + \ldots \\ \xi(b_U) \left| \tilde{G}_{M,B}(b,b) - G_{M,B}(b,b) \right| \end{array} \right\} \end{aligned}$$

where $\int_{\tilde{b}_L}^{b} \xi(t) \frac{g_{M,B}(t,b)}{G_{M,B}(b,b)} dt \leq \xi(b_U)$ and $\left| \tilde{G}_{M,B}(b,b) \right| \geq |G_{M,B}(b,b)| - \left| \tilde{G}_{M,B}(b,b) - G_{M,B}(b,b) \right|$. It is shown in *Lemma B3* above that $\inf_{b \in C_\delta(B)} |G_{M,B}(b,b)| \geq \alpha h_g^{n-1} + o_p(h_g^{n-1})$ with $R > n$ and $\sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C^2_\delta(B)} |\tilde{G}_{M,B}(b,b) - G_{M,B}(b,b)| = O_p(h_g^{R-1})$. Furthermore for all $b \geq \tilde{b}_L$, and

$(\tilde{b}_L, b) \in C_\delta^2(B)$,

$$\begin{aligned} & \left| \int_{\tilde{b}_L}^{b} \hat{\xi}(t)\hat{g}_{M,B}(t,b)dt - \xi(t)g_{M,B}(t,b)dt \right| \\ \leq & \int_{\tilde{b}_L}^{b} \left| \hat{\xi}(t) - \xi(t) \right| \left| \hat{g}_{M,B}(t,b) \right| dt + \int_{\tilde{b}_L}^{b} \left| \xi(t) \right| \left| \hat{g}_{M,B}(t,b) - g_{M,B}(t,b) \right| dt \end{aligned}$$

The boundedness of $g_{M,B}$ and ξ implies $\sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C_\delta^2(B)} \left| \int_{\tilde{b}_L}^{b} \hat{\xi}(t)\hat{g}_{M,B}(t,b)dt - \xi(t)g_{M,B}(t,b)dt \right| = O_p(h_g^{R-(n-1)})$, which is the rate of convergence of $\sup_{b \in C_\delta(B)} \left| \hat{\xi} - \xi \right|$. As a result, $\sup_{\tilde{b}_L \leq b, (\tilde{b}_L, b) \in C_\delta^2(B)}$ $\left| \frac{\int_{\tilde{b}_L}^{b} \hat{\xi}(t)\hat{g}_{M,B}(t,b)dt}{\tilde{G}_{M,B}(b,b)} - \frac{\int_{\tilde{b}_L}^{b} \xi(t)g_{M,B}(t,b)dt}{G_{M,B}(b,b)} \right| = O_p(h_g^{R-2(n-1)})$ converges to zero when $R > 2(n-1)$. □

Proof of Lemma B5. The proof is similar to that of $\sup_{\tilde{b}_L \leq b, b \in C_\delta(B)} \left| \frac{\hat{G}_{M,B}(\tilde{b}_L, b)}{\hat{G}_{M,B}(b,b)} - \frac{G_{M,B}(\tilde{b}_L, b)}{G_{M,B}(b,b)} \right|$ $= O_p(h_g^{R-n})$. On the support of $C_\delta^2(B)$,

$$\left| \frac{\hat{g}_{M,B}}{\hat{G}_{M,B}} - \frac{g_{M,B}}{G_{M,B}} \right| \leq \frac{1}{\left| \hat{G}_{M,B} \right| \left| G_{M,B} \right|} \left(\left| G_{M,B} \right| \left| \hat{g}_{M,B} - g_{M,B} \right| + \left| g_{M,B} \right| \left| \hat{G}_{M,B} - G_{M,B} \right| \right)$$

It is shown $\sup_{C_\delta^2(B)} |\hat{G}_{M,B} - G_{M,B}| = O_p(h_G^{R-1})$ and $\sup_{C_\delta^2(B)} |\hat{g}_{M,B} - g_{M,B}| = O_p(h_g^{R-1})$. Besides, $\sup_{C_\delta^2(B)} |G_{M,B}| < \infty$ and $\sup_{C_\delta^2(B)} |g_{M,B}| < \infty$ implies supremum of the term in the bracket is $O_p(h_g^{R-1})$ as $h_g > h_G$ for L large enough. Hence $\sup_{b \in C_\delta(B)} \left| \frac{\hat{g}_{M,B}(b,b)}{\hat{G}_{M,B}(b,b)} - \frac{g_{M,B}(b,b)}{G_{M,B}(b,b)} \right| \leq$ $\frac{O_p(h_g^{R-1})}{\inf_{b \in C_\delta(B)} \left| \hat{G}_{M,B} \right| \inf_{b \in C_\delta(B)} \left| G_{M,B} \right|}$, where the two terms in the denominator are bounded below by $\alpha h_g^{n-1} + o(h_g^{n-1})$ and $\beta h_g^{n-1} + o(h_g^{n-1})$ respectively by some constant α and β. It follows the denominator is bounded below by $\gamma h_g^{2n-2} + o(h_g^{2n-2})$. Hence $\sup_{b \in C_\delta(B)} \left| \frac{\hat{g}_{M,B}(b,b)}{\hat{G}_{M,B}(b,b)} - \frac{g_{M,B}(b,b)}{G_{M,B}(b,b)} \right|$ $= O_p(h^{R-2n+1})$. □

1.3. Sharpness of the bounds

This appendix discusses the sharpness of the bounds on the revenue distributions in counterfactual first-price auctions in the benchmark model. (The proof of sharpness of bounds in counterfactual second-price auctions is similar and omitted.) Specifically, I show that for a given counterfactual first-price auction and a given revenue level t, each point within the bounds constructed from $G^0_{\mathbf{B}}$ will be the true counterfactual revenue distribution under a certain structure in the identified set. Let ψ be a generic structure $(\theta, F_{\mathbf{X}})$, and let $\Psi(G^0_{\mathbf{B}})$ denote the set of ψ that are observationally equivalent relative to a rationalizable distribution of bids $G^0_{\mathbf{B}}$. Formally, the sharpness of bounds means for any given rationalizable $G^0_{\mathbf{B}}$, any given counterfactual reserve price r, and any revenue level t, $\exists \psi \in \Psi(G^0_{\mathbf{B}})$ such that $F_{R^I(r)}(t;\psi) = p$ for all $p \in [F^l_{R^I(r)}(t; G^0_{\mathbf{B}}), F^u_{R^I(r)}(t; G^0_{\mathbf{B}}))$, where $F^k_{R^I(r)}(t; G^0_{\mathbf{B}})$ are robust bounds proposed in the paper.

The proof takes three steps. In the first step, I show the sharpness of bounds on the hypothetical marginal bid $b_0(x^*(r;\psi);\psi)$. Let $b^0_{k,r}(G^0_{\mathbf{B}}(\psi))$ be a shorthand for $b_0(x_k(r;\psi);\psi)$ for $k = l, h$, and $b^0_r(\psi)$ for $b_0(x^*(r;\psi);\psi)$. The following lemma gives the sharpness of bounds $[b^0_{l,r}, b^0_{h,r}]$ on $b^0_r(\psi)$.

Lemma C1 Consider any rationalizable distribution $G^0_{\mathbf{B}}$. For all $b \in [b^0_{l,r}(G^0_{\mathbf{B}}), b^0_{h,r}(G^0_{\mathbf{B}}))$, $\exists\ \psi \in \Psi(G^0_{\mathbf{B}})$ such that $b^0_r(\psi) = b$.

Proof. Consider the case where $\theta(\mathbf{X}) = \alpha X_i + (1-\alpha)\max_{j\neq i} X_j$, and $v_h(x) = x$ for all x and $\alpha \in (0,1)$. Let $F_{\mathbf{X}}$ be defined as $G^0_{\mathbf{B}}(\xi^{-1}(x_1; G^0_{\mathbf{B}}), .., \xi^{-1}(x_n; G^0_{\mathbf{B}}))$. Then $(\alpha, F_{\mathbf{X}})$ generates $G^0_{\mathbf{B}}$ in a Bayesian Nash equilibrium for all $\alpha \in (0,1)$. Also note the equilibrium bidding strategy is invariant in $\alpha \in (0,1)$ for the given $F_{\mathbf{X}}$. Then an argument similar to *Lemma 3* can be used to prove the sharpness of bounds on $b^0_r(\psi)$. $\square$

The second step shows bounds on $b^0_r(t;\psi) \equiv b_0(b^{-1}_r(t;\psi);\psi)$ are also sharp.

Lemma C2 Consider any rationalizable distribution $G^0_{\mathbf{B}}$, any counterfactual reserve price r. For any revenue level $t > r$, and for all $b \in [\delta^{-1}_{r,l}(t; G^0_{\mathbf{B}}), \delta^{-1}_{r,h}(t; G^0_{\mathbf{B}}))$, $\exists\ \psi \in \Psi(G^0_{\mathbf{B}})$ such that $b^0_r(t;\psi) = b$.

Proof. Note that for $t > r$, $b_r^0(t;\psi)$ is defined as the solution to the following minimization problem:

$$b_r^0(t;\psi) = \arg\min_{s\in[b_r^0(\psi),b_U^0]}[t - \delta_r(s; b_r^0(\psi), G_{\mathbf{B}}^0)]^2$$

where $b_r^0(\psi)$ and δ_r are defined as before. For a given t and $G_{\mathbf{B}}^0$, $b_r^0(\psi)$ is a parameter that determines the constraint set and the criterion function. The criterion function is continuous in its first argument s and the parameter $b_r^0(\psi)$. Therefore, the Theorem of Maximum can be used to show $b_r^0(t;\psi)$ is a continuous function of $b_r^0(\psi)$ for all t. Then the sharpness of bounds on $b_r^0(\psi)$ translates into the sharpness of bounds on $b_r^0(t;\psi)$. $\square$

For the final step, it suffices to note that $G_{\mathbf{B}}^0$ is invariant for all $\psi \in \Psi(G_{\mathbf{B}}^0)$ by construction, and that the bounds on $b_r^0(\psi)$ and $b_r^0(t;\psi)$ are both sharp.

2. Appendix for Chapter 2

2.1. Proofs of identification and consistency

Proof. *(Lemma 1)* Suppose $\mathbf{b}$ is such that $\Pr(x \in \xi_b') > 0$. Then by the definition of Γ, for all $\mathbf{x} \in \xi_b'$ *s.t.* $(-\mathbf{x}^T\mathbf{b} \le L(\mathbf{x}) \wedge P_{1|\mathbf{x}}^* < \frac{1}{2})$, $P_{1|\mathbf{x}}(\mathbf{b}, F_{\varepsilon|\mathbf{X}}) = \int 1(\varepsilon \ge -\mathbf{x}^T\mathbf{b})dF_{\varepsilon|\mathbf{X}=\mathbf{x}} \ge \frac{1}{2}\ \forall F_{\varepsilon|\mathbf{X}} \in \Gamma$. Likewise $\forall \mathbf{x} \in \xi_b'$ *s.t.* $(-\mathbf{x}^T\mathbf{b} \ge U(\mathbf{x}) \wedge P_{1|\mathbf{x}}^* > \frac{1}{2})$, $P_{1|\mathbf{x}}(\mathbf{b}, F_{\varepsilon|\mathbf{X}}) = \int 1(\varepsilon \ge -\mathbf{x}^T\mathbf{b})dF_{\varepsilon|\mathbf{X}=\mathbf{x}} \le \frac{1}{2}$ $\forall F_{\varepsilon|\mathbf{X}} \in \Gamma$. As a result $\forall \mathbf{x} \in \xi_b', P_{1|\mathbf{x}}^* \ne P_{1|\mathbf{x}}(\mathbf{b}, F_{\varepsilon|\mathbf{X}})\ \forall F_{\varepsilon|\mathbf{X}} \in \Gamma$. By our supposition, $\Pr(\mathbf{x} \in \xi_b') > 0$ and it implies $\Pr(\mathbf{x} \in X(\mathbf{b}, F_{\varepsilon|\mathbf{X}})) > 0\ \forall\ F_{\varepsilon|\mathbf{X}} \in \Gamma$. Hence $\boldsymbol{\beta}$ is identified relative to $\mathbf{b}$. Now Suppose $\mathbf{b}$ is such that $\Pr(\mathbf{x} \in \xi_b') = 0$. Then $\Pr(\mathbf{x} \in S(\mathbf{X}) \setminus \xi_b') = 1$ where $S(\mathbf{X})\backslash\xi_b' \equiv \{\mathbf{x} \in S(\mathbf{X}) : (-\mathbf{x}^T\mathbf{b} \le L(\mathbf{x}) \wedge P_{1|\mathbf{x}}^* \ge \frac{1}{2}) \vee (-\mathbf{x}^T\mathbf{b} \ge U(\mathbf{x}) \wedge P_{1|\mathbf{x}}^* \le \frac{1}{2}) \vee (L(\mathbf{x}) < -\mathbf{x}^T\mathbf{b} < U(\mathbf{x}))\}$. Then $\forall \mathbf{x}$ *s.t.* $-\mathbf{x}^T\mathbf{b} \le L(\mathbf{x}) \wedge P_{1|\mathbf{x}}^* \ge \frac{1}{2}$, pick $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ *s.t.* *(i)* $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ is continuous in ε and $L(\mathbf{x}) \le \sup Med(\varepsilon|\mathbf{x})$ and $\inf Med(\varepsilon|\mathbf{x}) \le U(\mathbf{x})$; and *(ii)* $\int 1(\mathbf{x}^T\mathbf{b} + \varepsilon \ge 0)dF_{\varepsilon|\mathbf{X}=\mathbf{x}} = P_{1|\mathbf{x}}^*$. This can be done because $-\mathbf{x}^T\mathbf{b} \le L(\mathbf{x}) \le \sup Med(\varepsilon|\mathbf{x})$

of $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ requires $\int 1(\mathbf{x}^T\mathbf{b}+\varepsilon \geq 0)dF_{\varepsilon|\mathbf{X}=\mathbf{x}} \geq \frac{1}{2}$. Likewise $\forall$ $\mathbf{x}$ $s.t.$ $-\mathbf{x}^T\mathbf{b} \geq U(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} \leq \frac{1}{2}$, we can pick $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ $s.t.$ (i) holds and $\int 1(\mathbf{x}^T\mathbf{b}+\varepsilon \geq 0)dF_{\varepsilon|\mathbf{X}=\mathbf{x}} = P^*_{1|\mathbf{x}}$. And $\forall$ $\mathbf{x}$ $s.t.$ $L(\mathbf{x}) < -\mathbf{x}^T\mathbf{b} < U(\mathbf{x})$, we can always pick $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ $s.t.$ $\int 1(\mathbf{x}^T\mathbf{b}+\varepsilon \geq 0)dF_{\varepsilon|\mathbf{X}=\mathbf{x}} = P^*_{1|\mathbf{x}}$ (regardless of the value of $P^*_{1|\mathbf{x}}$) while (i) still holds. Finally let $F_{\varepsilon|\mathbf{X}}$ be such that $F_{\varepsilon|\mathbf{X}=\mathbf{x}}$ is picked as above $\forall$ $\mathbf{x} \in S(\mathbf{X}) \setminus \xi'_b$. We have shown $P_{1|x}(\mathbf{b}, F_{\varepsilon|\mathbf{X}}) = P^*_{1|\mathbf{x}}$ $\forall \mathbf{x} \in S(\mathbf{X}) \setminus \xi_b$ (equivalent to $a.e.F_{\mathbf{X}}$ since $\Pr(\mathbf{x} \in \xi'_b) = 0$). Hence $\exists$ $F_{\varepsilon|\mathbf{X}} \in \Gamma$ $s.t.$ $\Pr(\mathbf{x} \in X(\mathbf{b}, F_{\varepsilon|\mathbf{X}})) = 0$ and $\mathbf{b}$ is observationally equivalent to $\boldsymbol{\beta}$. □

Proof. *(Corollary 1 of Lemma 1)* The proof is similar to Lemma 1 and is omitted. □

Proof. *(Corollary 2 of Lemma 1)* $\Theta'_I \neq \varnothing$ and $\Theta_I \neq \varnothing$ because the true coefficient β belongs to both. To prove convexity, suppose $\mathbf{b}_1 \in \Theta'_I, \mathbf{b}_2 \in \Theta'_I$. Then $\Pr(\mathbf{x} \in \xi'_{b_1}) = \Pr(\mathbf{x} \in \xi'_{b_2}) = 0$. Let $\mathbf{b}_\alpha \equiv \alpha\mathbf{b}_1 + (1-\alpha)\mathbf{b}_2 \in \Theta'_I$ for some $\alpha \in (0,1)$ and ξ_{b_α} be defined as before for b_α. Note $\forall$ $\mathbf{x} \in \xi'_{b_\alpha}$, either $(-\mathbf{x}^T\mathbf{b}_\alpha \geq U(\mathbf{x}) \wedge P^*_{1|x} > 1/2)$ or $(-\mathbf{x}^T\mathbf{b}_\alpha \leq L(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} < 1/2)$. Consider the former case. Then it must be $P^*_{1|\mathbf{x}} > 1/2$ and either $-\mathbf{x}^T\mathbf{b}_1 \geq U(\mathbf{x})$ or $-\mathbf{x}^T\mathbf{b}_2 \geq U(\mathbf{x})$. This implies either $\mathbf{x} \in \xi'_{b_1}$ or $\mathbf{x} \in \xi'_{b_2}$. Symmetric argument applies to the case $(-\mathbf{x}^T b_\alpha \leq L(\mathbf{x}) \wedge P^*_{1|\mathbf{x}} < 1/2)$. It follows that $\xi'_{b_\alpha} \subseteq (\xi'_{b_1} \cup \xi'_{b_2})$. Then $\Pr(\mathbf{x} \in \xi'_{b_\alpha}) \leq \Pr(\mathbf{x} \in \xi'_{b_1}) + \Pr(\mathbf{x} \in \xi'_{b_2}) = 0$, and $\mathbf{b}_\alpha \in \Theta'_I$. The convexity of Θ_I is proven in the same way. □

Proof. *(Proposition 1)* By *BCQ-2* and *Lemma 1*, it suffices to show $\Pr(\mathbf{X} \in \xi_{\mathbf{b}}) > 0$ for all $\mathbf{b} \neq \boldsymbol{\beta}$, where $\xi_{\mathbf{b}} \equiv \{\mathbf{x} : (-\mathbf{x}'\mathbf{b} \leq L(\mathbf{x}) \wedge -\mathbf{x}'\boldsymbol{\beta} > U(\mathbf{x})) \vee (-\mathbf{x}'\mathbf{b} \geq U(\mathbf{x}) \wedge -\mathbf{x}'\boldsymbol{\beta} < L(\mathbf{x}))\}$. By *SX1-(a)*, $\Pr(\mathbf{X}'_{-J}(\boldsymbol{\beta}_{-J} - \mathbf{b}_{-J}) \neq \mathbf{0}) > 0$. Without loss of generality, let $\Pr(\mathbf{X}'_{-J}\boldsymbol{\beta}_{-J} < \mathbf{X}'_{-J}\mathbf{b}_{-J}) > 0$. Then by *SX1-(b),(c)*, $\Pr(-\mathbf{X}'\boldsymbol{\beta} < L(\mathbf{X}) \leq \mathbf{U}(\mathbf{X}) < -\mathbf{X}'\mathbf{b}) > 0$. □

Proof. *(Corollary 1 of Proposition 1)* The parameter space is compact by supposition. For player 1, *BCQ-2* is satisfied with $L(\mathbf{x})$ =$U(\mathbf{x})$ =$p^2(\mathbf{x})$ a.e. $F_{\mathbf{X}}$, and $L = 0$, $U = 1$. By *REG-(i)*, X_l impacts $p^2(\mathbf{X})$ and $p^1(\mathbf{X})$ but not $\mathbf{X}'\boldsymbol{\beta}_1$, and $\Pr\{\mathbf{X}'_{-l}(\mathbf{b}_{1,-l} - \tilde{\mathbf{b}}_{1,-l}) \neq 0\} > 0$ for all $\mathbf{b}_1, \tilde{\mathbf{b}}_1 \in \Theta_1$. Hence *SX-1 (a)* is satisfied. Suppose $\Pr\{\mathbf{X}'_{-l}(\mathbf{b}_{-l} - \tilde{\mathbf{b}}_{-l}) \neq 0 \wedge sgn(\mathbf{X}'_{-l}\mathbf{b}_{-l}) \neq sgn(\mathbf{X}'_{-l}\tilde{\mathbf{b}}_{-l})) > 0$, then *SX1-(b)* is satisfied. Otherwise without loss of generality consider the case $\Pr(\mathbf{X}'_{-l}\mathbf{b}_{1,-l} > \mathbf{X}'_{-l}\tilde{\mathbf{b}}_{1,-l} > 0) > 0$. Then *REG-(ii)* and the closedness under scalar multiplications in *REG-(v)* guarantee that *SX1-(b)* is satisfied. Note for all $\bar{\mathbf{x}}_{-l}$ and all $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ in BNE, substitution implies

$$p_1(\bar{\mathbf{x}}_{-l}, x_l) = F_{\epsilon_1|\bar{\mathbf{x}}_{-l},x_l}[\bar{\mathbf{x}}_{-l}\boldsymbol{\beta}_{1,-l} + F_{\epsilon_2|\bar{\mathbf{x}}_{-l},x_l}(\bar{\mathbf{x}}_{-l}\boldsymbol{\beta}_{2,-l} + x_l\beta_{2,l} + p_1(\bar{\mathbf{x}}_{-h_1}, x_{h_1}))]$$

Under the regularity conditions in *ERR*, an application of Schauder's Fixed Point Theorem shows $p_1(\bar{\mathbf{x}}_{-l}, x_l)$ is continuous in x_l for all $\bar{\mathbf{x}}_{-l}$. Also note

$$p^2(x_l, \bar{\mathbf{x}}_{-l}) = F_{\varepsilon_2|\mathbf{X}=\mathbf{x}}[\bar{\mathbf{x}}'_{-l}\boldsymbol{\beta}_{2,-l} + x_l\beta_{2l} + p^1(x_l, \bar{\mathbf{x}}_{-l})]$$

Since $\beta_{2l} \neq 0$ and $p^1(x_l, \mathbf{x}_{-l})$ is continuous in x_l, *REG-(iii),(iv)* then implies that *SX1-(c)* is satisfied. □

Proof. *(Proposition 2)* The proof requires slight changes from that of Lemma 2 in Manski (1985), as one of the regressors is a known function of the other regressors, and the coefficient in front of it is normalized to be 1. The proof is omitted for brevity. □

Proof. *(Lemma 2)* By law of iterated expectations,

$$\begin{aligned} Q(\mathbf{b}) &= E[(-U(\mathbf{X}) - \mathbf{X}'\mathbf{b})^2_+ | P^*_{1|\mathbf{X}} \geq 1/2] \Pr(P^*_{1|\mathbf{X}} \geq 1/2) \\ + & \quad E[(-L(\mathbf{X}) - \mathbf{X}'\mathbf{b})^2_- | P^*_{1|\mathbf{X}} \leq 1/2] \Pr(P^*_{1|\mathbf{X}} \leq 1/2) \end{aligned}$$

By definition, $\forall \mathbf{b} \in \Theta_I$, $\Pr(-\mathbf{X}'\mathbf{b} - U(\mathbf{X}) > 0 \wedge P^*_{1|\mathbf{X}} \geq 1/2) = 0$, and $\Pr(-\mathbf{X}'\mathbf{b} - L(\mathbf{X}) < 0 \wedge P^*_{1|\mathbf{X}} \leq 1/2) = 0$. Therefore $Q(\mathbf{b}) = 0$ for all $\mathbf{b} \in \Theta_I$. On the other hand, for any $\mathbf{b} \notin \Theta_I$, at least one of the four following events have positive probability: "$-\mathbf{X}'\mathbf{b} < L(\mathbf{X}) \wedge P^*_{1|\mathbf{X}} = 1/2$", "$-\mathbf{X}'\mathbf{b} > U(\mathbf{X}) \wedge P^*_{1|\mathbf{X}} = 1/2$", "$-\mathbf{X}'\mathbf{b} \geq U(\mathbf{X}) \wedge P^*_{1|\mathbf{X}} > 1/2$" or "$-\mathbf{X}'\mathbf{b} \leq L(\mathbf{X}) \wedge P^*_{1|\mathbf{X}} < 1/2$". Without loss of generality, let the last event occur with positive probability. Then *SX-2* ensures the inequality is strict with positive probability. This implies the second term in $Q(\mathbf{b})$ will be strictly positive. Similar arguments can be applied to prove $Q(\mathbf{b}) > 0$ if any of the other events has positive probability. □

Proof. *(Theorem 1)* It can be shown that the objective function $Q(\mathbf{b})$ is continuous in $\mathbf{b}$ under the regularity conditions. Below I will show that $\hat{Q}_n(\mathbf{b})$ converges to $Q(\mathbf{b})$ in probability uniformly over Θ. The rest of the proof follows similar steps in Proposition 3 of Manski and Tamer (2002) and is omitted for brevity. First note by Lemma 8.10 in Newey and McFadden (1994), under *RD-1*, *TF* and *K*,

$$\sup_{\mathbf{x} \in S(\mathbf{X})} |\hat{p}(\mathbf{x}) - p(\mathbf{x})| = o_p(n^{-1/4})$$

By the mean value expansion of $\Lambda(\hat{p}(\mathbf{x}) - \frac{1}{2})$ and $\Lambda(\frac{1}{2} - \hat{p}(\mathbf{x}))$ around $\Lambda(p(\mathbf{x}) - \frac{1}{2})$ and $\Lambda(\frac{1}{2} - p(\mathbf{x}))$, the uniform consistency of the first step estimator of $\hat{p}_i$, and the Law of Large Numbers, we have for all $\mathbf{b} \in \Theta$,

$$\hat{Q}_n(\mathbf{b}) \xrightarrow{p} Q(\mathbf{b})$$

Note that $\hat{Q}_n(\mathbf{b})$ is convex in $\mathbf{b}$ for all n and the parameter space is compact and convex by *PAR*. Then by Theorem 2.7 in Newey and McFadden (1994), point-wise convergence of convex functions implies that $\hat{Q}_n(\mathbf{b})$ converges uniformly in probability to $Q(\mathbf{b})$ over Θ. □

2.2. Asymptotic normality under point identification

Throughout this subsection, I maintain that all conditions for point identification of $\boldsymbol{\beta}$ (i.e. *BCQ-2, PAR and SX-1*) and the regularity conditions *RD*, *TF* and *K* are satisfied. Define $G_n(\tau, p) \equiv \sum_{i=1}^{n} g_i(\tau, p; n)$, where

$$g_i(\tau, p; n) \equiv \Lambda^u(p_i)[(\nu_i^u - n^{-1/2}\mathbf{x}_i'\tau)_+^2 - (v_i^u)_+^2] + \Lambda^l(p_i)[(\nu_i^l - n^{-1/2}\mathbf{x}_i'\tau)_-^2 - (v_i^l)_-^2]$$

where $\nu_i^l \equiv -L(\mathbf{x}_i) - \mathbf{x}_i'\boldsymbol{\beta}_0$, $\nu_i^u \equiv -U(\mathbf{x}_i) - \mathbf{x}_i'\boldsymbol{\beta}_0$, and p_i, $\Lambda^u(p_i)$, $\Lambda_i^l(p_i)$ denote the true choice probability $p(\mathbf{x}_i) \equiv \Pr(Y = 1|\mathbf{X} = \mathbf{x}_i)$, $\Lambda(p_i - \frac{1}{2})$, $\Lambda(\frac{1}{2} - p_i)$ respectively. By definition, $G_n(\tau, \hat{p})$ is a convex and continuously differentiable function in τ and is minimized at $\tau_n = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Note $g_i(\tau, p; n)f(\mathbf{x})$ is continuously differentiable in τ for all i and n, and under conditions *RD* and *TF*, $\int \sup_{\tau \in N_0} ||\nabla_\tau g_i(\tau, p; n)f(\mathbf{x})||d\mathbf{x} < \infty$ for a small open neighborhood N_0 around the zero vector. Therefore by *Lemma 3.6* in Newey and McFadden (1994) and an application of dominated convergence theorem, it can be shown for all i, n,

$$\nabla_\tau E(g_i(\tau, p; n))|_{\tau=0} = -2n^{-1/2}E\{[\Lambda^l(p_i)(V_i^l)_- + \Lambda^u(p_i)(V_i^u)_+]\mathbf{X}_i\} = 0 \tag{.5}$$

It can also be shown by direct calculation of gradients that under conditions *RD* and *TF* for all i, n, the Hessian matrix

$$\nabla_{\tau\tau'}^2 E(g_i(\tau, p; n))|_{\tau=0} = 2n^{-1}E\{[\Lambda^l(p_i)1(V_i^l < 0) + \Lambda^u(p_i)1(V_i^u > 0)]\mathbf{X}_i\mathbf{X}_i'\}$$

exists and is continuous at $\tau = 0$. Thus by a Taylor expansion, $\Gamma_n(\tau, p) = \frac{1}{2}\tau' J\tau + o(1)$, where $J \equiv \nabla^2_{\tau\tau'} E(g_i(\tau, p; n))|_{\tau=0}$. Let $\hat{p}_i$, $\hat{\Lambda}^u_i$, $\hat{\Lambda}^l_i$ denote $\hat{p}(\mathbf{x}_i)$, $\Lambda(\hat{p}_i - \frac{1}{2})$, $\Lambda(\frac{1}{2} - \hat{p}_i)$. Define

$$\begin{aligned}
\omega_n(p) &\equiv n^{-1/2} \textstyle\sum_{i=1}^n \xi(\mathbf{X}_i, p) \\
\xi(\mathbf{X}_i, p) &\equiv 2[\Lambda^l(p_i)(V_i^l)_- + \Lambda^u(p_i)(V_i^u)_+]\mathbf{X}_i \\
\rho_n(\mathbf{X}_i, p, \tau) &\equiv \Lambda^u(p_i)[(V_i^u - n^{-1/2}\mathbf{X}_i'\tau)_+^2 - (V_i^u)_+^2] + ... \\
&\quad \Lambda^l(p_i)[(V_i^l - n^{-1/2}\mathbf{X}_i'\tau)_-^2 - (V_i^l)_-^2] + n^{-1/2}\tau'\xi(\mathbf{X}_i, p)
\end{aligned}$$

Since $E[\xi(\mathbf{X}_i, p)] = 0$, we can rewrite for any generic choice probability $\tilde{p}$

$$G_n(\tau, \tilde{p}) = \Gamma_n(\tau, p) - \tau'\omega_n(\tilde{p}) + \textstyle\sum_{i=1}^n \{\rho_n(\mathbf{X}_i, \tilde{p}, \tau) - E[\rho_n(\mathbf{X}_i, p, \tau)]\}$$

Below I will show that $G_n(\tau, \hat{p}) = \frac{1}{2}\tau' J\tau - \tau'\omega_n(\hat{p}) + o_p(1)$. This will enable us to use the Convexity Lemma in Pollard (1991) and an argument similar to Buchinsky and Hahn (1998) to show τ_n is asymptotically equivalent to the asymptotically normal "maximizer" of $\frac{1}{2}\tau' J\tau - \tau'\omega_n(\hat{p})$.

Lemma B1 *Under the identifying conditions BCQ-2, PAR, SX-1,2 and regularity conditions RD-1,2, TF-1,2 and K,* $\omega_n(\hat{p}) \xrightarrow{d} N(0, \Sigma)$.

Proof. The proof proceeds by checking conditions of Theorem 8.11 in Newey and McFadden (1994). That is, it suffices to show that there exists a vector of functionals $\Psi(\mathbf{x}; h, f)$, which is linear in (h, f) and satisfies:(i) For (h, f) with $||(h, f) - (h_0, f_0)||_{\text{sup}}$ small, there exists some $b(\mathbf{x}) : S(\mathbf{X}) \to \mathbb{R}^1$ such that

$$||\xi(\mathbf{x}_i; h, f) - \xi(\mathbf{x}_i; h_0, f_0) - \Psi(\mathbf{x}; h - h_0, f - f_0)|| \le b(\mathbf{x})||(h, f) - (h_0, f_0)||^2_{\text{sup}}$$

with $E[b(\mathbf{X})] < \infty$, where $||.||_{\sup}$ denotes the sup of the Euclidean norm of a vector-valued function defined on $S(\mathbf{X})$; (ii) $||\Psi(\mathbf{x}; h, f)|| \leq c(\mathbf{x})||(h, f)||_{\sup}$, with $E[c(\mathbf{x})^2] < \infty$; (iii) $\exists$ $\lambda_1(\mathbf{x})$ and $\lambda_2(\mathbf{x})$ s.t.

$$E[\Psi(\mathbf{x}; h, f)] = \int \lambda_1(\mathbf{x})h(\mathbf{x}) + \lambda_2(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

(iv) For $j = 1, 2$, $\lambda_j(\mathbf{x})$ is continuous almost everywhere, $\int |\lambda_j(\mathbf{x})|d\mathbf{x} < \infty$, and $\exists\ \eta > 0$ s.t.

$$E[\sup_{|\iota| \leq \eta} ||\lambda_j(\mathbf{X} + \boldsymbol{\iota})||^4] < \infty$$

To verify condition (i), define the linear functional

$$\Psi(\mathbf{x}_i; h, f) \equiv \left\{ \frac{a_1(\mathbf{x}_i; h_0, f_0)}{f_0(\mathbf{x}_i)} h(\mathbf{x}_i) - \frac{h_0(\mathbf{x}_i) a_1(\mathbf{x}_i; h_0, f_0)}{f_0(\mathbf{x}_i)^2} f(\mathbf{x}_i) \right\} \mathbf{x}_i$$

where

$$a_1(\mathbf{x}_i; h, f) \equiv 2 \left\{ \Lambda' \left(\frac{h(\mathbf{x})}{f(\mathbf{x})} - \frac{1}{2} \right) (v_i^u)_+ - \Lambda' \left(\frac{1}{2} - \frac{h(\mathbf{x})}{f(\mathbf{x})} \right) (v_i^l)_- \right\}$$

Then by 2nd order Taylor expansion,

$$\begin{aligned} & \xi(\mathbf{x}_i; h, f) - \xi(\mathbf{x}_i; h_0, f_0) - \Psi(\mathbf{x}; h - h_0, f - f_0) \\ = \;& [h(\mathbf{x}_i) - h_0(\mathbf{x}_i) \;\; f(\mathbf{x}_i) - f_0(\mathbf{x}_i)] \begin{pmatrix} A_1(\mathbf{x}_i; \tilde{h}, \tilde{f}) & A_2(\mathbf{x}_i; \tilde{h}, \tilde{f}) \\ A_2(\mathbf{x}_i; \tilde{h}, \tilde{f}) & A_3(\mathbf{x}_i; \tilde{h}, \tilde{f}) \end{pmatrix} \begin{bmatrix} h(\mathbf{x}_i) - h_0(\mathbf{x}_i) \\ f(\mathbf{x}_i) - f_0(\mathbf{x}_i) \end{bmatrix} \end{aligned}$$

where $\tilde{h}$, $\tilde{f}$ are on the line segment connecting (h, f) and (h_0, f_0), and

$$\begin{aligned} A_1(\mathbf{x}; h, f) &\equiv \frac{a_2(\mathbf{x}; h, f)}{f(\mathbf{x})^2}\mathbf{x} \\ A_2(\mathbf{x}; h, f) &\equiv \left\{ -\frac{h(\mathbf{x})a_2(\mathbf{x}; h, f)}{f(\mathbf{x})^3} - \frac{a_1(\mathbf{x}; h, f)}{f(\mathbf{x})^2} \right\}\mathbf{x} \\ A_3(\mathbf{x}; h, f) &\equiv \left\{ \frac{h(\mathbf{x})^2 a_2(\mathbf{x}; h, f)}{f(\mathbf{x})^4} + \frac{2h(\mathbf{x})a_1(\mathbf{x}; h, f)}{f(\mathbf{x})^3} \right\}\mathbf{x} \end{aligned}$$

where

$$a_2(\mathbf{x}; h, f) \equiv 2\left\{ \Lambda'' \left(\frac{h(\mathbf{x})}{f(\mathbf{x})} - \frac{1}{2} \right) (v_i^u)_+ + \Lambda'' \left(\frac{1}{2} - \frac{h(\mathbf{x})}{f(\mathbf{x})} \right) (v_i^l)_- \right\}$$

Then by regularity conditions in *RD-2*, there exists some positive constant C where

$$||\xi(\mathbf{x}_i; h, f) - \xi(\mathbf{x}_i; h_0, f_0) - \Psi(\mathbf{x}; h - h_0, f - f_0)|| \le C||\mathbf{x}_i|| \cdot |B(\mathbf{x}_i)| \cdot ||(h, f) - (h_0, f_0)||_{\sup}^2$$

and condition (i) above is satisfied. Also condition (ii) is satisfied by conditions in *RD-2*. Then note

$$\begin{aligned} E[\Psi(\mathbf{X}; h, f)] &= \int \left\{ \frac{a_1(\mathbf{x}; h_0, f_0)}{f_0(\mathbf{x})} h(\mathbf{x}) - \frac{h_0(\mathbf{x})a_1(\mathbf{x}; h_0, f_0)}{f_0(\mathbf{x})^2} f(\mathbf{x}) \right\} \mathbf{x} f_0(\mathbf{x}) d\mathbf{x} \\ &= \int \left\{ a_1(\mathbf{x}; h_0, f_0)h(\mathbf{x}) - \frac{h_0(\mathbf{x})}{f_0(\mathbf{x})} a_1(\mathbf{x}; h_0, f_0) f(\mathbf{x}) \right\} \mathbf{x} d\mathbf{x} \end{aligned}$$

Hence condition (iii) is satisfied with $\lambda_1(\mathbf{x}) = a_1(\mathbf{x}; h_0, f_0)\mathbf{x}$ and $\lambda_2(\mathbf{x}) = \frac{h_0(\mathbf{x})}{f_0(\mathbf{x})} a_1(\mathbf{x}; h_0, f_0)\mathbf{x}$. Condition (iv) is also satisfied because $||\mathbf{X}|| \cdot |\max(0, -V_i^l)|$ and $||\mathbf{X}|| \cdot |\max(0, V_i^u)|$ both have finite 4th moments, Λ' is bounded, and $f_0(\mathbf{x})$ is bounded away from zero. Let $\Delta(\mathbf{x}, y; h_0, f_0) = \lambda_1(\mathbf{x})y + \lambda_2(\mathbf{x})$. It then follows from Theorem 8.11 in Newey and McFadden (1994) that

$$\omega_n(\hat{p}) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma \equiv Var[\xi(\mathbf{X}; h_0, f_0)+\Delta(\mathbf{X}, Y; h_0, f_0)]$. *Q.E.D.* □

The rest of the proof shows that $\tau_n = J^{-1}\omega_n(\hat{p}) + o_p(1)$ and therefore $n^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution.

Lemma B2 *Under the conditions of Lemma B1 above, for any fixed* τ *in the compact parameter space* Θ,

$$\textstyle\sum_{i=1}^n \{\rho_n(\mathbf{X}_i; \hat{p}, \tau) - E[\rho_n(\mathbf{X}_i, p, \tau)]\} = o_p(1)$$

Proof. First consider the term

$$\begin{aligned} & \textstyle\sum_{i=1}^n [\rho_n(\mathbf{X}_i; \hat{p}, \tau) - \rho_n(\mathbf{X}_i, p, \tau)] \\ = & \textstyle\sum_{i=1}^n S_n^u(\mathbf{X}_i; \tau)[\Lambda^u(\hat{p}_i) - \Lambda^u(p_i)] + \sum_{i=1}^n S_n^l(\mathbf{X}_i; \tau)[\Lambda^l(\hat{p}_i) - \Lambda^l(p_i)] \end{aligned}$$

where

$$\begin{aligned} S_n^u(\mathbf{X}_i; \tau) &= (v_i^u - n^{-1/2}\mathbf{X}_i'\tau)_+^2 - (v_i^u)_+^2 + 2n^{-1/2}\mathbf{X}_i'\tau(v_i^u)_+ \\ S_n^l(\mathbf{X}_i; \tau) &= (v_i^l - n^{-1/2}\mathbf{X}_i'\tau)_-^2 - (v_i^l)_-^2 + 2n^{-1/2}\mathbf{X}_i'\tau(v_i^l)_- \end{aligned}$$

Hence

$$\begin{aligned} & \textstyle|\sum_{i=1}^n [\rho_n(\mathbf{X}; \hat{p}, \tau) - \rho_n(\mathbf{X}, p, \tau)]| \\ \leq \; & \textstyle C_1 \sup_{\mathbf{x}\in S(\mathbf{X})} |\hat{p}(\mathbf{x}) - p(\mathbf{x})| \cdot \left\{|\sum_{i=1}^n S_n^l(\mathbf{X}_i; \tau)| + |\sum_{i=1}^n S_n^u(\mathbf{X}_i; \tau)|\right\} \end{aligned}$$

where C is the bound on the first derivative on Λ. By Lemma 8.10 and the regularity conditions *RD-2* on the kernel, the bandwidth and higher moments of Y_i, $\sup_{\mathbf{x}\in S(\mathbf{X})} |\hat{p}(\mathbf{x}) -$

$p(\mathbf{x})| = o_p(1)$. Also note

$$\begin{aligned} S_n^u(\mathbf{X};\tau) &= 1\left\{\min(v^u, v^u - n^{-1/2}\mathbf{X}'\tau) \geq 0\right\} n^{-1}(\mathbf{X}'\tau)^2 + ... \\ & 1\{v^u - n^{-1/2}\mathbf{X}'\tau < 0 < v^u\}[-(v^u)^2 + 2n^{-1/2}\mathbf{X}'\tau] + ... \\ & 1\{v^u - n^{-1/2}\mathbf{X}'\tau > 0 > v^u\}(v^u - n^{-1/2}\mathbf{X}'\tau)^2 \end{aligned}$$

Note $-(v^u)^2 + 2n^{-1/2}\mathbf{X}'\tau v^u \geq 0$ both in the second and third event. Therefore $S_n^u(\mathbf{X};\tau) \leq n^{-1}(\mathbf{X}'\tau)^2$. By *RD-2*, $\sum_{i=1}^n S_n^u(\mathbf{X}_i;\tau) = O_p(1)$. By similar arguments, $\sum_{i=1}^n S_n^l(\mathbf{X}_i;\tau) = O_p(1)$, and

$$|\textstyle\sum_{i=1}^n[\rho_n(\mathbf{X}_i;\hat{p},\tau) - \rho_n(\mathbf{X}_i,p,\tau)]| \leq o_p(1)O_p(1) = o_p(1)$$

Now consider the second term $\sum_{i=1}^n T_{n,i} \equiv \sum_{i=1}^n \{\rho_n(\mathbf{X}_i;p,\tau) - E[\rho_n(\mathbf{X}_i,p,\tau)]\}$. Note

$$E[(\textstyle\sum_{i=1}^n T_{n,i})^2] \leq \sum_{i=1}^n E[\rho_n(\mathbf{X}_i;p,\tau)^2] \leq n^{-2}\sum_{i=1}^n E\{[\Lambda^u(p_i) + \Lambda^l(p_i)]^2(\mathbf{X}'\tau)^4\}]$$

where the first inequality is due to the cancellation of cross-product terms as a result of independence across observations, and the second inequality follows from the above bounds on S_n^u and S_n^l. Since both Λ^u and Λ^l are positive and bounded, $E[(\sum_{i=1}^n T_{n,i})^2] = O(n^{-1})$. Since $E(\sum_{i=1}^n T_{n,i}) = 0$ by construction, this implies $(\sum_{i=1}^n T_{n,i})^2$ converges in mean square and therefore is $o_p(1)$. *Q.E.D.* □

The 2nd-order Taylor expansion $\Gamma_n(\tau,p) = \frac{1}{2}\tau' J\tau + o(1)$ and the preceding two lemmae imply for each fixed τ,

$$G_n(\tau,\hat{p}) = \tfrac{1}{2}\tau' J\tau - \tau'\omega_n(\hat{p}) + o_p(1)$$

Then this point-wise convergence is strengthened into uniform convergence on the support of τ, which is a convex, compact subset of $\mathbb{R}^K$. This enables us to rewrite

$$G_n(\tau,\hat{p}) = \tfrac{1}{2}(\tau-\eta_n)J'(\tau-\eta_n) - \frac{1}{2}\eta_n' J\eta_n + r_n(\tau)$$

where $\eta_n \equiv J^{-1}\omega_n(\hat{p})$. and $\sup_{\tau\in T}|r_n(\tau)| = o_p(1)$ for any compact $T \subset \mathbb{R}^K$. Then it suffices to show $\tau_n = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically equivalent to η_n, i.e. $\tau_n = \eta_n + o_p(1)$. This result follows from standard arguments in *Theorem 1* in Pollard (1991) and *Lemma 3* in Buchinsky and Hahn (1998) and the proof is omitted.

3. Appendix for Chapter 3

Let $\mathbf{B}^2(S(\mathbf{X}))$ denote the space of bounded, continuous $\mathbb{R}^2$-valued functions, i.e. $\mathbf{B}^2(S(\mathbf{X})) = B(S(\mathbf{X})) \otimes B(S(\mathbf{X}))$, where $B(S(\mathbf{X}))$ is the space of bounded, real-valued functions defined on $\mathbf{X}$.Define the norm on $\mathbf{B}^2(S(\mathbf{X}))$ as $||\mathbf{f}(\mathbf{x})|| = \sup_{j\in\{0,1\},\mathbf{x}\in S(\mathbf{X})}|f_j(\mathbf{x})|$.

Lemma A1 $(\mathbf{B}^2(S(\mathbf{X})), ||.||)$ is a complete normed vector space.

Proof. The proof is standard and omitted for brevity. □

Lemma A2 Suppose the operator $T : \mathbf{B}^2(S(\mathbf{X})) \to \mathbf{B}^2(S(\mathbf{X}))$ satisfies (a) $\forall \mathbf{f}, \mathbf{g} \in \mathbf{B}^2(S(\mathbf{X}))$, $\mathbf{f}(\mathbf{x}) \le \mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in S(\mathbf{X})$ implies $(T\circ\mathbf{f})(\mathbf{x}) \le (T\circ\mathbf{g})(\mathbf{x})$ for all $\mathbf{x} \in S(\mathbf{X})$ (where the inequality is component-wise in $\mathbb{R}^2$); (b) $\exists\beta \in (0,1).s.t.T\circ(\mathbf{f}(\mathbf{x}) + a\mathbf{1}_2) \le T\circ\mathbf{f}(x) + \beta a$, $\forall \mathbf{f} \in \mathbf{B}^2(S(\mathbf{X}))$, $a \ge 0$, $\mathbf{x} \in S(\mathbf{X})$ (where $\mathbf{1}_2 \equiv [1\ 1]'$). Then T is an contraction mapping with modulus β.

Proof. We need to show that $\forall \mathbf{f}, \mathbf{g} \in \mathbf{B}^2(S(\mathbf{X})), ||T \circ \mathbf{f} - T \circ \mathbf{g}|| \leq \beta ||\mathbf{f} - \mathbf{g}||$.Note:

$$\begin{aligned}
\mathbf{f} \quad &\leq \quad \mathbf{g} + ||\mathbf{f} - \mathbf{g}||\mathbf{1}_2 \\
&\implies T \circ \mathbf{f} \leq T \circ (\mathbf{g} + ||\mathbf{f} - \mathbf{g}||\mathbf{1}_2) \leq T \circ \mathbf{g} + \beta ||\mathbf{f} - \mathbf{g}||\mathbf{1}_2 \\
&\implies T \circ \mathbf{f} - T \circ \mathbf{g} \leq \beta ||\mathbf{f} - \mathbf{g}||\mathbf{1}_2
\end{aligned}$$

Likewise by interchanging the role of $\mathbf{f}$ and bg, we have $T \circ \mathbf{g} - T \circ \mathbf{f} \leq \beta ||\mathbf{f} - \mathbf{g}||\mathbf{1}_2$. Combining the two inequalities proves $||T \circ \mathbf{f} - T \circ \mathbf{g}|| \leq \beta ||\mathbf{f} - \mathbf{g}||$. □

Lemma A3 *(Contraction Mapping)* Define the operator $T \circ \mathbf{f}(\mathbf{x}) \equiv [T_1(\mathbf{x}; \mathbf{f}) \; T_0(\mathbf{x}; \mathbf{f})]$, where

$$T_j(\mathbf{x}) \equiv u_j(\mathbf{x}) + \beta \int \max_{k \in \{0,1\}} \{f_k(\mathbf{x}') + \varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x})$$

Under *REG*, T is a contraction mapping that maps from $\mathbf{B}^2(S(\mathbf{X}))$ into $\mathbf{B}^2(S(\mathbf{X}))$.

Proof. Note $\max_{k \in \{0,1\}}\{f_k(\mathbf{x})\}$ is bounded since $\mathbf{f} \in \mathbf{B}^2(S(\mathbf{X}))$. Also:

$$\begin{aligned}
&\int \max_{k \in \{0,1\}} \{f_k(x') + \varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) \\
\leq &\int \max_{k \in \{0,1\}} \{f_k(x')\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) + \int \max_{k \in \{0,1\}} \{\varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x})
\end{aligned}$$

Both terms as well as $\mathbf{u}(\mathbf{x})$ are bounded and continuous under *REG*. Hence $T \circ (\mathbf{f}(\mathbf{x}))$ is bounded and continuous. Suppose $\mathbf{f}, \mathbf{g} \in \mathbf{B}^2(S(\mathbf{X}))$, and $\mathbf{f}(\mathbf{x}) \leq \mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in S(\mathbf{X})$. Then

$$\begin{aligned}
T_j(\mathbf{x}; \mathbf{f}) \quad &= \quad u_j(\mathbf{x}) + \beta \int \max_{k \in \{0,1\}} \{f_k(\mathbf{x}') + \varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) \\
&\leq \quad u_j(\mathbf{x}) + \beta \int \max_{k \in \{0,1\}} \{g_k(\mathbf{x}') + \varepsilon_k'\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x}) = T_j(\mathbf{x}; \mathbf{g})
\end{aligned}$$

And:

$$T_j(\mathbf{x};\mathbf{f}+a\mathbf{1}_2) = u_j(\mathbf{x}) + \beta \int \max_{k\in\{0,1\}}\{f_k(\mathbf{x}') + a + \varepsilon'_k\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$
$$= u_j(\mathbf{x}) + \beta \int \max_{k\in\{0,1\}}\{f_k(x') + \varepsilon'_k\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x}) + \beta a$$

By *Lemma A2*, the operator T is a contraction mapping. □

Proof of Lemma 1. By definition, the Bellman Equation is:

$$V(\mathbf{s}) = \max_{j\in\{0,1\}} U(\mathbf{s}, j) + \beta \int V(\mathbf{s}')dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Under *AS* and *CI*, $V(\mathbf{s}) = \max_{j\in\{0,1\}}\{\delta_j(\mathbf{x}) + \varepsilon_j\}$, where

$$\delta_j(\mathbf{x}) \equiv u_j(\mathbf{x}) + \beta \int V(\mathbf{x}', \boldsymbol{\varepsilon}')dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Substitute expression for $V(\mathbf{s})$ into the definition of $\delta_j(\mathbf{x})$ for $j \in \{0, 1\}$,

$$\delta_j(\mathbf{x}) = u_j(\mathbf{x}) + \beta \int \max_{j\in\{0,1\}}\{\delta_k(\mathbf{x}') + \varepsilon'_k\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

It follows from *Lemma A3* that under *REG*, the operator is well-defined for any $\{u, \beta, F_{\boldsymbol{\varepsilon}|\mathbf{X}}\}$, and that a fixed point $\delta(\mathbf{x})$ exists. □

Proof of Proposition 1. We need to show that any $(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ can generate observed choice probabilities $p(\mathbf{x})$ if and only if it satisfies conditions in the proposition. *(Sufficiency)* Suppose $\mathbf{u}$,$F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ satisfies the conditions in the proposition. Then for $j = 0, 1$, $\delta_j(\mathbf{x}; u_j, F_{\Delta\varepsilon|\mathbf{X}}) \equiv \omega_j(\mathbf{x}; u_j) + \xi_j(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}})$ is the unique fixed point for the following operator:

$$T_j \circ \delta_j(\mathbf{x}; u_j, F_{\Delta\varepsilon|\mathbf{X}}) = u_j(\mathbf{x}) + \beta \int \delta_j(\mathbf{x}; u_j, F_{\Delta\varepsilon|\mathbf{X}}) + \kappa_j(\mathbf{x}'; p, F_{\Delta\varepsilon|\mathbf{X}})dG_j(\mathbf{x}'|\mathbf{x})$$

By our supposition, for all $\mathbf{x} \in S(\mathbf{X})$, $\Delta\delta(\mathbf{x}) = \Delta\omega(\mathbf{x}) + \Delta\xi(\mathbf{x}) = F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$. Substitution implies $\kappa_0(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}}) = \int \max\{\Delta\delta(\mathbf{x}; \mathbf{u}, F_{\Delta\varepsilon|\mathbf{X}}) - s, 0\} dF_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x})$ and $\kappa_1(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}}) = \int \max\{s - \Delta\delta(\mathbf{x}; \mathbf{u}, F_{\Delta\varepsilon|\mathbf{X}}), 0\} dF_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x})$. Then $E(\varepsilon_j|\mathbf{x}) = 0$ for all $\mathbf{x} \in S(\mathbf{X})$ implies

$$\delta_j(\mathbf{x}; u_j, F_{\Delta\varepsilon|\mathbf{X}}) + \kappa_j(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}}) = \int \max_{k\in\{0,1\}}\{\delta_j(\mathbf{x}; u_j, F_{\Delta\varepsilon|\mathbf{X}}) + \varepsilon_j\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}|\mathbf{x}).$$

Therefore $\boldsymbol{\delta}(\mathbf{x}) \equiv [\delta_1(\mathbf{x})\ \delta_0(\mathbf{x})]'$ is the unique fixed point of the operator $T \circ \boldsymbol{\delta}(\mathbf{x}) \equiv [T_1(\mathbf{x}; \boldsymbol{\delta})\ T_0(\mathbf{x}; \boldsymbol{\delta})]$, where

$$T_j(\mathbf{x}) \equiv u_j(\mathbf{x}) + \beta \int \max_{k\in\{0,1\}}\{\delta_k(\mathbf{x}') + \varepsilon'_k\} dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}') dG_j(\mathbf{x}'|\mathbf{x})$$

Then the proof of sufficiency is completed by the supposition $\Delta\delta(\mathbf{x}) = F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$ for all $\mathbf{x} \in S(\mathbf{X})$. *(Necessity)* Now suppose $(\mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}})$ generates $p(\mathbf{x})$. This requires $\Delta\delta(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) = F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$ for all $\mathbf{x} \in S(\mathbf{X})$, where $\boldsymbol{\delta}(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) \equiv [\delta_1(\mathbf{x})\ \delta_0(\mathbf{x})]'$ is the unique fixed point of the operator T. Recursive substitution of $\boldsymbol{\delta}(\mathbf{x})$ into the definition of T suggests $\delta_j(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\varepsilon}|\mathbf{X}}) = \omega_j(\mathbf{x}; u_j) + \xi_j(\mathbf{x}; p, F_{\Delta\varepsilon|\mathbf{X}})$ for $j = 0, 1$. (See Aguirregabiria 2007 for more details.) It follows immediately $F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x}) = \Delta\omega(\mathbf{x}; \mathbf{u}) + \Delta\xi(\mathbf{x}; F_{\Delta\varepsilon|\mathbf{X}})$ for all $\mathbf{x} \in S(\mathbf{X})$. □

Proof of Corollary 1 *(Proposition 1).* Suppose $\mathbf{u}$,$F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ satisfies (3.2). Then it follows from Proposition 1 that the joint identification region is characterized by the system of linear equations

$$[\mathbf{I} + \mathbf{G}^1_\infty]\mathbf{u}_1 - [I + \mathbf{G}^0_\infty]\mathbf{u}_0 = \mathbf{C}(F_{\boldsymbol{\varepsilon}|\mathbf{X}}, p)$$

where $\mathbf{G}^j_\infty \equiv \lim_{T\to\infty} \sum_{t=1}^{T} \beta^t (G^j)$ and $\mathbf{C}(F_{\boldsymbol{\varepsilon}|\mathbf{X}}, p)$ is a K-by-1 vector of constants calculated from knowledge of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ and $p(\mathbf{x})$. Since $\mathbf{u}_0$ is normalized to a K-by-1 zero vector, $\mathbf{u}_1$ has unique solutions if and only if $\mathbf{I} + \mathbf{G}^j_\infty$ has full rank. Note $\mathbf{I} + \mathbf{G}^j_\infty = (\mathbf{I} - \beta\mathbf{G}^j)^{-1}$. Hence $\mathbf{u}_1$ has unique solutions if and only if $(\mathbf{I} - \beta\mathbf{G}^j)$ has full rank. □

Proof of Proposition 2. Suppose $\boldsymbol{\theta}_0$ is not locally identified, and there is a sequence of $\boldsymbol{\theta}_n$ in a sequence of shrinking neighborhood around $\boldsymbol{\theta}_0$ that $p(\mathbf{x}_k;\boldsymbol{\theta}_0) = p(\mathbf{x}_k;\boldsymbol{\theta}_n)$ for all $\mathbf{x}_k$. By the mean value theorem, there exists a sequence $\boldsymbol{\theta}_n^*$ on the line segment linking $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_0$ such that $\sum_l \nabla_{\boldsymbol{\theta}_l} p(\mathbf{x}_k;\boldsymbol{\theta}_n^*)d_{n,l} = 0$ for all $\mathbf{x}_k$, where l is an index for coordinates in $\boldsymbol{\theta}_0$ and $d_{n,l} \equiv \frac{\boldsymbol{\theta}_{n,l}-\boldsymbol{\theta}_{0,l}}{||\boldsymbol{\theta}_n-\boldsymbol{\theta}_0||}$. The sequence $\{d_{n,1}\}$ is an infinite sequence on the unit sphere and therefore there exists a non-zero limit $\mathbf{d}_0$. As $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$, $\mathbf{d}_n$ approaches $\mathbf{d}_0$ in the limit and we have $\sum_l \nabla_{\boldsymbol{\theta}_l} p(\mathbf{x}_k;\boldsymbol{\theta}_0)d_{0,l} = 0$ for all $\mathbf{x}_k$. This implies $\nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k;\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}} p(\mathbf{x}_k;\boldsymbol{\theta}_0)'$ is singular for all $\mathbf{x}_k$. This completes the proof. □

Proof of Corollary 2 *(Proposition 1).* Under the assumptions of the corollary, the joint identification region is given by the linear system:

$$(\mathbf{I} - \beta\mathbf{G}^1)^{-1}\mathbf{u}_1 = \mathbf{Q} - [(\mathbf{I} - \beta\mathbf{G}^1)^{-1}\beta\mathbf{G}^1\boldsymbol{\kappa}^1 - (\mathbf{I} - \beta\mathbf{G}^0)^{-1}\beta\mathbf{G}^0\boldsymbol{\kappa}^0]$$

where $\mathbf{Q}$, $\boldsymbol{\kappa}^j$ are K-by-1 vectors with $Q_k \equiv F_{\Delta\varepsilon}^{-1}(p(\mathbf{x}_k))$ and $\kappa_k^j \equiv \kappa_j(\mathbf{x}_k; F_{\Delta\varepsilon})$ for $j = 0, 1$, $k = 1, ., K$. Note that $\boldsymbol{\kappa}^0 - \boldsymbol{\kappa}^1 = \mathbf{Q}$. Substitution gives (3.7). Since $1, ., k, ., K$ is ranked in ascending order in $p(\mathbf{x}_k)$, the statistical independence restriction requires Q_k's are also in ascending order. Furthermore, note for $k \geq 2$, $\kappa_k^0 = \kappa_{k-1}^0 + \int_{Q_{k-1}}^{Q_k} F_{\Delta\varepsilon}(s)ds$, where the second term is bounded between $p(\mathbf{x}_{k-1})(Q_k - Q_{k-1})$ and $p(\mathbf{x}_k)(Q_k - Q_{k-1})$. Hence the inequality constraints must also hold. It follows that any $\mathbf{u}_1$ that keeps the feasibility of this linear system of equalities and inequalities will not be identified relative to the true parameter under $\mathcal{F}_{SI}$ (the set of $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ that satisfies statistical independence). □

Proof of Corollary 3 *(Proposition 1).* The equalities in the linear system follows from the same argument as in the proof of Corollary 3. The inequalities in the first two rows follow immediately from the definitions: $Q_k \equiv F_{\Delta\varepsilon|\mathbf{X}}^{-1}(p(\mathbf{x}_k)|\mathbf{x}_k)$, and ε_M as the median conditional on all $\mathbf{x}_k$. Those in the last two rows uses the fact that $\Delta\varepsilon$ has a support of length $2C$ and is symmetrically distributed around ε_M for all $\mathbf{x}_k$, and the definition of $\kappa_k^0 = \int_{\varepsilon_L}^{Q_k} F_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x}_k)ds$ (the area beneath neath the distribution $F_{\Delta\varepsilon|\mathbf{X}}(s|\mathbf{x}_k)$ up to Q_k). The details in algebra are omitted for brevity. $\square$