Inference in ordered response games with complete information

Andres Aradillas-Lopez
Adam Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP33/13

Inference in Ordered Response Games with Complete Information.*

Andres Aradillas-Lopez† University of Wisconsin

Adam M. Rosen‡ UCL and CeMMAP

July 26, 2013

Abstract

We study econometric models of complete information games with ordered action spaces, such as the number of store fronts operated in a market by a firm, or the daily number of flights on a city-pair offered by an airline. The model generalizes single agent models such as ordered probit and logit to a simultaneous model of ordered response. We characterize identified sets for model parameters under mild shape restrictions on agents' payoff functions. We then propose a novel inference method for a parametric version of our model based on a test statistic that embeds conditional moment inequalities implied by equilibrium behavior. Using maximal inequalities for U-processes, we show that an asymptotically valid confidence set is attained by employing an easy to compute fixed critical value, namely the appropriate quantile of a chi-square random variable. We apply our method to study capacity decisions measured as the number of stores operated by Lowe's and Home Depot in geographic markets. We demonstrate how our confidence sets for model parameters can be used to perform inference on other quantities of economic interest, such as the probability that any given outcome is an equilibrium and the propensity with which any particular outcome is selected when it is one of multiple equilibria.

Keywords: Discrete games, ordered response, partial identification, conditional moment inequalities.

JEL classification: C01, C31, C35.

*This paper has benefited from comments from seminar audiences at NYU, Berkeley, UCLA, Cornell, and the June 2013 conference on Recent Contributions to Inference on Game Theoretic Models co-sponsored by CeMMAP and the Breen Family Fund at Northwestern. We thank Eleni Aristodemou, Andrew Chesher, Jeremy Fox, Ariel Pakes, Aureo de Paula, Elie Tamer, and especially Francesca Molinari for helpful discussion and suggestions. Financial support from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and through the funding of the "Programme Evaluation for Policy Analysis" node of the UK National Centre for Research Methods, and from the European Research Council (ERC) grant ERC-2009-StG-240910-ROMETA is gratefully acknowledged.

†Address: Department of Economics, University of Wisconsin at Madison, 1180 Observatory Drive, Madison, WI 53706, United States. Email: aaradill@ssc.wisc.edu.

‡Address: Adam Rosen, Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: adam.rosen@ucl.ac.uk.

1 Introduction

This paper provides set identification and inference results for a model of simultaneous ordered response. These are settings in which multiple economic agents simultaneously choose actions from a discrete ordered action space so as to maximize their payoffs. Agents have complete information regarding each others' payoff functions, which depend on both their own and their rivals' actions. The agents' payoff-maximizing choices are observed by the econometrician, whose goal is to infer their latent payoff functions and the distribution of unobserved heterogeneity. Given the degree of heterogeneity allowed and the dependence of each agent's payoffs on their rivals' actions, the model generally admits multiple equilibria. We remain agnostic as to the selection of multiple equilibria, thus rendering the model incomplete, and its parameters set identified.

Although our model applies generally to the econometric analysis of complete information games in which players' actions are discrete and ordered, our motivation lies in application to models of firm entry. Typically, empirical models of firm entry have either allowed for only binary entry decisions, or have placed restrictions on firm heterogeneity that limit strategic interactions.[1] Yet in many contexts firms may decide not only whether to be in a market, but also how many shops or store fronts to operate. In such settings, the number of stores operated by each firm may reflect important information on firm profitability, and in particular on strategic interactions. Such information could be lost by only modeling whether the firm is present in the market and not additionally how many stores it operates. Consider, for example, a setting in which there are two firms, A and B, with (a, b) denoting the number of stores each operates in a given market. Observations of $(a, b) = (1, 3)$ or $(a, b) = (3, 1)$ are intrinsically different from observations with e.g. $(a, b) = (2, 2)$, the latter possibly reflecting less firm heterogeneity or more fierce competition relative to either of the former. Yet each of these action profiles appear identical when only firm presence is considered, as then they would all be coded as $(a, b) = (1, 1)$.

Classical single-agent ordered response models such as the ordered probit and logit have the property that, conditional on covariates, the observed outcome is weakly increasing in an unobservable profit-shifter. Our model employs shape restrictions on payoff functions, namely diminishing marginal returns in own action and increasing differences in own action and a player-specific unobservable, that deliver an analogous property for each agent. These restrictions facilitate straightforward characterization of regions of unobservable payoff shifters over which observed model outcomes are feasible. This in turn enables the transparent development of a system of conditional moment inequalities that characterize the identified set of agents' payoff functions.

When the number of actions and/or players is sufficiently large, the characterization of the identified set can comprise a computationally overwhelming number of moment inequalities. While ideally one would wish to exploit all of these moment restrictions in order to produce the sharpest

[1] See e.g. Berry and Reiss (2006) for a detailed overview of complications that arise from and methods for dealing with heterogeneity in such models.

possible set estimates, this may in some cases be infeasible. We thus also characterize outer sets that embed a subset of the full system of moment inequalities. Although less restrictive, the use of this system of inequalities can be computationally much easier for use in estimation and inference. As shown in our application such outer sets can sometimes be used to achieve economically meaningful inference.

We develop a novel approach for inference that is computationally attractive for the model at hand. Specifically, we rely on an unconditional mean-zero restriction implied by the conditional moment inequalities to develop a criterion function based approach for inference as advocated by Chernozhukov, Hong, and Tamer (2007). We show that, when evaluated at points in the identified set, our criterion function is asymptotically distributed chi-square. To construct confidence intervals for model parameters, one can thus use level sets of the criterion function with critical value given by the quantile of a chi-square random variable with appropriately specified degrees of freedom. Inference based on conditional moment inequalities is an active area of research and other possible approaches for inference include those of Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013), Ponomareva (2010), Kim (2009), Menzel (2011), Armstrong (2011a), Armstrong (2011b), and Chetverikov (2012).

We apply a parametric version of our model to study capacity decisions (number of stores) in geographic markets by Lowe's and Home Depot. We show that if pure-strategy behavior is maintained, a portion of the parameters of interest are point-identified under mild conditions. We provide point estimates for these and then apply our inference procedure to construct a confidence set for the entire parameter vector by exploiting the conditional moment inequalities implied by the model. In applications primary interest does not always rest on model parameters, but rather quantities of economic interest which can typically be written as (possibly set-valued) functions of these parameters. We illustrate in our application how our model also allows us to perform inference on such quantities, such as the likelihood that particular action profiles are equilibria, and the propensity of the underlying equilibrium selection mechanism to choose certain equilibria among multiple possibilities.

The paper proceeds as follows. In Section 1.1 we discuss the related literature on simultaneous discrete models, with particular attention to econometric models of games. In Section 2 we define the structure of the underlying complete information game and shape restrictions on payoff functions. In Section 3 we derive observable implications, including characterization of the identified set and computationally simpler outer sets. In Section 4 we provide specialized results for a parametric model of a two player game with strategic substitutes, including point identification of a subset of model parameters. In Section 5 we introduce a novel approach to inference based on conditional moment equalities and inequalities which is computationally attractive for inference on elements of the sets characterized in Section 3. In Section 6 we apply our method to model capacity (number of stores) decisions by Lowe's and Home Depot. Section 7 concludes. All proofs are provided in the Appendix.

1.1 Related Literature

We consider an econometric model of a discrete game of complete information. Our work follows the strand of literature on empirical models of entry initiated by Bresnahan and Reiss (1990, 1991a), and Berry (1992). Additional early papers on the estimation of complete information discrete games include Bjorn and Vuong (1984) and Bresnahan and Reiss (1991b). These models often allow the possibility of multiple or even no equilibria for certain realizations of unobservables, and the related issues of coherency and completeness have been considered in a number of papers, going back at least to Heckman (1978), see Chesher and Rosen (2012) for a thorough review. These issues can be and have been dealt with in a variety of different ways. Berry and Tamer (2007) discuss the difficulties these problems pose for identification in entry models, in particular with heterogeneity in firms' payoffs, and Berry and Reiss (2006) survey the various approaches that have been used to estimate such models.

The approach we take in this paper, common in the recent literature, is to abstain from imposing further restrictions simply to complete the model. Rather, we work with observable implications that may only set identify the model parameters, a technique fruitfully employed in a variety of contexts, see e.g. Manski (2003), Manski (2007), and Tamer (2010) for references to numerous examples. In the context of entry games with multiple equilibria, this tact was initially proposed by Tamer (2003), who showed how an incomplete simultaneous equations binary choice model implies a system of moment equalities and inequalities that can be used for estimation and inference. Ciliberto and Tamer (2009) apply this approach to an entry model of airline city-pairs, employing inferential methods from Chernozhukov, Hong, and Tamer (2007). Andrews, Berry, and Jia (2004) also consider a bounds approach to the estimation of entry games, based on necessary conditions for equilibrium. Pakes, Porter, Ho, and Ishii (2006) show how empirical models of games in industrial organization can generally lead to moment inequalities, and provide additional inference methods for bounds. Aradillas-López and Tamer (2008) show how weaker restrictions than Nash Equilibrium, in particular rationalizability and finite levels of rationality, can be used to set identify the parameters of discrete games. Beresteanu, Molchanov, and Molinari (2011) use techniques from random set theory to elegantly characterize the identified set of model parameters in a class of models including entry games. Galichon and Henry (2011) use optimal transportation theory to likewise achieve a characterization of the identified set applicable to discrete games. Chesher and Rosen (2012) build on concepts in both of these papers to compare identified sets obtained from alternative approaches to deal with incompleteness and in particular incoherence in simultaneous discrete outcome models.

What primarily distinguishes our work from most of the aforementioned papers is the particular focus on a simultaneous discrete model with *non-binary*, *ordered* outcomes. Simultaneous binary models are empirically relevant and have also proved an excellent expository tool in this literature.

However, as discussed above, the extension to ordered action spaces is important from a practical standpoint. Relevant examples of such outcomes include the number of store fronts a firm operates in a market, or the number of daily flights an airline offers for a particular city-pair.

Also related are a recent strand of papers on network economies faced by chain stores when setting their store location profiles, including Jia (2008), Holmes (2011), Ellickson, Houghton, and Timmins (2013), and Nishida (2012). These papers study models that allow for the measurement of payoff externalities from store location choices *across* different markets, which, like most of the aforementioned literature, our model does not incorporate. On the other hand, our model incorporates aspects that these do not, by both not imposing an equilibrium selection rule and by allowing for firm-specific unobserved heterogeneity. [2]

Some other recent papers specifically consider alternative models of ordered response with endogeneity. Davis (2006) also considers a simultaneous model with a game-theoretic foundation. He takes an alternative approach, employing enough additional structure on equilibrium selection so as to complete the model and achieve point-identification. Also related is Ishii (2005), who studies ATM networks. She uses a structural model of a multi-stage game that enables estimation of banks' revenue functions via GMM. These estimates are then used to estimate bounds for a single parameter that measures the cost of ATMs in equilibrium. Chesher (2010), provides set identification results for a single equation ordered response model with endogenous regressors and instrumental variables. Indeed, Chesher's analysis would apply if one were to use a model comprising only one of our system's simultaneous equations, and use excluded regressors from the others as instruments. Here we exploit the structure provided by the simultaneous (rather than single equation) model. Aradillas-López (2011) and Aradillas-López and Gandhi (2013) also consider simultaneous models of ordered response. In contrast to this paper, Aradillas-López (2011) focusses on nonparametric estimation of bounds on Nash outcome probabilities, and Aradillas-López and Gandhi (2013) on a model with *incomplete* information.

2 The Model

Our model consists of J economic agents or players $\mathcal{J} = \{1, ..., J\}$ who each simultaneously choose an action Y_j from the ordered action space $\mathcal{Y}_j = \{0, ..., M_j\}$. Each set $\mathcal{Y}_j$ is discrete but M_j can be arbitrarily large, possibly infinite. $Y \equiv (Y_1, ..., Y_J)'$ denotes the action profile of all J players, and for any player $j \in \mathcal{J}$ we adopt the common convention that Y_{-j} denotes the vector of

[2] Of the papers in this literature, only Ellickson, Houghton, and Timmins (2013) and Nishida (2012) also allow an ordered but non-binary within-market action space. Nishida (2012), in similar manner to Jia (2008), employs an equilibrium selection rule to circumvent the identification problems posed by multiple equilibria. We explicitly allow for multiple equilibria, without imposing restrictions on equilibrium selection. Ellickson, Houghton, and Timmins (2013) allow for multiple equilibria and partial identification, but employ a very different payoff structure. In particular, they model unobserved heterogeneity in market-level payoffs through a single scalar unobservable shared by all firms. In our model, within each market each firm has its own unobservable.

actions of j's rivals, $Y_{-j} \equiv (Y_1, ..., Y_{j-1}, Y_{j+1}, ..., Y_J)'$. As shorthand we sometimes write (Y_j, Y_{-j}) to denote an action profile Y with j^{th} component Y_j and all other components given by Y_{-j}. We use $\mathcal{Y} \equiv \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_J$ to denote the space of feasible action profiles, and for any player j, $\mathcal{Y}_{-j} \equiv \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{j-1} \times \mathcal{Y}_{j+1} \times \cdots \times \mathcal{Y}_J$ to denote the space of feasible rival action profiles.

The actions of each agent are observed across a large number n of separate environments, e.g. markets, networks, or neighborhoods, depending on the application at hand. The payoff of action Y_j for each agent j is affected by observable and unobservable payoff shifters $X_j \in \mathcal{X}_j \subseteq \mathbb{R}^{k_j}$ and $U_j \in \mathbb{R}$, respectively, as well as their rivals' actions Y_{-j}. We assume throughout that (Y, X, U) are realized on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We use $\mathbb{P}_0$ to denote the corresponding marginal distribution of observables (Y, X), and P_U to denote the marginal distribution of unobserved heterogeneity $U = (U_1, ..., U_J)'$, so that $P_U(\mathcal{U})$ denotes the probability that U is realized on the set $\mathcal{U}$. We assume throughout that U is continuously distributed with respect to Lebesgue measure with everywhere positive density on $\mathbb{R}^J$. The data comprise a random sample of observations $\{y_i, x_i : i = 1, ..., n\}$ of (Y, X) distributed $\mathbb{P}_0$, where X denotes the composite vector of observable payoff shifters X_j, $j \in J$, without repetition of any common components. The random sampling assumption guarantees identification of $\mathbb{P}_0$.[3]

For each player $j \in \mathcal{J}$ there is a payoff function $\pi_j(y, x_j, u_j)$ mapping action profile $y \in \mathcal{Y}$ and payoff shifters $(x_j, u_j) \in \mathcal{X}_j \times \mathbb{R}$ to payoffs satisfying the following restrictions.

Restriction SRP (Shape Restrictions on Payoffs): The collection of payoff functions $(\pi_1, ..., \pi_J)$ belong to a class of payoff functions $\mathbf{\Pi} = \Pi_1 \times \cdots \times \Pi_J$ such that for each $j \in \mathcal{J}$,
$\pi_j(\cdot, \cdot, \cdot) : \mathcal{Y} \times \mathcal{X}_j \times \mathbb{R} \to \mathbb{R}$ satisfies the following conditions.
(i) Payoffs are strictly concave in y_j:

$$\begin{aligned}\forall y_j \in \mathcal{Y}_j,\ & \pi_j((y_j + 1, y_{-j}), x_j, u_j) - \pi_j((y_j, y_{-j}), x_j, u_j) \\ & < \pi_j((y_j, y_{-j}), x_j, u_j) - \pi_j((y_j - 1, y_{-j}), x_j, u_j),\end{aligned}$$

where by convention $\pi_j(-1, x_j, u_j) = \pi_j(M_j + 1, x_j, u_j) = -\infty$.
(ii) For each $(y_{-j}, x) \in \mathcal{Y}_{-j} \times \mathcal{X}$, $\pi_j((y_j, y_{-j}), x, u_j)$ exhibits strictly increasing differences in (y_j, u_j), namely that if $u_j' > u_j$ and $y_j' > y_j$, then

$$\pi_j\left(\left(y_j', y_{-j}\right), x, u_j\right) - \pi_j\left(\left(y_j, y_{-j}\right), x, u_j\right) < \pi_j\left(\left(y_j', y_{-j}\right), x, u_j'\right) - \pi_j\left(\left(y_j, y_{-j}\right), x, u_j'\right). \blacksquare$$

Restriction SRP(i) imposes that marginal payoffs are decreasing in each player's own action y_j. It also implies that, given a rival pure strategy profile y_{-j}, agent j's best response correspondence is unique with probability one. Restriction SRP(ii) imposes that the payoff function exhibits

[3] We impose random sampling for simplicity and expositional ease, but our results can be generalized to less restrictive sampling schemes. For instance our identification results require that $\mathbb{P}_0$ is identified, for which random sampling is a sufficient, but not necessary, condition.

strictly increasing differences in (y_j, u_j). It plays a similar role to the monotonicity of latent utility functions in unobservables in single agent decision problems, implying that the optimal choice of y_j is weakly increasing in unobservable u_j, as in classical ordered choice models. This restriction aids in our identification analysis by guaranteeing the existence of intervals for u_j within which any y_j maximizes payoffs for any fixed (y_{-j}, x).

We focus attention on models where the distribution of unobserved heterogeneity is restricted to be independent of payoff shifters, as is common in the literature. This restriction can be relaxed, see e.g. Kline (2012), though at the cost of weakening the identifying power of the model, or requiring stronger restrictions otherwise.

Restriction I (Independence): U and X are stochastically independent, with the distribution of unobserved heterogeneity P_U belonging to some class of distributions $\mathcal{P}_U$. ■

3 Equilibrium Behavior and Observable Implications

In the model set out in Section 2 above a *structure* that generates a distribution of observables (Y, X) is a collection of payoff functions $(\pi_1, ..., \pi_J) \in \mathbf{\Pi}$ and a distribution of unobserved heterogeneity P_U. The goal of identification analysis is to deduce what structures $(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U$, and what relevant features of these structures, i.e. functionals of (π, P_U), are admitted by the model and compatible with the distribution of observables $\mathbb{P}_0$.

In order to close the model and relate structures (π, P_U) to the distribution $\mathbb{P}_0$ of (Y, X), we must additionally specify how players possessing payoff functions π play joint action profiles Y. We assume in this paper that players have complete information, and thus know the realizations of all payoff shifters (X, U) when they choose their actions. That is, there is no private information.[4] What then remains is to specify a solution concept for the underlying complete information game.

We restrict attention to Pure Strategy Nash Equilibrium (PSNE) as our solution concept to simplify the exposition. Yet our inference approach applies to other solution concepts too. This is due to the fact that for inference we exploit observable implications of PSNE that take the form of conditional moment inequalities. Observable implications of alternative solution concepts, such as rationalizability and (mixed or pure strategy) Nash Equilibrium also give rise to conditional moment inequalities, as shown for example by Aradillas-López and Tamer (2008), Aradillas-López (2011), Galichon and Henry (2011), and Beresteanu, Molchanov, and Molinari (2011), and the inference approach developed in Section 5 can also be readily applied to these alternative systems of conditional moment inequalities. Given our payoff restrictions we wish to emphasize that mixed-strategy Nash Equilibrium behavior can be readily handled through conditional moment inequalities that follow as special cases of the results in Aradillas-López (2011).

[4]For econometric analysis of incomplete information binary and ordered games see for example Aradillas-López (2010), de Paula and Tang (2012), Aradillas-López and Gandhi (2013) and the references therein.

A concern in the literature, and a motivation for considering these alternative solution concepts, is the possibility of non-existence of PSNE. However, in games where all actions are strategic complements, or in 2 player games where actions are either strategic substitutes or complements, a PSNE always exists. This follows from observing that in these cases the game is supermodular, or can be transformed into an equivalent representation as a supermodular game. This was shown for the binary outcome game by Molinari and Rosen (2008), based on the reformulation used by Vives (1999, Chapter 2.2.3) for Cournot duopoly. Tarski's Fixed Point Theorem, see e.g. Theorem 2.2 of Vives (1999) or Section 2.5 of Topkis (1998), then implies the existence of at least one PSNE. Our empirical example of Section 6 is a two player game of strategic substitutes, so the existence of PSNE is guaranteed in this context, further motivating our focus on this solution concept. Nonetheless, in other settings it is possible that no PSNE exists, in which case one could adopt an alternative solution concept and base inference on the resulting conditional moment inequalities. Alternatively, one could consider explicit approaches for dealing with non-existence, or *incoherence*, as in Chesher and Rosen (2012).

For clarity and completeness, we now formalize the restriction to PSNE behavior. To economize on notation, we define each player j's best response correspondence as

$$\boldsymbol{y}_j^* \left(y_{-j}, x_j, u_j\right) \equiv \underset{y_j \in \mathcal{Y}_j}{\arg\max} \quad \pi_j \left(\left(y_j, y_{-j}\right), x_j, u_j\right), \tag{3.1}$$

which delivers the set of payoff maximizing alternatives y_j for player j as a function of (y_{-j}, x_j, u_j).

Restriction PSNE (Pure Strategy Nash Equilibrium): With probability one, for all $j \in \mathcal{J}$, $Y_j = \boldsymbol{y}_j^* \left(Y_{-j}, X_j, U_j\right)$. ■

Strict concavity of each player j's payoff in her own action under Restriction SRP(i) guarantees that $\boldsymbol{y}_j^* \left(y_{-j}, X_j, U_j\right)$ is unique with probability one for any y_{-j}, though it does not imply that the *equilibrium* is unique. It also provides a further simplification of the conditions for PSNE, as summarized in the following Lemma.

Lemma 1 *Suppose Restriction SRP(i) holds. Then Restriction PSNE holds if and only if with probability one, for all $j \in \mathcal{J}$,*

$$\pi_j \left(Y, X_j, U_j\right) \geq \max \left\{\pi_j \left(\left(Y_j + 1, Y_{-j}\right), X_j, U_j\right), \pi_j \left(\left(Y_j - 1, Y_{-j}\right), X_j, U_j\right)\right\}, \tag{3.2}$$

where we define $\pi_j \left(\left(-1, Y_{-j}\right), X_j, U_j\right) = \pi_j \left(\left(M_j + 1, Y_{-j}\right), X_j, U_j\right) = -\infty$.

The proof of Lemma 1 is simple and thus omitted. That Restriction PSNE implies (3.2) is immediate. The other direction follows from noting that if (3.2) holds then violation of (3.1) would contradict strict concavity of $\pi_j \left(\left(y_j, Y_{-j}\right), X_j, U_j\right)$ in y_j. The import of this simple result is a rather large reduction in the number of inequalities required for characterization of PSNE, and hence the identified set of structures (π, P_U).

With these restrictions in hand, we now characterize the identified set of structures (π, P_U). Define

$$\Delta\pi_j(Y, X, U_j) \equiv \pi_j(Y, X_j, U_j) - \pi_j((Y_j - 1, Y_{-j}), X_j, U_j),$$

as the incremental payoff of action Y_j relative to $Y_j - 1$ for any (Y_{-j}, X, U_j). From Restriction SRP (ii) we have that $\Delta\pi_j(Y, X, U_j)$ is strictly increasing in U_j and thus invertible. Combining this with Lemma 1 allows us to deduce that for each player j there is for each (y_{-j}, x) an increasing sequence of non-overlapping thresholds, $\left\{u_j^*(y_j, y_{-j}, x) : y_j = 0, ..., M_{j+1}\right\}$ with

$$u_j^*(M_{j+1}, y_{-j}, x) = -u_j^*(0, y_{-j}, x) = \infty,$$

such that

$$\boldsymbol{y}_j^*(y_{-j}, x_j, u_j) = y_j \Leftrightarrow u_j^*(y_j, y_{-j}, x) < u_j \leq u_j^*(y_j + 1, y_{-j}, x). \tag{3.3}$$

That is, given (y_{-j}, x), player j's best response y_j is uniquely determined by within which of the non-overlapping intervals $\left(u_j^*(y_j, y_{-j}, x), u_j^*(y_j + 1, y_{-j}, x)\right]$ unobservable U_j falls. This holds for all j, so under Restriction PSNE each player is best responding to their rivals' actions. It follows that with probability one

$$U \in \mathcal{R}_\pi(Y, X) \equiv \underset{j \in \mathcal{J}}{\times} \left(u_j^*(Y_j, Y_{-j}, X), u_j^*(Y_j + 1, Y_{-j}, X)\right].$$

In other words, Y is an equilibrium precisely if U belongs to the rectangle $\mathcal{R}_\pi(Y, X)$. The notation makes explicit the dependence of the edges of the rectangle on the payoff functions π, through their implied threshold functions u_j^*.

We now use this result to characterize the identified set for (π, P_U). Before doing so we further define for any set $\tilde{\mathcal{Y}} \subseteq \mathcal{Y}$ and all $x \in \mathcal{X}$,

$$\mathcal{R}_\pi\left(\tilde{\mathcal{Y}}, x\right) \equiv \underset{y \in \tilde{\mathcal{Y}}}{\cup} \mathcal{R}_\pi(y, x),$$

which is the union of all rectangles $\mathcal{R}_\pi(y, x)$ across $y \in \tilde{\mathcal{Y}}$, and

$$\overline{\mathsf{R}^{\cup}}(x) \equiv \left\{\mathcal{U} \subseteq \mathbb{R}^J : \mathcal{U} = \mathcal{R}_\pi\left(\tilde{\mathcal{Y}}, x\right) \text{ for some } \tilde{\mathcal{Y}} \subseteq \mathcal{Y}\right\},$$

to be the collection of all such unions for any $x \in \mathcal{X}$.

Theorem 1 *Let Restrictions SRP, I, and PSNE hold. Then the identified set of structures is*

$$\mathcal{S}^* = \left\{(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \forall \mathcal{U} \in \mathsf{R}^{\cup}(x), \; P_U(\mathcal{U}) \geq \mathbb{P}_0[\mathcal{R}_\pi(Y, X) \subseteq \mathcal{U} | X = x] \;\; a.e. \;\; x \in \mathcal{X}\right\}, \tag{3.4}$$

where, for any $x \in \mathcal{X}$*,* $\mathsf{R}^{\cup}(x) \subseteq \overline{\mathsf{R}^{\cup}}(x)$ *denotes the collection of sets*

$$\mathsf{R}^{\cup}(x) \equiv \left\{ \mathcal{U} \subseteq \mathbb{R}^J : \begin{array}{c} \mathcal{U} = \mathcal{R}_\pi\left(\tilde{\mathcal{Y}}, x\right) \textit{ for some } \tilde{\mathcal{Y}} \subseteq \mathcal{Y} \textit{ such that } \forall \textit{ nonempty } \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2 \subseteq \mathcal{Y} \textit{ with} \\ \tilde{\mathcal{Y}}_1 \cup \tilde{\mathcal{Y}}_2 = \tilde{\mathcal{Y}} \textit{ and } \tilde{\mathcal{Y}}_1 \cap \tilde{\mathcal{Y}}_2 = \emptyset,\ P_U\left(\mathcal{R}_\pi\left(\tilde{\mathcal{Y}}_1, x\right) \cap \mathcal{R}_\pi\left(\tilde{\mathcal{Y}}_2, x\right)\right) > 0 \end{array} \right\}. \tag{3.5}$$

The above characterization is sharp. That is, the set of pairs (π, P_U) that satisfy (3.4) all satisfy the restrictions of the model and are compatible with the observed distribution of (Y, X). The characterization (3.4) makes use of results from Chesher and Rosen (2012, Theorem 5) to express the identified set as those (π, P_U) such that the random set $\mathcal{R}_\pi(Y, X)$ satisfies the conditional containment functional inequality

$$P_U(\mathcal{U}) \geq \mathbb{P}_0\left[\mathcal{R}_\pi(Y, X) \subseteq \mathcal{U} | X = x\right], \text{ a.e. } x \in \mathcal{X},$$

over the collection of sets $\mathcal{U} \in \mathsf{R}^{\cup}(x)$.[5] When $\mathcal{Y}$ is finite the identified set coincides with that of other characterizations given in the literature, e.g. Galichon and Henry (2011, Theorem 1) and Beresteanu, Molchanov, and Molinari (2011, Theorem D.2), which incorporate inequalities equivalent to those in (3.4), but over $\mathcal{U} \in \overline{\mathsf{R}^{\cup}}(x)$. The collection $\mathsf{R}^{\cup}(x)$ is a sub-collection of core-determining test sets, as defined by Galichon and Henry (2011, Theorem 1), shown to be sufficient by Chesher and Rosen (2012, Theorem 5) to imply (3.4) for all closed $\mathcal{U} \subseteq \mathbb{R}^J$. This characterization comprises fewer conditional moment inequalities while retaining sharpness.

Nonetheless, the identified set $\mathcal{S}^*$ characterized by Theorem 1 may comprise a rather large number of conditional moment inequalities, namely as many as belong to $\mathsf{R}^{\cup}(x)$, for each x. More inequality restrictions will in general produce smaller identified sets. Yet the incorporation of a very large number of inequalities may pose challenges for inference, both with regard to the quality of finite sample approximations as well as computation. As stated in the following Corollary, consideration of those structures satisfying inequality (3.4) applied to only an *arbitrary* sub-collection of those in $\overline{\mathsf{R}^{\cup}}(x)$, or indeed any arbitrary collection of sets in $\mathcal{U}$, will produce an outer region that contains the identified set.

Corollary 1 *Let* $\mathsf{U}(x) : \mathcal{X} \to 2^{\mathcal{U}}$ *map from values of* x *to collections of closed subsets of* $\mathcal{U}$*. Let*

$$\mathcal{S}^*(\mathsf{U}) = \left\{(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \forall \mathcal{U} \in \mathsf{U}(x),\ P_U(\mathcal{U}) \geq \mathbb{P}_0\left[\mathcal{R}_\pi(Y, X) \subseteq \mathcal{U} | X = x\right] \ a.e.\ x \in \mathcal{X}\right\}. \tag{3.6}$$

Then $\mathcal{S}^* \subseteq \mathcal{S}^*(\mathsf{U})$*.*

Because $\mathcal{S}^*(\mathsf{U})$ contains the identified set, it can be used to estimate valid, but potentially non-sharp bounds on functionals of (π, P_U), i.e. parameters of interest. Although $\mathcal{S}^*(\mathsf{U})$ is a

[5] Note that by definition the collection $\mathsf{R}^{\cup}(x)$ contains all sets of the form $\mathcal{R}_\pi(y, x)$ for some $y \in \mathcal{Y}$, since the requirement regarding subsets $\tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2 \subseteq \mathcal{Y}$ holds vacuously when $\tilde{\mathcal{Y}} = \{y\}$.

larger set than $\mathcal{S}^*$, its reliance on fewer inequalities can lead to significant computational gains for bound estimation and inference relative to the use of $\mathcal{S}^*$. Even in cases where the researcher wishes to estimate $\mathcal{S}^*$, it may be faster to first base estimation on $\mathcal{S}^*(\mathsf{U})$. If estimation or inference based on this outer set delivers sufficiently tight set estimates to address the empirical questions at hand, a researcher may be happy to stop here. If it does not, the researcher can refine set estimates or confidence sets based on $\mathcal{S}^*(\mathsf{U})$ by then incorporating additional restrictions, either proceeding to use $\mathcal{S}^*(\mathsf{U}')$ for some superset U' of U, or by using $\mathcal{S}^*$ itself. Typically, checking the imposed inequality restrictions involves searching over a multi-dimensional parameter space, so the computational advantage can be substantial.

In the two-player parametric model introduced in the following Section, and used in the application of Section 6, we show that a particular $\mathsf{U}(\cdot)$ is sufficient to point identify all but three of the model parameters, and we are able to achieve useful inferences based on an outer region that makes use of this and other conditional moment restrictions.

4 A Two-Player Game of Strategic Substitutes

In this section we introduce a parametric specification satisfying Restriction SRP for a two-player game with $\mathcal{J} = \{1, 2\}$. We use this specification in our empirical application, and thus focus special attention on analysis of this model. We continue to maintain Restrictions I and PSNE. In this model, existence of at least one PSNE a.e. (X,U), is guaranteed by e.g. Theorem 2.2 of Vives (1999) or Section 2.5 of Topkis (1998), as discussed in Section 3.

4.1 A Parametric Specification

For each $j \in \mathcal{J}$ we specify

$$\pi_j(Y, X_j, U_j) = Y_j \times (\delta + X_j\beta - \Delta_j Y_{-j} - \eta Y_j + U_j), \tag{4.1}$$

where we impose the restriction that $\eta > 0$ to ensure that payoffs are strictly concave in Y_j, ensuring Restriction SRP(i). Given this functional form, Restriction SRP(ii) also holds. In this specification the parameters of the two player's payoff functions differ only in the strategic interaction parameters (Δ_1, Δ_2), though this is not required for our identification analysis. We additionally impose that $\Delta_1, \Delta_2 \geq 0$, so that actions are strategic substitutes, and existence of PSNE follows as previously discussed.

Given this functional form, each player j's best response function takes the form (3.3), namely

$$\boldsymbol{y}_j^*(y_{-j}, x_j, u_j) = y_j \Leftrightarrow u_j^*(y_j, y_{-j}, x_j) < u_j \leq u_j^*(y_j + 1, y_{-j}, x_j),$$

where for $\tilde{y}_j = 0$, $u_j^*(\tilde{y}_j, y_{-j}, x_j) = -\infty$, for $\tilde{y}_j = M_j + 1$, $u_j^*(\tilde{y}_j, y_{-j}, x_j) = \infty$, and for all

$\tilde{y}_j \in \{1, ..., M_j\}$,

$$u_j^* \left(\tilde{y}_j, y_{-j}, x_j\right) \equiv \eta \left(2\tilde{y}_j - 1\right) + \Delta_j y_{-j} - \delta - x_j \beta. \tag{4.2}$$

In addition we restrict the distribution of bivariate unobserved heterogeneity U to the Farlie-Gumbel-Morgenstern (FGM) copula indexed by parameter $\lambda \in [-1, 1]$.[6] Specifically U_1 and U_2 each have the logistic marginal CDF

$$G\left(u_j\right) = \frac{\exp\left(u_j\right)}{1 + \exp\left(u_j\right)}, \tag{4.3}$$

and their joint cumulative distribution function is

$$F\left(u_1, u_2; \lambda\right) = G\left(u_1\right) \cdot G\left(u_2\right) \cdot \left[1 + \lambda \left(1 - G\left(u_1\right)\right)\left(1 - G\left(u_2\right)\right)\right]. \tag{4.4}$$

The parameter λ measures the degree of dependence between U_1 and U_2 with correlation coefficient given by $\rho = 3\lambda/\pi^2$. This copula restricts the correlation to the interval $[-0.304, 0.304]$. This is clearly a limitation, but one which appears to be reasonable in our application in Section 6. Note that ρ captures the correlation remaining after controlling for X. Thus with sufficiently many variables included in X a low "residual" correlation may be reasonable. Naturally, we could use alternative specifications, such as bivariate normal, but the closed form of $F\left(u_1, u_2; \lambda\right)$ is easy to work with and provides computational advantages. Compared to settings with a single agent ordered choice model, our framework offers a generalization of the ordered logit model, whereas multivariate normal U generalizes the ordered probit model.

For notational convenience we define $\alpha \equiv \eta - \delta$ and collect parameters into a composite parameter vector $\theta \equiv \left(\theta_1', \theta_2'\right)'$ where $\theta_1 \equiv \left(\alpha, \beta', \lambda\right)'$ and $\theta_2 = (\eta, \Delta_1, \Delta_2)'$. We show in the following Section that under fairly mild conditions the parameter subvector θ_1 is point identified, another advantage of the specification for the distribution of U given in (4.4).[7]

4.2 Observable Implications of Pure Strategy Nash Equilibrium

Given a parametric model, we re-express the sets $\mathcal{R}_\pi\left(Y, X\right)$ described in (4.5) as $\mathcal{R}_\theta\left(Y, X\right)$ in order to indicate explicitly their dependence on the finite-dimensional parameter θ. It follows from (4.2) that observed (Y, X, U) correspond to PSNE if and only if $U \in \mathcal{R}_\theta\left(Y, X\right)$ where

$$\mathcal{R}_\theta\left(Y, X\right) \equiv \left\{ U : \begin{array}{c} \eta\left(2Y_1 - 1\right) + \Delta_1 Y_2 - \delta - X_1\beta < U_1 \leq \eta\left(2Y_1 + 1\right) + \Delta_1 Y_2 - \delta - X_1\beta \\ \eta\left(2Y_2 - 1\right) + \Delta_2 Y_1 - \delta - X_2\beta < U_2 \leq \eta\left(2Y_2 + 1\right) + \Delta_2 Y_1 - \delta - X_2\beta \end{array} \right\}. \tag{4.5}$$

[6]See Farlie (1960), Gumbel (1960), and Morgenstern (1956).

[7]Results from Kline (2012) can be used to establish point identification of (α, β) under alternative distributions of unobserved heterogeneity, e.g. multivariate normal, if X is continuously distributed.

and from Theorem 1 we have the inequality

$$P_U(\mathcal{U}) \geq \mathbb{P}_0[\mathcal{R}_\theta(Y,X) \subseteq \mathcal{U}|X=x] \tag{4.6}$$

for each $\mathcal{U} \in \mathsf{R}^{\cup}(x)$, a.e. $x \in \mathcal{X}$ (see the definition of $\mathsf{R}^{\cup}(x)$ in (3.5)). However, it is straightforward to see that $Y=(0,0)$ is a PSNE if and only if

$$U \in (-\infty, \alpha - X_1\beta) \times (-\infty, \alpha - X_2\beta). \tag{4.7}$$

and that when this holds, $Y=(0,0)$ is the unique PSNE. This follows by the same reasoning as in the simultaneous binary outcome model, see for example Bresnahan and Reiss (1991a) and Tamer (2003), and this observation implies that the conditional moment inequality (4.6) using $\mathcal{U} = (-\infty, \alpha - X_1\beta) \times (-\infty, \alpha - X_2\beta)$ in fact holds with *equality*.[8]

Specifically, we have from (4.7) with this $\mathcal{U}$ that for $\tilde{\beta} \equiv (\alpha, \beta')'$, and $Z_j \equiv (1, -X_j)$,

$$\mathbb{P}_0[Y=(0,0)|X=x] = F\left(Z_1\tilde{\beta}, Z_2\tilde{\beta}; \lambda\right),$$

with $F\left(Z_1\tilde{\beta}, Z_2\tilde{\beta}; \lambda\right)$ defined in (4.4). Based on this we can construct the partial log-likelihood for the event $Y=(0,0)$ and its complement as

$$\mathcal{L}(b,\lambda) = \sum_{i=1}^{n} \ell(b,\lambda; z_i, y_i), \tag{4.8}$$

where

$$\ell(b,\lambda;z,y) \equiv 1[y=(0,0)]\log F(z_1b_1, z_2b_2;\lambda) + 1[y \neq (0,0)]\log(1-F(z_1b_1,z_2b_2;\lambda)).$$

The following theorem establishes that under suitable conditions $E[\mathcal{L}(b,\lambda)]$ is uniquely maximized at the population values for $\left(\tilde{\beta}, \lambda\right)$, which we denote $\left(\tilde{\beta}^*, \lambda^*\right)$. Thus there is point identification of the parameter subset θ_1, which is consistently estimated by the maximizer of (4.8) at the parametric rate.

Theorem 2 *For each player $j \in \{1,2\}$ let payoffs take the form (4.1), with $U \perp\!\!\!\perp X$, and let Restriction PSNE hold. Furthermore, assume that (i) for each $j \in \{1,2\}$ there exists no proper linear subspace of the support of $Z_j \equiv (1, -X_j)$ that contains Z_j with probability one, and (ii) For all conformable column vectors c_1, c_2 with $c_2 \neq 0$, we have that either $\mathbb{P}\{Z_2c_2 \leq 0|Z_1c_1 < 0\} > 0$ or $\mathbb{P}\{Z_2c_2 \geq 0|Z_1c_1 > 0\} > 0$. Then:*

[8]See also Chesher and Rosen (2012) for general conditions whereby the inequality in (4.6) can be strengthened to equality in simultaneous equations discrete outcome models.

1. *If U has known CDF F, then $\tilde{\beta}^*$ is identified. If the CDF of U is only known to belong to some class of distribution functions $\{F_\lambda : \lambda \in \Gamma\}$, then the identified set for $\left(\tilde{\beta}^*, \lambda^*\right)$ takes the form $\{(b(\lambda), \lambda) : \lambda \in \Gamma'\}$ for some function $b(\cdot) : \Gamma \to B$ and some $\Gamma' \subseteq \Gamma$.*

2. *If U has CDF $F(\cdot, \cdot; \lambda)$ given in (4.4) for some $\lambda \in [-1, 1]$, then $\left(\tilde{\beta}^*, \lambda^*\right)$ is point identified and uniquely maximizes $E[\mathcal{L}(b, \lambda)]$. Moreover,*

$$\sqrt{n}\left(\hat{\theta}_1 - \theta_1^*\right) \xrightarrow{d} \mathcal{N}\left(0, H_0^{-1}\right), \tag{4.9}$$

where

$$H_0 = E\left[\frac{\partial \ell(\theta_1; Z, Y)}{\partial \theta_1}\frac{\partial \ell(\theta_1; Z, Y)'}{\partial \theta_1}\right]. \tag{4.10}$$

Theorem 2 makes use of two conditions on the variation in X. The first condition is standard, requiring that for each j, $Z = (1, -X_j)$ is contained in no proper linear subspace with probably one. This rules out the possibility that X contains a constant component. The second condition restricts the joint distribution of Z_1 and Z_2, requiring that conditional on $Z_j c_j < 0$ (> 0), $Z_{-j}c_{-j}$ is nonpositive (nonnegative) with positive probability. This condition is automatically satisfied under well-known semiparametric large support restrictions, for example that X_j has a component X_{jk} that, conditional on all other components of X_j, has everywhere positive density on $\mathbb{R}$, with $\beta_{1k} \neq 0$. However, it is a less stringent restriction and does not require large support. For example, it immediately applies to the case where $Z_1 = Z_2$, i.e. with no player-specific payoff shifters, even if all covariates are discrete.

With these restrictions in place, the theorem provides a number of useful results. First, there is point identification of the parameters θ_1 if the distribution of unobserved heterogeneity is known. Generally econometric models only restrict the distribution of unobserved heterogeneity to be known (i.e. assumed) to belong to some set of distributions, here $\mathcal{P}_U$, indexed by $\lambda \in \Gamma$ with corresponding cumulative distributions F_λ. In this case there is, for each fixed distribution, equivalently each $\lambda \in \Gamma$, a unique $\beta = b(\lambda)$ that maximizes the expected log-likelihood when the CDF of unobserved heterogeneity is F_λ. Thus, the identified set for θ_1 belongs to the set of pairs $(b(\lambda), \lambda)$ such that $\lambda \in \Gamma$. This can simplify characterization and estimation of the identified set, since for each $\lambda \in \Gamma$ there is only one value of β to consider as a member of the identified set. Thus, for estimation, one need only scan over $\lambda \in \Gamma$ and compute the corresponding maximum likelihood estimator for each such value, rather than search over all values of $\beta \in B$. We further show that when F_λ is restricted to the FGM family, there is in fact point identification of λ^* and hence also of θ_1, which can be consistently estimated via maximum likelihood using the coarsened outcome $1[Y = (0, 0)]$. The parameter vector $\theta_2 = (\eta, \Delta_1, \Delta_2)'$ remains in general only partially identified.

5 Inference on the Full Parameter Vector

To perform inference on θ we combine the results of Theorem 2 above with conditional moment inequalities of the form in Theorem 1 and Corollary 1 over collections of test sets $\mathcal{U}$. In the following section we describe how we incorporate these moment inequalities into a test statistic for inference. In the subsequent section we then describe how we combine the use of the moment inequalities with the results of Theorem 2.[9]

5.1 Conditional Moment Inequalities

For inference we employ a statistic incorporating density-weighted versions of conditional moment inequalities, conditioning on each realization x_i observed in the sample, namely

$$T_k\left(y_i, x_i; \theta\right) \equiv E\left[m_k\left(Y, y_i, x_i; \theta\right)|X = x_i\right] \cdot f_X\left(x_i\right) \leq 0,$$

where $m_k\left(Y, X, y_i; \theta\right)$ is a "moment function" and where $f_X\left(\cdot\right)$ is the density of x. For instance, for moment inequalities of the form given by Theorem 1 and Corollary 1 we have

$$m_k\left(Y, y_i, x_i; \theta\right) = 1\left[\mathcal{R}_\theta\left(Y, x_i\right) \subseteq \mathcal{U}_k\left(x_i, y_i\right)\right] - P_U\left(\mathcal{U}_k\left(x_i, y_i\right); \theta\right), \tag{5.1}$$

where we can allow for the test sets $\mathcal{U}_k\left(x_i, y_i\right) : k = 1, ..., K$ to depend on the observations $\left(y_i, x_i\right)$. K denotes the number of conditional moment inequalities incorporated for inference.[10] For example, we may define

$$\mathsf{R}_i \equiv \mathsf{R}\left(x_i, y_i\right), \quad \mathsf{R}\left(\cdot, \cdot\right) : \mathcal{X} \times \mathcal{Y} \rightarrow 2^{\mathcal{U}},$$

where $\mathsf{R}\left(\cdot, \cdot\right)$ is a pre-specified mapping from values of $\left(x_i\, y_i\right)$ to collections of subsets of $\mathcal{U}$, with $\mathsf{R}\left(x_i, y_i\right)$ the collection of test sets incorporated in our test statistic for each observation i. Given Theorem 1 and Corollary 1 it follows that if $\mathsf{R}\left(\cdot, \cdot\right)$ is chosen such that $\mathsf{R}\left(x_i, y_i\right) = \mathsf{R}^{\cup}\left(x_i\right)$ for all i, then inference is based on the identified set, while for other choices of $\mathsf{R}\left(\cdot, \cdot\right)$ it is based on an outer set. As discussed in Section 3, the sets $\mathsf{R}^{\cup}\left(x_i\right)$ may have extremely large cardinality, rendering their use impractical. The use of other collections of test sets or moment inequalities implied by the characterization of the identified set given in Theorem 1 may in some cases be computationally advantageous.

[9]More generally, our inference approach can accommodate more than two players and alternative solution concepts, such as those allowing for mixed strategies equilibria. Though it is useful when it occurs, we do not require that a subset of the parameter vector θ be point identified, i.e. we can set $\theta_1 = \emptyset$. Rather, the essential ingredient is that the set of restrictions to be employed comprise conditional moment inequalities, in which case the approach in the following Section can be used in the construction of $\hat{R}\left(\theta\right)$ defined below.

[10]The number of inequalities used can also be allowed to vary with $\left(y_i, x_i\right)$. In this case we could write $K\left(y_i, x_i\right)$ for the number of conditional moment inequalities for $\left(y_i, x_i\right)$ and set $m_k\left(Y, y_i, x_i; \theta\right) = 0$ for each i, k with $K\left(y_i, x_i\right) < k \leq \bar{K} \equiv \max_i K\left(y_i, x_i\right)$.

Consider the function

$$R(\theta) \equiv E\left[\sum_{k=1}^{K} \left(T_k(Y, X; \theta)\right)_+\right],$$

where the expectation is taken with respect to the joint distribution of (Y, X), and $(\cdot)_+ \equiv \max\{\cdot, 0\}$. The function $R(\theta)$ is nonnegative, and positive only for θ that violate the conditional moment inequality $E[m_k(Y, y_i, x_i; \theta) | X = x_i] \leq 0$ for some $k = 1, ..., K$ with positive probability.

For the purpose of inference we employ an estimator for $R(\theta)$ that incorporates a kernel estimator for $T_k(Y, X; \theta)$, and replaces the use of the function $(\cdot)_+ = \max\{\cdot, 0\}$ with the function $\max\{\cdot, -b_n\}$ for an appropriately chosen sequence $b_n \searrow 0$. The estimator is thus of the form

$$\hat{R}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K} \hat{T}_k(y_i, x_i; \theta) \cdot 1\left[\hat{T}_k(y_i, x_i; \theta) \geq -b_n\right]\right),$$

where $1_{X_i} \equiv 1[x_i \in \mathcal{X}_i^*]$ and the estimators $\hat{T}_k(y_i, x_i; \theta)$ are defined below. The use of the sequence b_n will allow us to deal with the "kink" at zero of the function $(\cdot)_+$ while obtaining asymptotically pivotal properties for $\widehat{R}(\theta)$.

To derive the properties of our estimator, we assume that each element of X has either a discrete or absolutely continuous distribution with respect to Lebesgue measure, and we write $X = \left(X^d, X^c\right)$, where X^d denotes the discretely distributed components and X^c the continuously distributed components. Convergence rates of conditional expectations estimators will therefore depend on $z \equiv \dim(X^c)$. For kernel-weighting incorporating all components of X we define

$$\mathbf{K}(x_i - x; h) \equiv \mathbf{K}^c\left(\frac{x_i^c - x^c}{h}\right) \cdot 1\left[x_i^d = x^d\right],$$

where $\mathbf{K}^c : \mathbb{R}^z \to \mathbb{R}$ is an appropriately defined kernel function for the continuous components of X. We specify the particular properties required of this function below. The estimators $\hat{T}_k(y_i, x_i; \theta)$, $k = 1, ..., K$ that appear in $\hat{R}(\theta)$ are given by

$$\hat{T}_k(y, x; \theta) \equiv \frac{1}{nh^z}\sum_{i=1}^{n} m_k(y_i, y, x; \theta)\, \mathbf{K}(x_i - x; h) \tag{5.2}$$

Among other conditions, our approach requires that the bias of these estimators disappears uniformly at the same rate over the range of values of x_i in the data. For this purpose we restrict the summand in $\hat{R}(\theta)$ to be positive only if x_i belongs to a pre-specified "inference range" $\mathcal{X}^*$ such that the projection of $\mathcal{X}^*$ onto the continuous components of X is contained in the interior of the projection of $\mathcal{X}$ onto the continuous components of $\mathcal{X}$. In principle we could allow $\mathcal{X}^*$ to depend on n and approach $\mathcal{X}$ at an appropriate rate as $n \to \infty$. For the sake of brevity, rather than formalize this argument, we presume fixed $\mathcal{X}^*$ and state results for the convergence of $\hat{R}(\theta)$

to an appropriately re-defined $R(\theta)$:

$$R(\theta) \equiv E\left[1_X \sum_{k=1}^{K} \left(T_k(Y,X;\theta)\right)_+\right]. \tag{5.3}$$

To characterize the asymptotic behavior of $n^{1/2}\left(\hat{R}(\theta) - R(\theta)\right)$, we first impose some further restrictions. These entail smoothness restrictions, bandwidth restrictions, and conditions that guarantee manageability of relevant empirical processes. We begin with smoothness restrictions.

Restriction I1 (Smoothness): As before, let $z \equiv dim(X^c)$. For some $M \geq 2z+1$, uniformly in $(y,x) \in \text{Supp}(Y,X)$ and $\theta \in \Theta$, $f_X(x)$ and $m_k(y_i,y,x;\theta)$ are almost surely M-times continuously differentiable with respect to x_c, with bounded derivatives. ■

Our goal is to characterize sufficient conditions for $\hat{R}(\theta)$ to converge to $R(\theta)$ at rate $n^{-1/2}$. To this end we combine these smoothness restrictions with the use of bias-reducing kernels, and we require the bandwidths h_n and b_n to converge to zero at appropriate rates, as follows.

Restriction I2 (Kernels and bandwidths): K^c is a bias-reducing kernel of order M with bounded support, exhibits bounded variation, is symmetric around zero, and $\sup_{v\in\mathbb{R}^z}|K(v)| \leq \bar{K} < \infty$. The positive bandwidth sequences b_n and h_n satisfy $n^{1/2}h_n^z b_n \to \infty$, and there exists $\epsilon > 0$, such that $h_n^{-z/2}b_n n^{\epsilon} \to 0$, and $n^{1/2+\epsilon}b_n^2 \to 0$. In addition, M is large enough such that $n^{1/2+\epsilon}b_n^M \to 0$. ■

Suppose our bandwidths satisfy $h_n \propto n^{-\alpha_h}$ and $b_n \propto n^{-\alpha_b}$. Then Restriction **I2** is satisfied if α_h and α_b are chosen to satisfy

$$\alpha_h = \frac{1}{4z} - \epsilon_h,\ \alpha_b = \frac{1}{4} + \epsilon_b, \text{ with } 0 < \epsilon_h \leq \frac{1}{4z(2z+1)},\ 0 < \epsilon_b < \epsilon_h.$$

With these bandwidths, the smallest value of M compatible with Restriction **I2** is $M = 2z+1$. Combined with the smoothness Restriction **I1**, our bandwidth and kernel restrictions will be used to establish convergence of $\hat{R}(\theta)$ to $R(\theta)$ at rate $n^{-1/2}$, and asymptotically pivotal properties of $\hat{R}(\theta) - R(\theta)$, appropriately studentized.

Our next restriction, illustrated in Figure 1, imposes a condition on the behavior of each $T_k(Y,X;\theta)$. This restriction admits the possibility that $\mathbb{P}(T_k(Y,X;\theta)=0) > 0$, i.e. that any of the conditional moment inequalities are satisfied with equality with positive probability. Although inference would be simplified by ruling this out, it is important to allow the possibility of binding inequalities. Our restriction thus allows this, but requires that the density of $T_k(Y,X;\theta)$ not "blow up" in a neighborhood to the left of zero. The restriction *does* allow for $T_k(Y,X;\theta)$ to have mass points.

Restriction I3 (Behavior of $T_k(Y,X,\theta)$ at zero from below): There exist constants $\bar{b} > 0$ and $\bar{A} < \infty$ such that for all positive $b < \bar{b}$ and each $k = 1,...,K$, $\sup_{\theta\in\Theta}\mathbb{P}(-b \leq T_k(Y,X;\theta) < 0) \leq b\bar{A}$. ■

We next impose a restriction on the manageability of relevant empirical processes, with manageability as defined in Definition 7.9 of Pollard (1990). In the context of our model, unless stated otherwise, m_k is as defined in (5.1).

Restriction I4 (Manageability of Empirical Processes I): For each $k = 1, ..., K$, (i) the process

$$\mathcal{M} \equiv \{m_k (Y_i, y, x; \theta) \cdot \mathbf{K} (X_i - x; h) : (y, x, \theta) \in \text{Supp}(Y, X) \times \Theta, h > 0, 1 \leq i \leq n\}$$

is manageable with respect to the constant envelope $\bar{K} \equiv \sup_{v \in \mathbb{R}} \mathbf{K}^c (v; h)$, and (ii) there exists a $\bar{c} > 0$ such that the process

$$\mathcal{I} \equiv \{1 \{-c \leq T_k (Y_i, X_i; \theta) < 0\} : \theta \in \Theta, 0 < c < \bar{c}, 1 \leq i \leq n\},$$

is manageable with respect to the constant envelope 1. ■

Sufficient conditions for manageability, which comprise restrictions on the classes of functions allowed, are abundant in the empirical process literature. For example, if the kernel function $\mathbf{K}^c$ is of bounded variation then Lemma 22 in Nolan and Pollard (1987) and Lemmas 2.4 and 2.14 in Pakes and Pollard (1989) imply that the the class of functions $\{\mathbf{K} (x - v; h) : v \in \mathcal{X}, h > 0\}$ is Euclidean, as defined in Pakes and Pollard (1989) Definition 2.7, with respect to the constant envelope $\bar{K}$. From here, manageability of $\mathcal{M}$ follows, for example, if the classes of functions $\{g (y) = m_k (y, y', x; \theta) : (y, x, \theta) \in \mathcal{S}_{Y,X} \times \Theta\}$ are Euclidean with respect to the constant envelope 1. Sufficient conditions for this property can be found in Nolan and Pollard (1987) and Pakes and Pollard (1989), for example.

Likewise, sufficient conditions for manageability of $\mathcal{I}$ can be established, for example, if the class of sets

$$\Psi_k \equiv \{(y, x) \in \mathcal{S}_{Y,X} : -c \leq T_k (y, x; \theta) < 0, \theta \in \Theta, 0 < c < \bar{c}\}$$

have polynomial discrimination (see Pollard (1984) Definition 13) of degree at most $r < \infty$. Lemma 1 of Asparouhova, Golanski, Kasprzyk, Sherman, and Asparouhov (2002) provides a sufficient condition for this to hold, namely that the number of points at which $T_k (y, x; \cdot)$ changes sign be uniformly bounded over $(y, x) \in \mathcal{S}_{Y,X}$ and $k = 1, ..., K$.

We now establish a linear representation for $\hat{R} (\theta)$ around $R (\theta)$, which will be key to our inference approach. This result relies on two parts. First we establish the effect of the use of the threshold b_n with respect to $\hat{T}_k (Y, X; \theta)$ in $\hat{R} (\theta)$ in place of the zero threshold for $T_k (y_i, x_i, \theta)$ in $R (\theta)$. We then use a Hoeffding (1948) projection and results from Sherman (1994) to establish the asymptotically linear representation, and the corresponding "influence function" for characterizing the limiting behavior of $n^{1/2} \left(\hat{R} (\theta) - R (\theta)\right)$.

To establish the first result, define

$$\tilde{R} (\theta) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{X_i} \left(\sum_{k=1}^{K} \hat{T}_k (y_i, x_i; \theta) \cdot 1 \{T_k (y_i, x_i, \theta) \geq 0\} \right), \tag{5.4}$$

which is equivalent to $\hat{R}(\theta)$ but for the replacement of $1\left\{\hat{T}_k(y_i, x_i; \theta) \geq -b_n\right\}$ with $1\{T_k(y_i, x_i, \theta) \geq 0\}$. With the following Lemma we establish that $\tilde{R}(\theta)$ and $\hat{R}(\theta)$ are uniformly close, specifically that they differ by no more that $o_p\left(n^{-1/2}\right)$ uniformly in θ.

Lemma 2 *Let Restrictions I1-I4 hold. Then there exists $a > 1/2$ such that*

$$\sup_{\theta\in\Theta}\left|\tilde{R}(\theta) - \hat{R}(\theta)\right| = O_p\left(n^{-a}\right).$$

With Lemma 2 established, the task of producing a linear representation for $n^{1/2}\left(\hat{R}(\theta) - R(\theta)\right)$ is simplified to establishing such a representation for $n^{1/2}\left(\tilde{R}(\theta) - R(\theta)\right)$, which does not depend on the bandwidth b_n. For notational convenience let us group

$$W \equiv (X, Y).$$

With some minor algebraic manipulation of $\tilde{R}(\theta)$ defined in (5.4) and use of the definition of $\hat{T}_k(w_i; \theta)$ given in (5.2) we obtain

$$\tilde{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i} \sum_{k=1}^{K} (T_k(w_i, \theta))_+ + \sum_{k=1}^{K} \frac{1}{n^2} \sum_{i=1}^{n}\sum_{\ell=1}^{n} v_k(w_\ell, w_i; \theta, h_n), \tag{5.5}$$

where

$$v_k(w_\ell, w_i; \theta, h_n) \equiv \left(\frac{1}{h_n^z} m_k(y_\ell, y_i, x_i; \theta) \cdot \mathbf{K}(x_i - x_\ell; h_n) - T_k(w_i, \theta)\right) 1_{X_i} 1\{T_k(w_i, \theta) \geq 0\}. \tag{5.6}$$

In the proof of the following Lemma, we write the second term of (5.5) as a sum of three component terms. We use our smoothness restrictions and a result of Sherman (1994) to establish that all but one of these is uniformly $O_p(n^{-a})$ for some $a > 1/2$. For the application of Sherman's result we impose a further restriction, namely that for each k the class of functions

$$\mathcal{V}_k = \{v : v(w_1, w_2) = v_k(w_1, w_2; \theta, h), \theta \in \Theta, h > 0\},$$

is Euclidean (see Definition 2.7 in Pakes and Pollard (1989) or Definition 3 in Sherman (1994)) with respect to an envelope $\bar{V}$ such that $E\left[\bar{V}^{2+\delta}\right] < \infty$ for some $\delta > 0$.[11] Primitive conditions to establish this property can be found, e.g., in Nolan and Pollard (1987), Pakes and Pollard (1989) and Sherman (1994).

[11] More generally we only require that the U-process produced by the class $\mathcal{V}_k$ satisfy the maximal inequality in Sherman (1994), for which the Euclidean property is sufficient.

Lemma 3 *Let Restrictions I1-I4 hold and suppose that for each $k = 1, ..., K$, the class of functions $\mathcal{V}_k$ is Euclidean with envelope $\bar{V}$. Then for some $a > 1/2$, and for each $k = 1, ..., K$, uniformly in $\theta \in \Theta$,*

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{\ell=1}^{n} v_k (w_\ell, w_i; \theta, h_n) = \frac{1}{n} \sum_{i=1}^{n} [\tilde{g}_k (w_i; \theta, h_n) - E [\tilde{g}_k (W; \theta, h_n)]] + O_p \left(n^{-a}\right).$$

where $v_k (w_\ell, w_i; \theta, h_n)$ is as defined in (5.6) and where

$$\tilde{g}_k (w; \theta, h) \equiv \int (v_k (w_1, w_2; \theta, h) + v_k (w_2, w_1; \theta, h)) \, dF_W \left(w'\right).$$

Finally, with Lemmas 2 and 3 in hand, we state the resulting linear approximation for $\hat{R}(\theta)$.

Theorem 3 *Let the restrictions of Lemma 3 hold. Then for some $a > 1/2$,*

$$\hat{R}(\theta) = R(\theta) + \frac{1}{n} \sum_{i=1}^{n} \psi_R (y_i, x_i; \theta, h_n) + \xi_n (\theta), \text{ where } \sup_{\theta \in \Theta} |\xi_n (\theta)| = O_p \left(n^{-a}\right),$$

and where

$$\psi_R (y_i, x_i; \theta, h) = \sum_{k=1}^{K} \left(1_{X_i} (T_k (w_i, \theta))_+ - E \left[1_{X_i} (T_k (W, \theta))_+\right]\right) + [\tilde{g} (w_i; \theta, h) - E [\tilde{g} (W; \theta, h)]].$$

5.2 Combination of Moment Inequalities and Partial Likelihood

We now combine the linear representation for $\hat{R}(\theta)$ given by Theorem 3 with the maximum likelihood estimator described in Theorem 2 for θ_1 to perform inference on the set of parameters

$$\Theta^* \equiv \left\{ \theta \in \Theta : \begin{array}{l} \forall k = 1, ..., K, \, \forall y \in \mathcal{Y}, \, E [m_k (Y, y, X; \theta) | X = x] \leq 0 \, \wedge \\ \mathbb{P}_0 [Y = (0,0) | X = x] = F \left(Z_1 \tilde{\beta}, Z_2 \tilde{\beta}; \lambda\right), \text{ a.e. } x \in \mathcal{X}^* \end{array} \right\}.$$

As described in Section 5.1,depending on the choice of moments $E [m_k (Y, y, X; \theta) | X = x] \leq 0$, Θ^* can be either the identified set, or an outer set.

As before, let us group $w_i \equiv (y_i, x_i)$. Let $\psi_M(w_i)$ denote the MLE influence function for $\widehat{\theta}_1$. From Theorems 2 and 3 we have that uniformly over $\theta \in \Theta$, for some $\epsilon > 0$,

$$\hat{V}(\theta) \equiv n^{1/2} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{R}(\theta) \end{pmatrix} = n^{1/2} \begin{pmatrix} \theta_1^* - \theta_1 \\ R(\theta) \end{pmatrix} + \begin{pmatrix} n^{-1/2} \sum_{i=1}^{n} \psi_M (w_i) \\ n^{-1/2} \sum_{i=1}^{n} \psi_R (w_i; \theta, h_n) \end{pmatrix} + \begin{pmatrix} o_p (1) \\ o_p (n^{-\epsilon}) \end{pmatrix}. \quad (5.7)$$

For inference we use the quadratic form

$$\hat{Q}_n (\theta) \equiv \hat{V}(\theta)' \hat{\Sigma}(\theta)^{-1} \hat{V}(\theta),$$

where

$$\hat{\Sigma}(\theta) \equiv \begin{pmatrix} \hat{\Sigma}_{MM}(\theta) & \hat{\Sigma}_{MR}(\theta) \\ \hat{\Sigma}'_{MR}(\theta) & \hat{\Sigma}_{RR}(\theta) \end{pmatrix},$$

is an estimator for the variance of $\hat{V}(\theta)$. Specifically, we set

$$\begin{aligned} \hat{\Sigma}_{MM}(\theta) &\equiv \left(n^{-1}\sum_{i=1}^{n}\widehat{\psi}_M(w_i)\widehat{\psi}_M(w_i)'\right)^{-1}, \\ \hat{\Sigma}_{MR}(\theta) &\equiv n^{-1}\sum_{i=1}^{n}\widehat{\psi}_M(w_i)\hat{\psi}_R(w_i;\theta,h_n)', \\ \hat{\Sigma}_{RR}(\theta) &\equiv \max\left\{n^{-1}\sum_{i=1}^{n}\hat{\psi}_R(w_i;\theta,h_n)^2, \kappa_n\right\}, \end{aligned}$$

where $\widehat{\psi}_M(w_i)$ and $\hat{\psi}_R(w_i;\theta,h_n)$ consistently estimate $\psi_M(w_i)$ and $\psi_R(w_i;\theta,h_n)$, respectively, and where $\kappa_n \searrow 0$ is a slowly decreasing sequence of nonnegative constants such that for all $\epsilon > 0$, $n^{\epsilon}\kappa_n \to \infty$, for example $\kappa_n = (\log n)^{-1}$. This ensures that for any n, $\hat{\Sigma}_{RR}(\theta)$ is bounded away from zero. In Appendix A we show that using κ_n in this way achieves valid inference, as it guarantees that for all $\theta \in \overline{\Theta}^*$, $\widehat{\Omega}^{-1}(\theta) - \widehat{\Sigma}^{-1}(\theta)$ is positive semidefinite with probability approaching one as $n \to \infty$, where $\hat{\Omega}(\theta)$ is the same as $\hat{\Sigma}(\theta)$, but with $\widehat{\sigma}_n^2(\theta) = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_R(w_i;\theta,h_n)^2$ in place of $\hat{\Sigma}_{RR}(\theta)$.

Under appropriate regularity conditions, the quadratic form $\hat{V}(\theta)$ is asymptotically distributed χ^2 for any $\theta \in \Theta^*$, where the degrees of freedom of the asymptotic distribution depend on whether any of the K conditional moment inequalities bind with positive probability $\mathbb{P}_X$. If $\theta \in \Theta^*$ and all of the conditional moment inequalities are satisfied strictly at θ, then $n^{1/2}\hat{R}(\theta) = o_p(1)$, and $\hat{Q}_n(\theta) \xrightarrow{d} \chi^2_r$, where $r \equiv \dim(\theta_1)$. If, on the other hand, $\theta \in \Theta^*$ but at least one of the conditional moment inequalities are satisfied with equality at θ with positive probability, i.e. if θ belongs to the set

$$\bar{\Theta}^* \equiv \left\{\theta \in \Theta^* : \begin{array}{c} \mathbb{P}\{x \in \mathcal{X}^* : E[m_k(Y,y,X;\theta)|X=x] = 0\} > 0, \\ \text{for at least one } k \in \{1,...,K\} \text{ and some } y \in \mathcal{Y} \end{array}\right\},$$

then $n^{1/2}\hat{R}(\theta)$ is asymptotically normal and shows up in the asymptotic distribution of $\hat{Q}_n(\theta)$ such that $\hat{Q}_n(\theta) \xrightarrow{d} \chi^2_{r+1}$. Finally, if $\theta \notin \Theta^*$, then $\hat{Q}_n(\theta)$ "blows up", i.e. for any $c > 0$, $\Pr\left\{\hat{Q}_n(\theta) > c\right\} \to 1$ as $n \to \infty$.

Theorem 4 below uses these results to provide an asymptotically valid confidence set for θ uniformly over $\theta \in \Theta^*$. The aforementioned distributional results on which it relies are provided in Appendix A. Before stating the Theorem, we provide an additional restriction, which imposes some mild regularity conditions on the influence function $\psi_R(w_i;\theta,h_n)$ over $\theta \in \bar{\Theta}^*$. We also require that $\hat{\Sigma}(\theta)$ be within $o_p(1)$ of its population counterpart

$$\Sigma(\theta) \equiv \begin{pmatrix} \Sigma_{MM}(\theta) & \Sigma_{MR}(\theta) \\ \Sigma'_{MR}(\theta) & \Sigma_{RR}(\theta) \end{pmatrix},$$

with $\Sigma_{MM}(\theta)$, $\Sigma_{MR}(\theta)$, $\Sigma_{RR}(\theta)$ defined identically to $\hat{\Sigma}_{MM}(\theta)$, $\hat{\Sigma}_{MR}(\theta)$, $\hat{\Sigma}_{RR}(\theta)$, respectively, but with population expectations $E[\cdot]$ rather than sample means and taking the limit as $h_n \to 0$ for Σ_{MR} and Σ_{RR}.

Restriction I5 (Regularity on $\bar{\Theta}^*$): $\Sigma_{MR}(\theta)$ and $\Sigma_{RR}(\theta)$ are continuous on $\bar{\Theta}^*$ and the estimator $\hat{\Sigma}(\theta)$ is uniformly consistent on $\bar{\Theta}^*$, namely

$$\sup_{\theta\in\bar{\Theta}^*} \left\| \hat{\Sigma}(\theta) - \Sigma_n(\theta) \right\| = o_p(1).$$

In addition, the following integrability and manageability conditions hold:
(i) For some $\bar{C} < \infty$ and $\delta > 0$,

$$\lim_{n\to\infty} \sup_{\theta\in\bar{\Theta}^*} E\left[\frac{|\psi_R(w_i;\theta,h_n)|^{2+\delta}}{\sigma_n^{2+\delta}(\theta)} \right] \le \bar{C},$$

where $\sigma_n^2(\theta) \equiv var(\psi_R(w_i;\theta,h_n))$.
(ii) The triangular array of processes

$$\left\{ \psi_R(w_i;\theta,h_n) : i \le n, n \ge 1, \theta \in \bar{\Theta}^* \right\}$$

is manageable with respect to an envelope $\bar{G}$ satisfying $E\left[\bar{G}^2\right] < \infty$. ■

Theorem 4 *Let the restrictions of Lemma 3 and Restriction* ***I5*** *hold. Then the set*

$$\mathrm{CS}_{1-\alpha} \equiv \left\{ \theta \in \Theta : \hat{Q}_n(\theta) \le c_{1-\alpha} \right\},$$

where $\alpha > 0$, $c_{1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_{r+1} distribution, and $r \equiv \dim(\theta_1)$ satisfies

$$\lim_{n\to\infty} \inf_{\theta\in\Theta^*} P(\theta \in \mathrm{CS}_{1-\alpha}) \ge 1-\alpha,$$

and for all $\theta \notin \Theta^*$,

$$\lim_{n\to\infty} P(\theta \in \mathrm{CS}_{1-\alpha}) = 0.$$

The confidence set $\mathrm{CS}_{1-\alpha}$ provides correct ($\ge 1-\alpha$) asymptotic coverage for fixed P uniformly over $\theta \in \Theta^*$, and the associated test for $\theta \in \Theta^*$ is consistent against all alternatives $\theta \notin \Theta^*$. It is worth noting that our $\mathrm{CS}_{1-\alpha}$ can attain good pointwise asymptotic properties, i.e.

$$\inf_{\theta \in \Theta^*} \lim_{n \to \infty} P\left(\theta \in \mathrm{CS}_{1-\alpha}\right) \geq 1 - \alpha,$$

under weaker regularity conditions than those stated here. In particular, with Restrictions I1, I2, I3 maintained we could relax Restrictions I4 and I5, as well as the Euclidean property invoked in Lemma 3, as only sufficient conditions for the asymptotically linear representation of Theorem 3 to hold pointwise in θ would be required.

6 An Application to Home Depot and Lowe's

We apply our model to the study of the home improvement industry in the United States. According to *IBISWorld*, this industry has two dominant firms: Home Depot and Lowe's, whose market shares in 2011 were 40.8% and 32.6%, respectively. We label these two players as

Player 1: Lowe's, Player 2: Home Depot.

We take the outcome of interest $y_i = (y_{i1}, y_{i2})$ to be the number of stores operated by each firm in geographic market i. We define a market as a core based statistical area (CBSA) in the contiguous United States.[12] Our sample consists of a cross section of $n = 954$ markets in April 2012. Table 1 summarizes features of the observed distribution of outcomes.

As Table 1 shows, roughly 75 percent of markets have at most 3 stores. However, more than 10 percent of markets in the sample have 9 stores or more. If we focus on markets with asymmetries in the number of stores operated by each firm, Table 1 suggests that Lowe's tends to have more stores than Home Depot in smaller markets and viceversa. Our justification for modeling this as a static game with PSNE as our solution concept is the assumption that the outcome observed is the realization of a long-run equilibrium.[13] Because there is no natural upper bound for the number of stores each firm could open in a market, we allowed $\bar{y}_j$ to be arbitrarily large. We maintained the assumptions of mutual strategic substitutes and pure-strategy Nash equilibrium behavior with the parametrization described in Section 4.

[12]The Office of Budget and Management defines a CBSA as an area that consists of one or more counties and includes the counties containing the core urban area, as well as any adjacent counties that have a high degree of social and economic integration (as measured by commute to work) with the urban core. Metropolitan CBSAs are those with a population of 50,000 or more. Some metropolitan CBSAs with 2.5 million people or more are split into divisions. We considered all such divisions as individual markets.

[13]The relative maturity of the home improvement industry suggests that the assumption that the market is in a PSNE, commonly used in the empirical entry literature, is relatively well-suited to this application. Although, as is the case in any industry, market structure evolves over time, 82% of markets in our data exhibited no change in store configuration between March 2009 and September 2012.

Table 1: Summary of outcomes observed in the data, including average, median, and percentiles for each of Y_1 and Y_2.

	Y_1	Y_2
Average	1.68	1.97
Median	1	1
75^{th} percentile	2	1
90^{th} percentile	4	5
95^{th} percentile	7	11
99^{th} percentile	17	25
Total	1,603	1,880
$(\%Y^1 > Y^2)$	33%	
$(\%Y^1 < Y^2)$	25%	
$(\%Y^1 + Y^2 > 0)$	74%	
$(\%Y^1 + Y^2 > 0,\, Y^1 = Y^2)$	16%	

player 1: Lowe's, player 2: Home Depot.

6.1 Observable Payoff Shifters

For each market, the covariates included in X_j were: population, total payroll per capita, land area, and distance to the nearest distribution center of player j for $j = \{1, 2\}$. The first three of these were obtained from Census data. Our covariates aim to control for basic socioeconomic indicators, geographic size, and transportation costs for each firm[14]. Note that X includes 5 covariates, 3 common to each player as well as the player-specific distances to their own distribution centers. All covariates were treated as continuously distributed in our analysis.

Table 1 suggests a pattern where Home Depot operates more stores than Lowe's in larger markets. In the data we found that median market size and payroll were 50% and 18% larger, respectively, in markets where Home Depot had more store than Lowe's relative to markets where the opposite held. Overall, Home Depot opened more stores than Lowe's in markets that were larger, with higher earnings per capita. Our methodology allows us to investigate whether these types of systematic asymmetries are owed to the structure of the game, the underlying equilibrium selection mechanism, or unobserved heterogeneity.

6.2 Inference on Model Parameters

We began by computing partial maximum likelihood estimates for θ_1, corresponding to those of equation (4.9), Theorem 2. These are shown in the first column of Table 2. Given the ordinal

[14]Payroll per capita is included both as a measure of income and as an indicator of the overall state of the labor market in each CBSA. We employed alternative economic indicators such as income per household, but they proved to have less explanatory power as determinants of entry in our estimation and inference results.

nature of our action space, these point estimates indicate that within a each market, all else equal, a population increase of 100,000 has roughly the same effect on per store profit as a $45 increase in payroll per capital, a 12,300 sq mile increase in land area, or a 400 mile decrease in distance to the nearest distribution center. The second column of Table 2 shows the corresponding 95% CI based on these estimates. Figure 2 depicts the estimated partial log-likelihood for each individual parameter in a neighborhood of the corresponding estimate. Comparing their curvatures, we see that the one for ρ was relatively flatter than those of the remaining parameters. This is reflected in the rather wide MLE 95% CI for ρ. The 95% CI for the coefficients on population and land area include only positive values, while the 95% CI for the coefficient on payroll, though most positive, contains some small negative values. The MLE 90% CI for this coefficient (not reported) contained only positive values.

The last column in Table 2 provides 95% projection CIs for each parameter using the approach described in Section 5.2, using statistic $\hat{V}(\theta)$ defined in (5.7). This statistic incorporated both the moment *equalities* corresponding to likelihood contributions for the events $Y = (0,0)$ and $Y \neq (0,0)$ in the partial log-likelihood, as well as moment *inequalities* implied by the characterization in Theorem 1. Given the action space, the number of inequalities comprising the identified set would be extremely large. In the interest of computational tractability the inference approach we used incorporated conditional moment inequalities for only two moment functions (≤ 0), specifically,

$$m_1(Y_i, y, x; \theta) = 1(y \neq (0,0)) \cdot (1[\mathcal{R}_\theta(Y,X) \subseteq \mathcal{R}_\theta(y,x)] - P_U(\mathcal{R}_\theta(y,x);\theta)), \quad (6.1)$$

$$m_2(Y_i, y, x; \theta) = 1(y \neq (0,0)) \cdot (1[\mathcal{R}_\theta(Y,X) \subseteq \mathcal{R}_\theta(y,x)] - P_U(\mathcal{R}_\theta^c((0,0),x);\theta)), \quad (6.2)$$

where $\mathcal{R}_\theta^c((0,0),x)$ denotes the complement of $\mathcal{R}_\theta((0,0),x)$.[15] The term $1(y \neq (0,0))$ appears because the likelihood for the event $Y = (0,0)$ is already incorporated through the partial likelihood estimator $\hat{\theta}_1$, so the inequalities with $y = (0,0)$ would be redundant. The moment function (6.1) corresponds to using the test set $\mathcal{U}_k(x_i, y_i) = \mathcal{R}_\theta(y,x)$ in the inequalities given in Theorem 1, and (6.2) is an implication of those provided by the Theorem.

Our covariate vector X comprised five continuous random variables. We employed a multiplicative kernel $\mathbf{K}(\psi_1, \ldots, \psi_5) = \mathbf{k}(\psi_1)\mathbf{k}(\psi_2)\cdots\mathbf{k}(\psi_5)$, where each $\mathbf{k}(\cdot)$ was given by

$$\mathbf{k}(u) = \sum_{\ell=1}^{10} c_\ell \cdot \left(1-u^2\right)^{2\ell} \cdot 1\{|u| \leq 30\},$$

with $c_1, \ldots, c_{10}$ chosen such that $\mathbf{k}(\cdot)$ is a bias-reducing Biweight-type kernel of order 20. This is the same type of kernel used by Aradillas-López, Gandhi, and Quint (2013). Let $z \equiv dim(X^c) = 5$,

[15] As indicated previously, in this application the payoff functions π and the distribution of unobserved heterogeneity P_U are known functions of parameters θ. We therefore write $\mathcal{R}_\theta(Y,X)$ in place of $\mathcal{R}_\pi(Y,X)$ defined in (4.5), and $P_U(\cdot;\theta)$ in place of $P_U(\cdot)$.

and denote

$$\epsilon \equiv \frac{9}{10} \cdot \frac{1}{4z(2z+1)}, \quad \alpha_h \equiv \frac{1}{4z} - \epsilon.$$

For each element of X, the bandwidth used was of the form $h_n = c \cdot \widehat{\sigma}(X) \cdot n^{-\alpha_h}$.[16] The order of the kernel and the bandwidth convergence rate were chosen to satisfy Restriction **I2**. The constant c was set at 0.25.[17] The bandwidth b_n was set to be 0.001 at our sample size ($n = 954$). The "regularization" sequence κ_n was set below machine precision. All the results that follow were robust to moderate changes in our tuning parameters. The inference region $\mathcal{X}^*$ was set to include our entire sample, so there was no trimming used in our results. Our CS was constructed through a grid search that included over 30 million points. The computational simplicity of our approach makes a grid search of this magnitude a feasible task on a personal computer.

The third column of Table 2 presents the resulting 95% confidence intervals for each component of θ, i.e. projections given by the smallest and largest values of each parameter in our CS. Relative to the MLE CIs shown in column 2, our confidence intervals are shifted slightly and in some cases larger while in other cases smaller. In classical models where there is point identification ML estimators are asymptotically efficient, and hence produce smaller confidence intervals than those based on other estimators. The comparison here however is not so straightforward. The MLE is based only on the observation of whether each player is in or out of the market, and not the ordinal value of the outcome. The statistic we employ incorporates these likelihood equations as moment equalities and additionally some moment inequalities. That is, these inequalities constitute additional information not used in the partial log-likelihood. Furthermore, the CIs in Table 2 are projections onto individual parameter components, including parameter components for which the profile likelihood carries no information such as the interaction coefficients, $\Delta^{1,2}$ and $\Delta^{2,1}$. For all of these reasons, neither approach is expected to provide tighter CIs than the other. Reassuringly, the CIs for point-identified parameter components using either method are in all cases reasonably close to each other, yielding qualitatively similar interpretations.

One and two-dimensional graphical inspections of our CS did not reveal any holes but we are not

[16] Note that the use of a different bandwidth for each element of X is compatible with our econometric procedure. This particular choice of bandwidth is in fact equivalent to one using the same bandwidth for each component of X, but where each is first re-scaled by its standard deviation.

[17] $c = 0.25$ is approximately equal to the one that minimizes

$$AMISE = plim \left\{ \int_{-\infty}^{\infty} E\left[\left(\widehat{f}(x) - f(x) \right)^2 \right] dx \right\},$$

if we employ Silverman's "rule of thumb", Silverman (1986), using the Normal distribution as the reference distribution. In this case the constant c simplifies to

$$c = 2 \cdot \left(\frac{\pi^{1/2} \, (M!)^3 \cdot R_{\mathbf{k}}}{(2M) \cdot (2M)! \cdot \left(\mathbf{k}_M^2\right)} \right)^{\frac{1}{2M+1}}, \quad \text{where} \quad R_{\mathbf{k}} \equiv \int_{-1}^{1} \mathbf{k}^2(u) du, \, \mathbf{k}_M \equiv \int_{-1}^{1} u^M \mathbf{k}(u) du.$$

Given our choice of kernel, the solution yields $c \approx 0.25$, the value we used.

sure about the robustness of this feature for our CS as a whole given its dimension. Population, land area and distance were the only payoff shifters with coefficient estimates statistically significantly different from zero at the 5% level. The 95% CS for the correlation coefficient ρ was again wide and included zero. The payoff-concavity coefficient η was significantly positive and well above the lower bound 0.001 of our parameter space, indicating decreasing returns to scale for new stores in a market. Figure 3 depicts joint confidence regions for pairs of parameters.

Table 2: Estimates and Confidence Intervals for each Parameter

	MLE Estimate	MLE 95% CI	Moment-inequalities 95% CI†
Population (100,000)	2.219	$[0.869, 3.568]$	$[1.739, 3.850]$
Payroll per capita ($5 USD)	0.244	$[-0.023, 0.510]$	$[-0.095, 0.691]$
Land Area (1,000 sq miles)	0.180	$[0.027, 0.333]$	$[0.041, 0.415]$
Distance (100 miles)	-0.544	$[-0.929, -0.159]$	$[-1.008, -0.395]$
ρ ($Corr(\varepsilon^1, \varepsilon^2)$)	-0.050	$[-0.304, 0.204]$	$[-0.284, 0.303]$
$\delta - \eta$ (Intercept minus concavity coefficient)	-1.309	$[-2.084, -0.534]$	$[-1.970, -0.169]$
δ (Intercept)	N/A	N/A	$[-0.432, 5.581]$
η (Concavity coefficient)	N/A	N/A	$[1.038, 6.711]$
Δ^{12} (Effect of Home Depot on Lowes)	N/A	N/A	$[0, 4.047]$
Δ^{21} (Effect of Lowes on Home Depot)	N/A	N/A	$[0.622, 4.102]$

(†) Denotes the individual "projection" from the joint 95% CS obtained as described in Theorem 4.

Figure 4 depicts the joint CS for the strategic interaction coefficients, $\Delta^{1,2}$ and $\Delta^{2,1}$. Our grid search for these parameters covered the two-dimensional rectangle $[0, 16] \times [0, 16]$. As we can see, our results did not provide definitive evidence as to whether the strategic interaction effect was larger for one of the two firms, as our CS included elements above and below the 45-degree line. It did exclude, however the point $\Delta^{1,2} = \Delta^{2,1} = 0$, so we can reject the assertion that no strategic effect is present. In particular, while our CS includes $\Delta^{1,2} = 0$, it excludes $\Delta^{2,1} = 0$, leading us to reject the assertion that Lowes' decisions have no effect on Home Depot[18]. Finally, Figure 5 depicts joint confidence sets for strategic interaction coefficients and slope parameters in the model. Once again taking the ML point estimate for the coefficient on population as our benchmark, our 95% CIs on the strategic interaction coefficients from Table 2 can be used to bound the relative effect of interactions on profitability. These indicate that, again all else equal and within a given market, the effect of an additional Home Depot store on Lowe's profit is bounded above by that of a population decrease of roughly 182,000. Similarly, the effect of an additional Lowe's store on

[18]We also tried variants of our payoff form specification where strategic interaction was allowed to be a function of market characteristics, including population, population density and relative distance. In all cases our results failed to reject that strategic interaction effect is constant for each firm across markets.

Home Depot's profit is the equivalent of a population decrease of anywhere from roughly 28,000 to 185,000.

6.3 Analysis of Equilibrium Likelihood, Selection, and Counterfactual Experiments

Primary interest may not lie in the value of underlying model parameters, but rather on quantities of economic interest that can typically be expressed as (sometimes set-valued) functionals of these parameters. Equipped with a confidence set for θ, we now construct confidence regions for several such quantities, namely (i) the probability that a given outcome y is an equilibrium, (ii) the probability that a given outcome y' is an equilibrium conditional on a realized outcome $Y = y$ and covariates $X = x$, (iii) the probability that an equilibrium is selected given it is an equilibrium, and (iv) counterfactual conditional outcome probabilities generated by economically meaningful equilibrium selection rules.

6.4 Likelihood of Equilibria

Let $P_{\mathcal{E}}(y|x)$ denote the probability that y is an equilibrium outcome given $X = x$. From Lemma 1 and (3.3) we have

$$P_{\mathcal{E}}(y|x) = P_U(\mathcal{R}_\theta(y,x);\theta).$$

This relation plays a role in addressing the question: given market characteristics x and the outcome y observed in a given market, what is the probability that some other action profile y' was simultaneously an equilibrium, but not selected? We define this as $P_{\mathcal{E}}(y'|y,x)$, which, using the rules of conditional probability, is given by

$$P_{\mathcal{E}}\left(y'|y,x\right) = \frac{P_{\mathcal{E}}(y',y|x)}{P_{\mathcal{E}}(y|x)} = \frac{P_U(\mathcal{R}_\theta(y',x) \cap \mathcal{R}_\theta(y,x);\theta)}{P_U(\mathcal{R}_\theta(y,x);\theta)},$$

when $\theta = \theta_0$, where $P_{\mathcal{E}}(y',y|x)$ denotes the conditional probability that both y' and y are equilibria given $X = x$. This expression is a known function of θ, and we can construct a 95% CI for $P_{\mathcal{E}}(y'|y,x)$ by collecting the corresponding value for each $\theta \in \mathrm{CS}_{1-\alpha}$, our confidence set for θ. For the sake of illustration, Table 3 presents results using the realized outcome $y = (2,2)$ and demographics x observed in CBSA 11100 (Amarillo, TX), a metropolitan market.

Every outcome y excluded from Table 3 had *zero* probability of co-existing with $(2,2)$ as a PSNE. Notice that the lower bound in our CI was zero in each case. Overall, 13 different equilibrium outcomes y' could have simultaneously been equilibria with the observed y with positive probability. In eight cases, the probability $P_{\mathcal{E}}(y',y|x)$ could be higher than 95%. If we consider all outcomes included in Table 3 and think of them as possible counterfactual equilibria in this market, we can see that the total number of stores could have ranged between 3 and 7. The actual number of stores observed here (4) was closer to the lower bound.

We now consider the unconditional probability that any $y \in \mathcal{Y}$ is an equilibrium, denoted $P_{\mathcal{E}}(y)$. By the law of iterated expectations we can write

$$P_{\mathcal{E}}(y) = E\left[P_{\mathcal{E}}(y|Y,X)\right],$$

where the expectation is taken over Y, X. For $\theta = \theta_0$, a consistent estimator for $P_{\mathcal{E}}(y)$ is given by

$$\hat{P}_{\mathcal{E}}(y,\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} P_{\mathcal{E}}(y|y_i,x_i,\theta), \quad P_{\mathcal{E}}(y|y_i,x_i,\theta) \equiv \frac{P_U\left(\mathcal{R}_\theta(y,x_i) \cap \mathcal{R}_\theta(y_i,x_i);\theta\right)}{P_U\left(\mathcal{R}_\theta(y_i,x_i);\theta\right)}.$$

Let $\hat{\sigma}(\theta)$ denote the sample variance of $P_{\mathcal{E}}(y|Y,X,\theta)$, i.e.

$$\hat{\sigma}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(P_{\mathcal{E}}(y|y_i,x_i,\theta) - \hat{P}_{\mathcal{E}}(y,\theta)\right)^2.$$

If θ_0 were known, then by a central limit theorem, the set $\left\{\hat{P}_{\mathcal{E}}(y,\theta_0) \pm n^{-1/2} z_\alpha \hat{\sigma}(\theta_0)\right\}$ with $z_\alpha \equiv \Phi^{-1}(1-\alpha/2)$ would provide an asymptotic $1-\alpha$ CI for $P_{\mathcal{E}}(y)$. In practice θ_0 is unknown, but we can use the union of such sets over θ values in $CS_{1-\alpha}$ to construct our CI

$$\mathrm{CI}\left(P_{\mathcal{E}}(y)\right) \equiv \bigcup_{\theta \in CS_{1-\alpha}} \left\{r \in [0,1] : \hat{P}_{\mathcal{E}}(y,\theta) - n^{-1/2} z_\alpha \hat{\sigma}(\theta) \leq r \leq \hat{P}_{\mathcal{E}}(y,\theta) + n^{-1/2} z_\alpha \hat{\sigma}(\theta)\right\}. \quad (6.3)$$

If it were known (i.e. with probability one) that $\theta_0 \in CS_{1-\alpha}$, then $\mathrm{CI}\left(P_{\mathcal{E}}(y)\right)$ would contain $P_{\mathcal{E}}(y)$ with at least probability $1-\alpha$ asymptotically. With $\theta_0 \in CS_{1-\alpha}$ with probability bounded below by $1-\alpha$ asymptotically, $\mathrm{CI}\left(P_{\mathcal{E}}(y)\right)$ provides a nominal $(1-\alpha)^2$ CI.

Table 4 presents the 0.9025 ($\alpha = 0.05$) CI for $P_{\mathcal{E}}(y)$ for the ten most frequently observed outcomes in the data.

Table 3: Outcomes y that could have co-existed as equilibria with the realized outcome $(2,2)$ in CBSA 11100 (Amarillo, TX).

y	95% CI for $P_{\mathcal{E}}(y\|Y_i,X_i)$
(0, 4)	[0, 0.9997]
(6, 0)	[0, 0.9996]
(4, 1)	[0, 0.9993]
(0, 3)	[0, 0.9987]
(5, 0)	[0, 0.9975]
(1, 3)	[0, 0.9973]
(3, 1)	[0, 0.9884]

y	95% CI for $P_{\mathcal{E}}(y\|Y_i,X_i)$
(4, 0)	[0, 0.9785]
(3, 0)	[0, 0.3386]
(5, 1)	[0, 0.1306]
(0, 5)	[0, 0.0844]
(7, 0)	[0, 0.0517]
(1, 4)	[0, 0.0102]

Table 4: Outcomes y with the largest aggregate probability of being equilibria, $P_{\mathcal{E}}(y)$

y	90.25% CI for $P_{\mathcal{E}}(y)$ ($\alpha = 0.05$)	Observed frequency for y
(0, 0)	[0.2351, 0.2910]	0.2631
(1, 0)	[0.1768, 0.3213]	0.2023
(1, 1)	[0.1047, 0.1635]	0.1257
(0, 1)	[0.0998, 0.3346]	0.1205
(2, 1)	[0.0328, 0.0823]	0.0461
(1, 2)	[0.0110, 0.0824]	0.0199
(2, 0)	[0.0070, 0.1703]	0.0146
(3, 1)	[0.0062, 0.0436]	0.0136
(2, 2)	[0.0062, 0.0411]	0.0136
(3, 2)	[0.0032, 0.0320]	0.0094
(2, 3)	[0.0032, 0.0324]	0.0094
(3, 3)	[0.0025, 0.0220]	0.0083

Outcomes ordered by observed frequency.

6.4.1 Propensity of Equilibrium Selection

Our model is silent as to how any particular market outcome is selected when there are multiple equilibria. Nonetheless, a confidence set for θ can be used to ascertain some information on various measures regarding the underlying equilibrium selection mechanism $\mathcal{M}$. Consider for example the propensity that a given outcome y is selected when it is an equilibrium,

$$P_{\mathcal{M}}(y) \equiv \frac{\mathbb{P}(Y = y)}{P_{\mathcal{E}}(y)}.$$

In similar manner to the construction of the CI (6.3) for the probability that some outcome y is an equilibrium, we can construct an asymptotic $(1-\alpha)^2$ CI for $P_{\mathcal{M}}(y)$ as

$$\mathrm{CI}(P_{\mathcal{M}}(y)) \equiv \bigcup_{\theta \in CS_{1-\alpha}} \left\{ r \in [0,1] : \hat{P}_{\mathcal{M}}(y;\theta) - z_{\alpha}\hat{s}(\theta) \leq r \leq \hat{P}_{\mathcal{M}}(y;\theta) + z_{\alpha}\hat{s}(\theta) \right\},$$

where now $\hat{s}(\theta)$ consistently estimates the standard deviation of

$$\hat{P}_{\mathcal{M}}(y;\theta) \equiv \frac{\mathbb{P}(Y = y)}{P_{\mathcal{E}}(y;\theta)}.$$

Recall that $(0,0)$ cannot coexist with any other equilibrium and therefore $P_{\mathcal{M}}(0,0) = 1$. In Table 5 we present a CI for the selection propensity $P_{\mathcal{M}}(y)$ for all other outcomes listed in Table 4. In all cases in Table 5 the upper bound of our CIs was 1, so only the lower bounds of our CIs on the selection probabilities are informative.

Table 5: Aggregate propensity $P_{\mathcal{M}}(y)$ to select y when it is a NE.

y	95% CI for $P_{\mathcal{M}}(y)$	Observed frequency for y
(1, 1)	[0.8402, 1]	0.1257
(1, 0)	[0.6384, 1]	0.2023
(2, 1)	[0.5801, 1]	0.0461
(0, 1)	[0.3382, 1]	0.1205
(3, 3)	[0.3187, 1]	0.0083
(2, 2)	[0.2690, 1]	0.0136
(3, 1)	[0.2470, 1]	0.0136
(3, 2)	[0.2089, 1]	0.0094
(2, 3)	[0.2049, 1]	0.0094
(1, 2)	[0.1863, 1]	0.0199
(2, 0)	[0.0495, 1]	0.0146

Outcomes ranked by the CI lower bound.

We can also make direct comparisons of the selection propensities $P_{\mathcal{M}}(y)$ across particular profiles. Figure 6 makes such comparisons by plotting $\hat{P}_{\mathcal{M}}(y;\theta)$ for each $\theta \in CS_{1-\alpha}$. With the exception of θ yielding selection propensities very close to one for both outcomes considered in each graph, which our analysis does not rule out, the comparisons in parts (A)-(C) of Figure 6 can be summarized as follows:

(1) *Equilibria with at most one store by each firm:* We compare the propensity of equilibrium selection for the outcomes $(0,1)$, $(1,0)$ and $(1,1)$. Our results yield two findings: (i) Comparing equilibria where only one store is opened, there is a higher selection propensity for Lowe's to have the only store than for Home Depot. (ii) There is a greater selection propensity for the equilibrium in which both firms operate one store than those where only one firm does.

(2) *Equilibria with a monopolist opening multiple stores:* We focus on the outcomes $(0,2)$, $(2,0)$, $(0,3)$ and $(3,0)$. Our results indicate that the selection propensity is higher for the outcome in which Lowe's operates two stores than those where Home Depot operates two stores. Our findings regarding selection propensities for $(0,3)$ and $(3,0)$ were less conclusive.

(3) *Equilibria where both firms enter with the same number of stores:* We focus on the outcomes $(1,1)$, $(2,2)$ and $(3,3)$. Although not illustrated in the figure, the propensity to select symmetric equilibria where both firms are present appeared to be comparably higher than the propensity to select equilibria where there is only one firm in the market. For most $\theta \in CS_{1-\alpha}$, the outcome $(1,1)$ was the most favored.

Without a structural model, the observed frequencies alone are not informative about selection propensities. For example, even though $(1,1)$ was observed in only 12.5% of markets while $(1,0)$ was observed in 20.2% of them, our results show that, except for some $\theta \in CS_{1-\alpha}$ with both selection propensities close to one, the selection propensity for $(1,1)$ when it was an equilibrium was higher than that of $(1,0)$. The fact that the latter is observed more frequently simply indicates that payoff realizations where $(1,1)$ is an equilibrium occurred relatively rarely.

6.4.2 Some counterfactual experiments

As explained above, our framework allows us to study the likelihood that other outcomes could have co-existed as equilibria along with the outcomes actually observed in each market in the data. With this information at hand we can do counterfactual analysis based on pre-specified (by us) equilibrium selection mechanisms. Here we generate counterfactual outcomes in each market based on four hypothetical equilibrium selection rules. We focus our analysis on those markets where at least one firm entered and each firm opened at most 15 stores.[19] This accounts for approximately 70% of the entire sample.

(A) Selection rule favoring Lowe's.- For each market i, a counterfactual outcome $y_i^c \equiv \left(y_i^{1,c}, y_i^{2,c}\right)$ was generated through the following steps:

1.– Find all the outcomes y for which

$$\overline{P}_{\mathcal{E}}(y|Y_i, X_i) = \max\left\{\frac{P_{\mathcal{E}}(y, Y_i, X_i|\theta)}{P_{\mathcal{E}}(Y_i, X_i|\theta)} \,:\, \theta \in CS_{1-\alpha}\right\}$$

(the upper bound within our CS for the probability of co-existing with Y_i as NE) was at least 95%. If there are no such outcomes, then set $y_i^c = Y_i$. Otherwise proceed to step 2.

2.– Choose the outcome y with the largest number of Lowe's stores. If there are ties, choose the one with the largest number of Home Depot stores.

(B) Selection rule favoring Home Depot.- Same as (A), but switching the roles of Home Depot and Lowe's.

(C) Selection rule favoring entry by both firms and largest total number of stores.- Here we took the following steps to determine y_i^c:

1.– As in (A) and (B), look for all the outcomes y for which $\overline{P}_{\mathcal{E}}(y|Y_i, X_i) \geq 0.95$. If no such $y \neq Y_i$ exists, set $y_i^c = Y_i$. Do the same if no y was found where *both* firms enter. Otherwise proceed to step 2.

[19] Recall again that observing $(0,0)$ in a given market implies that no other counterfactual equilibrium was possible.

2.– Among the outcomes y found in step 1, look for the one that maximizes the total number of stores $y^1 + y^2$. If there are ties, then choose the one that minimizes $|y^1 - y^2|$. If more than one such outcome exists, choose randomly among them using uniform probabilities.

(D) Selection rule favoring symmetry.- Each y_i^c was generated as follows:

1.– As in (A)-(C), look for all the outcomes y for which $\overline{P}_{\mathcal{E}}(y|Y_i, X_i) \geq 0.95$. If no such $y \neq Y_i$ exists, set $y_i^c = Y_i$. Otherwise proceed to step 2.

2.– Among the outcomes y found in step 1, look for the one that minimizes $|y^1 - y^2|$. If more than one such outcomes exist, choose randomly among them using uniform probabilities.

Table 6: Results of counterfactual equilibrium selection experiments

	Observed data†		Selection rules							
			(A)		(B)		(C)		(D)	
	Y^1	Y^2	$y^{1,c}$	$y^{2,c}$	$y^{1,c}$	$y^{2,c}$	$y^{1,c}$	$y^{2,c}$	$y^{1,c}$	$y^{2,c}$
Average	1.76	1.62	4.34	0.48	0.41	3.09	3.08	1.11	1.74	1.78
Median	1	1	1	0	0	1	1	1	1	1
75^{th} percentile	2	1	4	0	0	3	2	1	1	1
90^{th} percentile	4	4	12	1	1	8	9	2	4	4
95^{th} percentile	6	7	20	1	1	15	17	2	8	8
Total	1,180	1,090	2,917	326	275	2,078	2,072	746	1,172	1,197
$\% \left(y^1 > y^2\right)$	47%		60%		15%		42%		26%	
$\% \left(y^1 = y^2\right)$	23%		26%		27%		40%		51%	

(†) The markets considered in this experiment where those where at least one firm entered and each firm opened at most 15 stores. This included approx. 70% of the entire sample.

Examining Table 6, the pattern of market outcomes that results from counterfactual selection rules (A), (B) and (C) is decisively different from the features of the observed outcomes in the data. This is less so for selection rule (D). Table 6 also suggests that a selection mechanism which maximizes the total number of stores in each market (rule (C)) would produce a pattern of outcomes heavily biased in favor of Lowe's. Overall, among these counterfactual experiments, the one employing selection rule (d) favoring symmetry most closely matches the observed pattern of store profiles in the data.

7 Conclusion

In this paper we have analyzed a simultaneous equations model for a complete information game in which agents' actions are ordered. This generalized the well-known simultaneous binary outcome model used for models of firm entry to cases where firms take ordered rather than binary actions, for example the number of store fronts to operate in a market, or the number of daily routes offered on a city pair by an airline.

We applied recently-developed methods from the literature to characterize (sharp) identified sets for model structures via conditional moment inequalities under easily interpreted shape restrictions. While one may ideally wish to incorporate all of the identifying information delivered by the model in performing estimation and inference, the number of implied conditional moment inequalities can be rather large, potentially posing significant challenges for both computation and the quality of asymptotic approximations in finite samples. However, the structure of this characterization lends itself readily to outer regions for model parameters, also characterized by conditional moment inequalities, which may be easier to use for estimation and inference. We further showed that in a parametric two player instance of our model, we achieve point identification of all but 3 parameters under fairly mild conditions, without using large support restrictions.

We proposed a novel method for inference based on a test statistic that employed density-weighted kernel estimators of conditional moments, summing over measured deviations of conditional moment inequalities. We used results from the behavior of U-processes to show that our test statistic behaves asymptotically as a chi-square random variable when evaluated at points in the identified set, with degrees of freedom dependent upon whether the conditional moments are binding with positive probability. This was then used to construct confidence sets for parameters, where the critical value employed is simply a quantile of a chi-square distribution with the appropriate degrees of freedom.

We applied our inference approach to data on the number of stores operated by Lowe's and Home Depot in different markets. We presented confidence sets for model parameters, and showed how these confidence sets could in turn be used to construct confidence intervals for other quantities of economic interest, such as equilibrium selection probabilities and the probability that counterfactual outcomes are equilibria jointly with observed outcomes in a given market. Our inference approach can be applied much more generally to models that comprise conditional moment inequalities, with or without identification of a subvector of parameters. Although we focused on Pure Strategy Nash Equilibrium as a solution concept, this was not essential to our inference method. It could alternatively be based on conditional moment inequalities implied by (mixed or pure strategy) Nash Equilibrium, or other solution concepts, such as rationalizability.

Moreover, in our application we only employed a small subset of the conditional moment inequalities comprising the (sharp) identified set. In principle our approach can be applied to sharp characterizations too, but the number of inequalities these incorporate can be rather large. The shear number of inequalities raises interesting questions regarding computational feasibility and the accuracy of asymptotic approximations in finite sample, both for our inference approach and others in the literature. Future research on these issues thus seems warranted.

References

ANDREWS, D. W. K., S. T. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," working paper, Yale University.

ANDREWS, D. W. K., AND X. SHI (2013): "Inference for Parameters Defined by Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

ARADILLAS-LÓPEZ, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157(2), 409–431.

——— (2011): "Nonparametric probability bounds for Nash equilibrium actions in a simultaneous discrete game," *Quantitative Economics*, 2, 135–171.

ARADILLAS-LÓPEZ, A., AND A. GANDHI (2013): "Robust Inference of Strategic Interactions in Static Games," working paper, University of Wisconsin.

ARADILLAS-LÓPEZ, A., A. GANDHI, AND D. QUINT (2013): "Identification and Inference in Ascending Auctions with Correlated Private Values," *Econometrica*, 81, 489–534.

ARADILLAS-LÓPEZ, A., AND E. TAMER (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business and Economic Statistics*, 26(3), 261–310.

ARMSTRONG, T. B. (2011a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," Working paper. Yale University.

——— (2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," Working paper. Stanford University.

ASPAROUHOVA, E., R. GOLANSKI, K. KASPRZYK, R. SHERMAN, AND T. ASPAROUHOV (2002): "Rank Estimators for a Transformation Model," *Econometric Theory*, 18, 1099–1120.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

BERRY, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

BERRY, S., AND P. REISS (2006): "Empirical Models of Entry and Market Structure," in *The Handbook of Industrial Organization*, ed. by M. Armstrong, and P. Rob, vol. 3, pp. 1845–1886. North-Holland.

Berry, S., and E. Tamer (2007): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume II*, ed. by R. Blundell, W. Newey, and T. Persson, pp. 46–85. Cambridge University Press.

Bjorn, P., and Q. Vuong (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," CIT working paper, SSWP 537.

Bresnahan, T. F., and P. J. Reiss (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

——— (1991a): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48(1-2), 57–81.

——— (1991b): "Entry and Competition in Concetrated Markets," *Journal of Political Economy*, 99(5), 977–1009.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2013): "Intersection Bounds, Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78(2), 575–601.

Chesher, A., and A. Rosen (2012): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification," CeMMAP working paper CWP21/12.

Chetverikov, D. (2012): "Adaptive Test of Conditional Moment Inequalities," Working paper. MIT.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Davis, P. (2006): "Estimation of Quantity Games in the Presence of Indivisibilities and Heterogeneous Firms," *Journal of Econometrics*, 134, 187–214.

de Paula, A., and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80(1), 143–172.

Ellickson, P. B., S. Houghton, and C. Timmins (2013): "Estimating Network Economies in Retail Chains: A Revealed Preference Approach," *Rand Journal of Economics*, 44(2), 169–193.

Farlie, D. (1960): "The Performance of Some Correlation Coefficients for a General Bivariate Distribution," *Biometrika*, 47, 307–323.

Galichon, A., and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78(4), 1264–1298.

Gumbel, E. (1960): "Bivariate Exponential Distributions," *Journal of the American Statistical Association*, 55, 698–707.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics*, 19, 293–325.

Holmes, T. J. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79(1), 253–302.

Ishii, J. (2005): "Interconnection Pricing, Compatibility, and Investment in Network Industries: An Empirical Study of ATM Surcharging in the Retail Banking Industry," working paper, Harvard University.

Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263–1316.

Kim, K.-I. (2009): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," Working Paper. University of Minnesota.

Kline, B. (2012): "Identification of Complete Information Games," Working Paper. University of Texas at Austin.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172(1), 14–32.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA.

Menzel, K. (2011): "Consistent estimation with many moment inequalities," Working Paper. New York University.

Molinari, F., and A. M. Rosen (2008): “The Identification Power of Equilibrium in Games: The Supermodular Case, a Comment on Aradillas-Lopez and Tamer (2008).,” *Journal of Business and Economic Statistics*, 26(3), 297–302.

Morgenstern, D. (1956): “Einfache Beispiele zweidimensionaler Verteilungen,” *Mathematische Statistik*, 8, 234–235.

Nishida, M. (2012): “Estimating a Model of Strategic Network Choice,” Working Paper, Johns Hopkins University.

Nolan, D., and D. Pollard (1987): “U-Processes: Rates of Convergence,” *Annals of Statistics*, 15, 780–799.

Pakes, A., and D. Pollard (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57(5), 1027–1057.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): “The Method of Moments with Inequality Constraints,” working paper, Harvard University.

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer Verlag.

——— (1990): *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.

Ponomareva, M. (2010): “Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates,” Working Paper. University of Western Ontario.

Romano, J. (2004): “On non-parametric testing, the uniform behaviour of the t-test and related problems,” *Scandinavian Journal of Statistics*, 31, 567–584.

Sherman, R. (1994): “Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators,” *Annals of Statistics*, 22, 439–459.

Silverman, B. (1986): *Density estimation for statistics and data analysis*. Chapman and Hall/CRC.

Tamer, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70(1), 147–167.

——— (2010): “Partial Identification in Economics,” *Annual Review of Economics*, 2, 167–195.

Topkis, D. M. (1998): *Supermodularity and Complementarity*. Princeton University Press.

Vives, X. (1999): *Oligopoly Pricing*. MIT Press.

A Proofs

In this section we provide proofs for the results stated in the main text and auxiliary lemmas.

A.1 Proofs of Results from the Main Text

Proof of Theorem 1. Let $\mathcal{R}_\pi(Y, X)$ be the rectangles described in (4.5). It follows from Theorem 1 of Chesher and Rosen (2012) that the identified set is given by

$$\mathcal{S}^* = \left\{(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \forall \mathcal{U} \in \mathsf{F}\left(\mathbb{R}^J\right), \, P_U(\mathcal{U}) \geq \mathbb{P}_0\left[\mathcal{R}_\pi(Y, X) \subseteq \mathcal{U} | X = x\right] \text{ a.e. } x \in \mathcal{X}\right\}, \quad \text{(A.1)}$$

where $\mathsf{F}\left(\mathbb{R}^J\right)$ denotes all closed sets in $\mathbb{R}^J$. This is equivalent to the characterizations of Galichon and Henry (2011, Theorem 1) and Beresteanu, Molchanov, and Molinari (2011, Theorem D.2) applicable with finite $\mathcal{Y}$, specifically

$$\mathcal{S}^* = \left\{ \begin{array}{c} (\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \\ \forall \mathcal{C} \in 2^{\mathcal{Y}}, \, P_U\left(\exists y \in \mathcal{C} : y \in \text{PSNE}(\pi, X, U) | X = x\right) \geq \mathbb{P}_0\left[Y \in \mathcal{C} | X = x\right] \text{ a.e. } x \in \mathcal{X} \end{array} \right\},$$

where $\text{PSNE}(\pi, X, U)$ denotes the set of PSNE when the payoff functions are π for the given (X, U). It follows from Chesher and Rosen (2012, Theorem 5) that (A.1) can be refined by replacing $\mathsf{F}\left(\mathbb{R}^J\right)$ with the sub-collection $\mathsf{R}^{\cup}(x)$. ■

Proof of Corollary 1. This follows from the observation that for any $x \in \mathcal{X}$,

$$\forall \mathcal{U} \in \mathsf{R}^{\cup}(x), \, P_U(\mathcal{U}) \geq \mathbb{P}_0\left[\mathcal{R}_\pi(Y, X) \subseteq \mathcal{U} | X = x\right] \quad \text{(A.2)}$$

implies that the same inequality holds for all $\mathcal{U} \in \overline{\mathsf{R}^{\cup}}(x)$, and in particular for all $\mathcal{U} \in \mathsf{U}(x)$. ■

Proof of Theorem 2. We prove parts 1 and 2 in the statement of the Theorem in separate steps.

Step 1. Suppose that F is known and define the sets

$$S_b^+ \equiv \left\{z : z_1(b_1 - \beta_1^*) > 0 \wedge z_2(b_2 - \beta_2^*) \geq 0\right\},$$

$$S_b^- \equiv \left\{z : z_1(b_1 - \beta_1^*) < 0 \wedge z_2(b_2 - \beta_2^*) \leq 0\right\}.$$

For any $z \in S_b^+$ we have that

$$F(z_1 b_1, z_2 b_2) > F(z_1 \beta_1^*, z_2 \beta_2^*) = \mathbb{P}\left\{Y = (0,0) | z\right\},$$

and likewise for any $z \in S_b^-$,

$$F(z_1 b_1, z_2 b_2) < F(z_1 \beta_1^*, z_2 \beta_2^*) = \mathbb{P}\left\{Y = (0,0) | z\right\}.$$

The probability that $Z \in S_b \equiv S_b^+ \cup S_b^-$ is

$$\begin{aligned}\mathbb{P}\{Z \in S_b\} &= \mathbb{P}\left\{Z \in S_b^+\right\} + \mathbb{P}\left\{Z \in S_b^-\right\} \\ &= \left(\begin{array}{c}\mathbb{P}\{Z_2(b_2-\beta_2^*) \geq 0 | Z_1(b_1-\beta_1^*) > 0\}\,\mathbb{P}\{Z_1(b_1-\beta_1^*) > 0\} \\ +\mathbb{P}\{Z_2(b_2-\beta_2^*) \leq 0 | Z_1(b_1-\beta_1^*) < 0\}\,\mathbb{P}\{Z_1(b_1-\beta_1^*) < 0\}\end{array}\right).\end{aligned}$$

Both $\mathbb{P}\{Z_1(b_1-\beta_1^*) > 0\}$ and $\mathbb{P}\{Z_1(b_1-\beta_1^*) < 0\}$ are strictly positive by (i), and at least one of $\mathbb{P}\{Z_2(b_2-\beta_2^*) \geq 0 | Z_1(b_1-\beta_1^*) > 0\}$ and $\mathbb{P}\{Z_2(b_2-\beta_2^*) \leq 0 | Z_1(b_1-\beta_1^*) < 0\}$ must be strictly positive by (ii). Therefore $\mathbb{P}\{Z \in S_b\} > 0$, implying that with λ^* known b is observationally distinct from β^* since for each $z \in S_b$, $\mathbb{P}\{Y = (0,0)|z\} \neq F(z_1 b_1, z_2 b_2)$.

If instead F is only known to belong to some class of distribution functions $\{F_\lambda : \lambda \in \Gamma\}$, the above reasoning implies that for each $\lambda \in \Gamma$, $E[\mathcal{L}(b,\lambda)]$ is uniquely maximized with respect to b. Then the conclusion of the first claim of the Theorem follows letting $b^*(\lambda)$ denote the maximizer of $E[\mathcal{L}(b,\lambda)]$ for any $\lambda \in \Gamma$.

Step 2. Suppose now that U has CDF $F(\cdot,\cdot;\lambda)$ of the form given in (4.4) for some $\lambda \in [-1,1]$.

To show that λ^* is identified, consider the expectation of the profiled log-likelihood:

$$\mathcal{L}_0(\lambda) \equiv E[\mathcal{L}(b^*(\lambda),\lambda)] = E[\ell(b^*(\lambda),\lambda;Z,Y)].$$

Note that because $\left(\tilde{\beta}^*, \lambda^*\right)$ maximizes $E[\mathcal{L}(b,\lambda)]$ with respect to (b,λ), it follows that λ^* maximizes $\mathcal{L}_0(\lambda) = \max_b E[\mathcal{L}(b,\lambda)]$. That λ^* is the unique maximizer of $\mathcal{L}_0(\lambda)$, and thus point-identified, follows from strict concavity of $\mathcal{L}_0(\lambda)$ in λ, shown in Lemma 4.

A standard mean value theorem expansion for maximum likelihood estimation then gives

$$\hat{\theta}_1 = \theta_1^* + \frac{1}{n}\sum_{i=1}^{n} \psi_M(y_i, x_i) + o_p\left(n^{-1/2}\right),$$

where

$$\psi_M(y_i, x_i) \equiv H_0^{-1} \frac{\partial \ell(\theta_1; z_i, y_i)}{\partial \theta_1}$$

is the maximum likelihood influence function satisfying

$$n^{-1/2} \sum_{i=1}^{n} \psi_M(y_i, x_i) \to \mathcal{N}\left(0, H_0^{-1}\right),$$

with H_0 as defined in (4.10). ■

Proof of Lemma 2. As in the main text, to simplify notation let $w \equiv (x,y)$ with support denoted $\mathcal{W}$. We abbreviate $\hat{T}_k(w_i;\theta)$ and $T_k(w_i;\theta)$ for $\hat{T}_k(y_i,x_i;\theta)$ and $T_k(y_i,x_i;\theta)$, respectively, $k = 1,...,K$. Suprema with respect to w,θ are to be understood to be taken with respect to $\mathcal{W} \times \Theta$

unless otherwise stated. Let

$$\begin{aligned}
\xi_n(\theta) &\equiv \hat{R}(\theta) - \tilde{R}(\theta) \\
&= \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K}\hat{T}_k(w_i;\theta)\cdot\left(1\left\{\hat{T}_k(w_i;\theta)\geq -b_n\right\} - 1\left\{T(w_i;\theta)\geq 0\right\}\right)\right).
\end{aligned}$$

Note that

$$|\xi_n(\theta)| \leq \xi_n^1(\theta) + \xi_n^2(\theta),$$

where

$$\begin{aligned}
\xi_n^1(\theta) &\equiv \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K}\left|\hat{T}_k(w_i;\theta)\right|\cdot 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right), \\
\xi_n^2(\theta) &\equiv \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K}\left|\hat{T}_k(w_i;\theta)\right|\cdot 1\left\{\left|\hat{T}_k(w_i;\theta) - T_k(w_i,\theta)\right| \geq b_n\right\}\right).
\end{aligned}$$

To complete the proof, we now show that each of these terms is $O_p(n^{-a})$ uniformly over $\theta \in \Theta$ for some $a > 1/2$.

Step 1 (Bound on $\left|\xi_n^1(\theta)\right|$).

We have

$$\begin{aligned}
\sup_\theta \left|\xi_n^1(\theta)\right| &\leq \sup_\theta \left\{\begin{array}{c} \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K}\left|T_k(w_i,\theta)\right|\cdot 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right) \\ +\frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K}\left|\hat{T}_k(w_i,\theta) - T_k(w_i,\theta)\right|\cdot 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right) \end{array}\right\} \\
&\leq \left(2b_n + \sup_{w,\theta}\left|\hat{T}_k(w_i,\theta) - T_k(w_i,\theta)\right|\right)\times \sup_\theta \left|\frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\left(\sum_{k=1}^{K} 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right)\right| \\
&= \left(2b_n + O_p\left(\frac{\log n}{\sqrt{nh_n^z}}\right)\right)\times \sup_{k,\theta}\left|\frac{1}{n}\sum_{i=1}^{n} 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right|, \qquad \text{(A.3)}
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second from elementary algebra, and the third from

$$\sup_{w,\theta}\left|\hat{T}_k(w_i,\theta) - T_k(w_i,\theta)\right| = O_p\left(\frac{\log n}{\sqrt{nh_n^z}}\right)$$

for all $k = 1, ..., K$ holding under I2 and I4. Now for $\bar{b}$ and $\bar{A}$ as defined in Restriction I3 we have for all $k = 1, ..., K$ and large enough n, $2b_n \leq \bar{b}$ and therefore

$$\sup_{\theta} E\left[1\left\{-2b_n \leq T_k(W,\theta) < 0\right\}\right] \leq 2\bar{A}b_n, \tag{A.4}$$

$$\bar{\Omega}_n \equiv \sup_{\theta} Var\left[1\left\{-2b_n \leq T_k(W,\theta) < 0\right\}\right] \leq 2\bar{A}b_n. \tag{A.5}$$

It now follows from the triangle inequality and (A.4) above that for any $k = 1, ..., K$,

$$\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n} 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right|$$
$$\leq \sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\} - E\left[-2b_n \leq T_k(w_i,\theta) < 0\right]\right)\right| + 2\bar{A}b_n.$$

The manageability Restriction I4 implies,

$$\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\} - E\left[-2b_n \leq T_k(w_i,\theta) < 0\right]\right)\right| = O_p\left(\frac{\sqrt{\bar{\Omega}_n}}{n}\right),$$

which is in fact $O_p\left(\sqrt{\frac{b_n}{n}}\right)$ by virtue of (A.5). Thus

$$\begin{aligned}
\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n} 1\left\{-2b_n \leq T_k(w_i,\theta) < 0\right\}\right| &= O_p\left(\sqrt{\frac{b_n}{n}}\right) + O(b_n) \\
&= b_n\left(O_p\left(\frac{1}{\sqrt{b_n n}}\right) + O(1)\right) \\
&= b_n\left(o_p(1) + O(1)\right) \\
&= O_p(b_n).
\end{aligned}$$

Plugging this into (A.3) we have

$$\begin{aligned}
\sup_{\theta}\left|\xi_n^1(\theta)\right| &= \left(2b_n + O_p\left(\frac{\log n}{\sqrt{nh_n^z}}\right)\right) \cdot O_p(b_n) \\
&= O_p\left(b_n^2\right) + O_p\left(\frac{b_n \log n}{\sqrt{nh_n^z}}\right) \\
&= O_p\left(n^{-a}\right)
\end{aligned}$$

for some $a > 1/2$ by the bandwidth conditions in Restriction I2.

Step 2 (Bound on $P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq c\right\}$).

From Restriction I4 and application of Theorem 3.5 and equation (7.3) of Pollard (1990), there exist positive constants κ_1, κ_2 such that for any $c > 0$, and any $\mathcal{U} \in$

$$P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq c\right\} \leq \kappa_1 \exp\left(-\left(nh_n^d\kappa_2 c\right)^2\right). \tag{A.6}$$

Our smoothness restriction I1 and an M^{th} order expansion imply the existence of a constant C such that

$$\sup_{w,\theta}\left|E\left[\hat{T}_k(w;\theta)\right] - T_k(w;\theta)\right| \leq Ch_n^M. \tag{A.7}$$

Thus

$$\begin{aligned}
& P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq b_n\right\} & \text{(A.8)} \\
\leq\ & P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta,\mathcal{U}) - E\left[\hat{T}_k(w;\theta)\right]\right| + \sup_{w,\theta}\left|E\left[\hat{T}_k(w;\theta,\mathcal{U})\right] - T_k(w;\theta)\right| \geq b_n\right\} & \\
\leq\ & P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta,\mathcal{U}) - T_k(w;\theta)\right| \geq b_n - Ch_n^M\right\}, &
\end{aligned}$$

where the first inequality follows by the triangle inequality and the second by (A.7). Under our bandwidth restrictions I2 we have for large enough n that $b_n > Ch_n^M$, and so application of (A.6) to (A.8) with $c = b_n - Ch_n^M$ gives

$$P\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq c\right\} \leq \kappa_1 \exp\left(-\left(nh_n^d\kappa_2\left(b_n - Ch_n^M\right)\right)^2\right). \tag{A.9}$$

Step 3 (Bound on $\left|\xi_n^2(\theta)\right|$).

We have

$$\begin{aligned}
\sup_{\theta}\left|\xi_n^2(\theta)\right| &\leq \sup_{w,\theta}\left|\hat{T}_k(w;\theta)\right| \cdot \sup_{w,\theta} 1\left\{\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq b_n\right\} \\
&= O_p(1) \times \sup_{w,\theta} 1\left\{\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq b_n\right\} \\
&= O_p(1) \times 1\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq b_n\right\}.
\end{aligned}$$

Let $\mathcal{D}_n \equiv 1\left\{\sup_{w,\theta}\left|\hat{T}_k(w;\theta) - T_k(w;\theta)\right| \geq b_n\right\}$. Now using Chebyshev's inequality we have

$$\left|\mathcal{D}_n - E\left[\mathcal{D}_n\right]\right| = O_p\left(\sqrt{var\left(\mathcal{D}_n\right)}\right) = O_p\left(\sqrt{E\left[\mathcal{D}_n\right]\left(1 - E\left[\mathcal{D}_n\right]\right)}\right) = \sqrt{E\left[\mathcal{D}_n\right]}O_p(1).$$

Therefore

$$\mathcal{D}_n \leq \sqrt{E\left[\mathcal{D}_n\right]} \cdot O_p\left(1\right) + E\left[\mathcal{D}_n\right] = \sqrt{E\left[\mathcal{D}_n\right]}\left(O_p\left(1\right) + \sqrt{E\left[\mathcal{D}_n\right]}\right) = \sqrt{E\left[\mathcal{D}_n\right]}O_p\left(1\right). \qquad \text{(A.10)}$$

From (A.9) in Step 2 we have

$$E\left[\mathcal{D}_n\right] \leq \kappa_1 \exp\left(-\left(nh_n^d\kappa_2\left(b_n - Ch_n^M\right)\right)^2\right),$$

which combined with (A.10) gives

$$\mathcal{D}_n = O_p\left(\sqrt{\kappa_1}\exp\left(-\frac{1}{2}\left(nh_n^d\kappa_2\left(b_n - Ch_n^M\right)\right)^2\right)\right),$$

from which it follows that $\mathcal{D}_n = O_p\left(n^{-a}\right)$, completing the proof. ■

Proof of Lemma 3. We prove the lemma for $K = 1$ and drop the subscript k notation for convenience. This suffices for the claim of the lemma since K is finite, and a finite sum of $O_p\left(n^{-a}\right)$ terms is $O_p\left(n^{-a}\right)$. Define

$$\begin{aligned} g\left(w_1, w_2; \theta, h\right) &\equiv v\left(w_1, w_2; \theta, h_n\right) + v\left(w_2, w_1; \theta, h_n\right), \\ \tilde{g}\left(w; \theta, h\right) &\equiv \textstyle\int g\left(w, w'; \theta, h\right) dF_W\left(w'\right), \quad \mu\left(\theta, h\right) \equiv \textstyle\int \tilde{g}\left(w; \theta, h\right) dF_W\left(w\right), \\ \tilde{v}\left(w_1, w_2; \theta, h\right) &\equiv g\left(w_1, w_2; \theta, h\right) - \tilde{g}\left(w_1; \theta, h\right) - \tilde{g}\left(w_2; \theta, h\right) + \mu\left(\theta, h\right). \end{aligned}$$

A Hoeffding (1948) decomposition of our U-process, making use of the relation $E\left[\tilde{g}\left(W; \theta, h\right)\right] = \mu\left(\theta, h\right)$, gives

$$\begin{aligned} &\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\ell=1}^{n} v_k\left(w_\ell, w_i; \theta, h_n\right) \\ &= \mu\left(\theta, h\right) + \frac{1}{n}\sum_{i=1}^{n}\left[\tilde{g}\left(w_i; \theta, h\right) - E\left[\tilde{g}\left(W; \theta, h\right)\right]\right] + \frac{1}{n\left(n-1\right)}\sum_{1\leq i<\ell\leq n}\tilde{v}\left(w_i, w_\ell; \theta, h\right) + o_p\left(n^{-1}\right). \end{aligned}$$

The third term above is a degenerate U-process of order 2. By Corollary 4 in Sherman (1994),

$$\sup_{\theta\in\Theta}\frac{1}{n\left(n-1\right)}\sum_{1\leq i<\ell\leq n}\tilde{v}\left(w_i, w_\ell; \theta, h\right) = O_p\left(nh_n^{-z}\right) = o_p\left(n^{-1/2-\epsilon}\right),$$

where the last equality follows from Restriction I2. Note that securing the above rate is the sole motivation for imposing that the class $\mathcal{V}$ be Euclidean. Any alternative restriction that could deliver this result would suffice.

Under the smoothness Restriction I1, using iterated expectations and an M^{th} order approximation,

$$\sup_{\theta\in\Theta} |\mu(\theta,h)| = Ch_n^M = O_p\left(n^{-1/2-\epsilon}\right),$$

for some $\epsilon > 0$. Thus

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\ell=1}^{n} v_k(w_\ell, w_i;\theta,h_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\tilde{g}(w_i;\theta,h) - E\left[\tilde{g}(W;\theta,h)\right]\right] + O_p\left(n^{1/2+\epsilon}\right). \blacksquare$$

Proof of Theorem 3. Let

$$\Delta_{g,i}(\theta,h) \equiv \sum_{k=1}^{K}\left[\tilde{g}_k(w_i;\theta,h) - E\left[\tilde{g}_k(W;\theta,h)\right]\right].$$

Combining Lemma 2 with the definition of $\tilde{R}(\theta)$ in (5.4) we have for some $a > 1/2$,

$$\begin{aligned}
\hat{R}(\theta) &= \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\sum_{k=1}^{K}(T_k(w_i,\theta))_+ + \sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\ell=1}^{n} v_k(w_\ell,w_i;\theta,h_n) + O_p\left(n^{-a}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n} 1_{X_i}\sum_{k=1}^{K}(T_k(w_i,\theta))_+ + \frac{1}{n}\sum_{i=1}^{n}\Delta_{g,i}(\theta,h_n) + O_p\left(n^{-a}\right)\\
&= R(\theta) + \frac{1}{n}\sum_{i=1}^{n}\left(1_{X_i}\sum_{k=1}^{K}(T_k(w_i,\theta))_+ - R(\theta)\right) + \frac{1}{n}\sum_{i=1}^{n}\Delta_{g,i}(\theta,h_n) + O_p\left(n^{-a}\right)\\
&= R(\theta) + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K} 1_{X_i}(T_k(w_i,\theta))_+ - E\left[1_{X_i}(T_k(w_i,\theta))_+\right]\right) + \frac{1}{n}\sum_{i=1}^{n}\Delta_{g,i}(\theta,h_n) + O_p\left(n^{-a}\right),
\end{aligned}$$

where the second line follows from Lemma 3, the third adding subtracting $R(\theta)$, and the fourth substituting for $R(\theta)$ using (5.3) and interchanging summation and expectation. $\blacksquare$

Proof of Theorem 4. We characterize the limiting behavior of $\hat{Q}_n(\theta) = \hat{V}(\theta)\hat{\Sigma}(\theta)^{-1}\hat{V}(\theta)$. From Theorems 2 and 3 we have from (5.7) that uniformly over $\theta \in \Theta$,

$$\hat{V}(\theta) = n^{1/2}\begin{pmatrix}\theta_1^* - \theta_1\\ R(\theta)\end{pmatrix} + \begin{pmatrix} n^{-1/2}\sum_{i=1}^{n}\psi_M(w_i)\\ n^{-1/2}\sum_{i=1}^{n}\psi_R(w_i;\theta,h_n)\end{pmatrix} + \begin{pmatrix} o_p(1)\\ o_p(n^{-\epsilon})\end{pmatrix}, \tag{A.11}$$

where $\epsilon > 0$. We consider each of the three cases (i) $\theta \in \Theta^*/\bar{\Theta}^*$, (ii) $\theta \in \bar{\Theta}^*$, and (iii) $\theta \notin \Theta^*$, which together prove the Theorem.

Case (i), $\theta \in \Theta^*/\bar{\Theta}^*$: Because $\theta \in \Theta^*$, $\theta_1^* - \theta_1 = 0$ and $R(\theta) = 0$. By definition of $\bar{\Theta}^*$, we have that

$$\inf_{\theta\in\Theta^*/\bar{\Theta}^*} P_W\left(\max_{k=1,\ldots,K} T_k(W,\theta) < 0\right) = 1.$$

It follows from the definition of $\psi_R(w_i;\theta,h_n)$ that $n^{-1/2}\sum_{i=1}^{n}\psi_R(w_i;\theta,h_n) = 0$ wp$\to$ 1 for all

$\theta \in \Theta^*/\bar{\Theta}^*$. Therefore

$$\hat{Q}_n(\theta) = n^{-1} \sum_{i=1}^{n} \psi_M(w_i) \hat{H}_0^{-1} \sum_{i=1}^{n} \psi_M(w_i) + o_p(1),$$

uniformly over $\theta \in \Theta^*/\bar{\Theta}^*$. Then by Theorem 2, (4.9), for any $c > 0$, and any sequence $\theta_n \in \Theta^*/\bar{\Theta}^*$

$$\lim_{n\to\infty} P\left(\hat{Q}_n(\theta_n) \leq c\right) = P\left(\chi_r^2 \leq c\right).$$

Case (ii): $\theta \in \bar{\Theta}^*$. Again, $\theta \in \Theta^*$ so $\theta_1^* - \theta_1 = 0$ and $R(\theta) = 0$. Let

$$\Omega(\theta) \equiv \begin{pmatrix} \Sigma_{MM}(\theta) & \Sigma_{MR}(\theta) \\ \Sigma'_{MR}(\theta) & \sigma^2(\theta) \end{pmatrix}, \quad \hat{\Omega}(\theta) \equiv \begin{pmatrix} \hat{\Sigma}_{MM}(\theta) & \hat{\Sigma}_{MR}(\theta) \\ \hat{\Sigma}'_{MR}(\theta) & \widehat{\sigma}_n^2(\theta) \end{pmatrix}$$

where

$$\sigma^2(\theta) \equiv \lim_{n\to\infty} \sigma_n^2(\theta), \quad \widehat{\sigma}_n^2(\theta) \equiv n^{-1} \sum_{i=1}^{n} \hat{\psi}_R(w_i; \theta, h_n)^2.$$

We assume $\Omega(\theta)$ to be well-defined and invertible at each $\theta \in \bar{\Theta}^*$. Part (i) of Restriction I5 suffices for a Lindeberg condition to hold, see Lemma 1 of Romano (2004). It allows for the limiting variance of $\hat{\psi}_R$ to become arbitrarily close to zero on $\theta \in \Theta^*$, but essentially dictates that its absolute expectation vanish faster. Combined with the manageability condition of Restriction I5 (ii), it follows that for any sequence $\theta_n \in \Theta^*$ such that $\sigma_n^2(\theta)$ has a well-defined limit,

$$n^{-1/2} \sum_{i=1}^{n} \frac{\hat{\psi}_R(w_i; \theta_n, h_n)}{\sigma_n(\theta_n)} \to \mathcal{N}(0, 1).$$

For a given $\theta \in \bar{\Theta}^*$, let

$$\breve{Q}_n(\theta) \equiv \hat{V}(\theta) \hat{\Omega}_n(\theta)^{-1} \hat{V}(\theta).$$

By construction $\hat{\Omega}^{-1}(\theta) - \hat{\Sigma}^{-1}(\theta)$ is positive semidefinite and therefore $\hat{Q}_n(\theta) \leq \breve{Q}_n(\theta)$ for all $\theta \in \bar{\Theta}^*$.

Now let θ_n be a sequence in Θ^*. Since Θ is compact, the sequence θ_n is bounded and has a convergent subsequence θ_{a_n}. By the continuity conditions in Restriction I5, $\hat{\Omega}_{a_n}^{-1}(\theta_{a_n})$ exists and has a well-defined limit. For any $c > 0$, parts (i) and (ii) of Restriction I5 then yield

$$\lim_{n\to\infty} P\left(\breve{Q}_n(\theta_{a_n}) \leq c\right) = P\left(\chi_{r+1}^2 \leq c\right),$$

and since for all $\theta \in \bar{\Theta}^*$ $\hat{Q}_n(\theta) \leq \breve{Q}_n(\theta)$,

$$\lim_{n\to\infty} P\left(\hat{Q}_n(\theta_{a_n}) \leq c\right) \geq P\left(\chi_{r+1}^2 \leq c\right).$$

To analyze the behavior of

$$\lim_{n\to\infty} \inf_{\theta\in\Theta^*} P\left(\hat{Q}_n\left(\theta_{a_n}\right) \leq c\right)$$

now choose a sequence $\theta_n \in \Theta^*$ such that for some $\delta_n \searrow 0$,

$$\left| P\left(\hat{Q}_n\left(\theta_n\right) \leq c\right) - \inf_{\theta\in\Theta^*} P\left(\hat{Q}_n\left(\theta\right) \leq c\right) \right| \leq \delta_n.$$

Note that we can always find such a sequence. Using Theorem 3, Restriction I5, and $P\left(\chi_r^2 \leq c\right) \geq P\left(\chi_{r+1}^2 \leq c\right)$, our previous arguments show that we can always find a subsequence θ_{a_n} such that

$$\lim_{n\to\infty} P\left(\hat{Q}_n\left(\theta_{a_n}\right) \leq c\right) \geq P\left(\chi_{r+1}^2 \leq c\right),$$

and from here we conclude that

$$\lim \inf_{\theta\in\Theta^*} P\left(\hat{Q}_n\left(\theta\right) \leq c\right) \geq P\left(\chi_{r+1}^2 \leq c\right),$$

which proves the first assertion of the Theorem.

Case (iii): $\theta \notin \Theta^*$. Now either $\theta_1^* - \theta_1 \neq 0$ or $R\left(\theta\right) \neq 0$, or both. It follows from (A.11) that for any $c > 0$,

$$\lim_{n\to\infty} P\left(\hat{V}\left(\theta\right) \leq c\right) = 0,$$

and therefore

$$\lim_{n\to\infty} P\left(\hat{Q}_n\left(\theta\right) \leq c\right) = 0,$$

completing the proof. ■

A.2 Auxiliary Lemmas and Proofs

Lemma 4 *Let the conditions of Theorem 2 hold and assume that U has CDF $F\left(\cdot,\cdot;\lambda^*\right)$ given in (4.4) for some $\lambda^* \in [-1,1]$. Then $\mathcal{L}_0\left(\lambda\right)$ defined in the proof of Theorem 2 is strictly concave in λ.*

Proof. By definition, for any $\lambda \in [-1,1]$, $\beta^*\left(\lambda\right)$ satisfies the first and second order necessary conditions:

$$\frac{\partial \mathcal{L}_0\left(\beta^*\left(\lambda\right),\lambda\right)}{\partial\beta} = 0, \quad \frac{\partial^2 \mathcal{L}_0\left(\beta^*\left(\lambda\right),\lambda\right)}{\partial\beta\partial\beta'} \leq 0, \tag{A.12}$$

where ≤ 0 denotes non-positive definiteness. These conditions require, respectively,

$$g\left(\lambda,\beta\right) \equiv E\left[m_1\left(\lambda,z\right)\frac{dF^*}{\partial\beta}\right] = 0, \tag{A.13}$$

at $\beta = \beta^*(\lambda)$, and

$$E\left[m_2(\lambda, z)\frac{\partial F^*}{\partial \beta}\frac{\partial F^*}{\partial \beta'} + m_1(\lambda, z)\frac{\partial^2 F^*}{\partial \beta \partial \beta'}\right] \leq 0, \tag{A.14}$$

where for ease of notation $p(z) \equiv \mathbb{P}_0[Y = (0,0)|Z = z]$, and for any parameter μ,

$$\frac{\partial F^*}{\partial \mu} \equiv \frac{dF(z_1\beta, z_2\beta; \lambda)}{d\mu}, \text{ evaluated at } \beta = \beta^*(\lambda),$$

$$m_1(\lambda, z) \equiv p(z) F(z_1\beta^*(\lambda), z_2\beta^*(\lambda); \lambda)^{-1} - (1 - p(z))(1 - F(z_1\beta^*(\lambda), z_2\beta^*(\lambda); \lambda))^{-1},$$

$$m_2(\lambda, z) \equiv -p(z) F(z_1\beta^*(\lambda), z_2\beta^*(\lambda); \lambda)^{-2} - (1 - p(z))(1 - F(z_1\beta^*(\lambda), z_2\beta^*(\lambda); \lambda))^{-2} < 0.$$

We now use these conditions to show concavity of $\mathcal{L}_0(\lambda)$. Using (A.13), equivalently the envelope theorem, we have that

$$\frac{\partial \mathcal{L}_0(\lambda)}{\partial \lambda} = E\left[\frac{\partial \mathcal{L}(\beta^*(\lambda), \lambda; z)}{\partial \lambda}\right].$$

The second derivative with respect to λ is

$$\frac{\partial^2 \mathcal{L}_0(\lambda)}{\partial \lambda^2} = E\left[\frac{\partial^2 \mathcal{L}(\beta^*(\lambda), \lambda; z)}{\partial \lambda \partial \beta'}\frac{\partial \beta^*(\lambda)}{\partial \lambda} + \frac{\partial^2 \mathcal{L}(\beta^*(\lambda), \lambda; z)}{\partial \lambda^2}\right] \tag{A.15}$$

We now proceed to solve for each term in (A.15).

To solve for $\dfrac{\partial \beta^*(\lambda)}{\partial \lambda}$ we apply the implicit function theorem to (A.13), obtaining

$$\frac{\partial \beta^*(\lambda)}{\partial \lambda} = \frac{\partial g}{\partial \beta}^{-1}\frac{\partial g}{\partial \lambda} = E\left[m_2\frac{\partial F^*}{\partial \beta}\frac{\partial F^*}{\partial \beta'} + m_1\frac{\partial^2 F^*}{\partial \beta \partial \beta'}\right]^{-1} E\left[m_2\frac{\partial F^*}{\partial \beta}\frac{\partial F^*}{\partial \lambda} + m_1\frac{\partial^2 F^*}{\partial \beta \partial \lambda}\right].$$

In addition we have

$$\frac{\partial^2 \mathcal{L}(\beta^*(\lambda), \lambda; z)}{\partial \lambda \partial \beta'} = m_2\frac{\partial F^*}{\partial \lambda}\frac{\partial F^*}{\partial \beta'} + m_1\frac{\partial^2 F^*(z, \lambda)}{\partial \lambda \partial \beta}, \tag{A.16}$$

and

$$\frac{\partial^2 \mathcal{L}(\beta^*(\lambda), \lambda; z)}{\partial \lambda^2} = m_2\frac{\partial^2 F^*}{\partial \lambda^2} = 0.$$

Putting these expressions together in (A.15) gives,

$$\frac{\partial^2 \mathcal{L}_0(\lambda)}{\partial \lambda^2} = AB^{-1}A' + D, \tag{A.17}$$

where

$$A = E\left[m_2\frac{\partial F^*}{\partial \lambda}\frac{\partial F^*}{\partial \beta'} + m_1\frac{\partial^2 F^*}{\partial \lambda \partial \beta'}\right], \qquad B = E\left[m_2\frac{\partial F^*}{\partial \beta}\frac{\partial F^*}{\partial \beta'} + m_1\frac{\partial^2 F^*}{\partial \beta \partial \beta'}\right],$$

and

$$D = E\left[m_2\left(\frac{\partial F^*}{\partial \lambda}\right)^2\right].$$

By (A.14), B is negative semi-definite so that $AB^{-1}A' \leq 0$. Then $m_2 \leq 0$ implies that

$$\frac{\partial^2 \mathcal{L}_0(\lambda)}{\partial \lambda^2} \leq 0,$$

and strictness of the inequality follows from $F(\cdot, \cdot; \lambda) \in (0,1)$. Therefore $\mathcal{L}_0(\lambda)$ is strictly concave in λ and consequently β^* and λ^* are point identified. ■

Figure 1: Illustration of Restriction I3

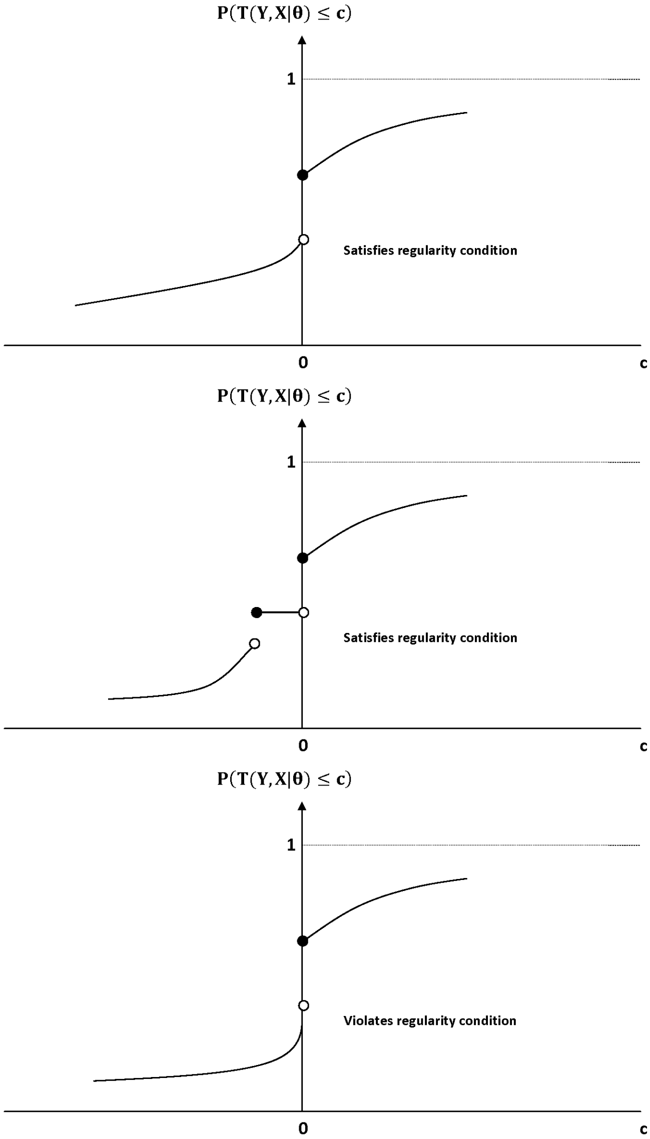


Figure 2: Profiled log-likelihood for each parameter in θ^I

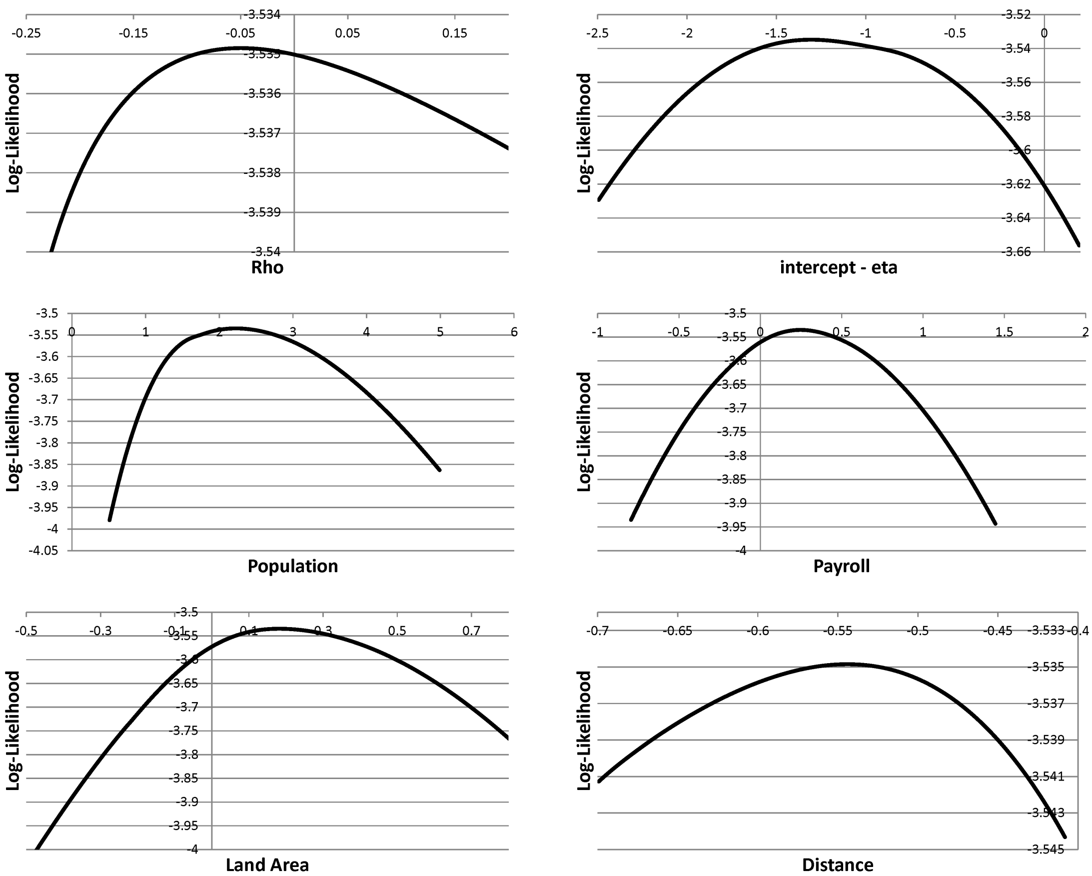


Figure 3: Joint 95% confidence regions for slopes, intercept, and η

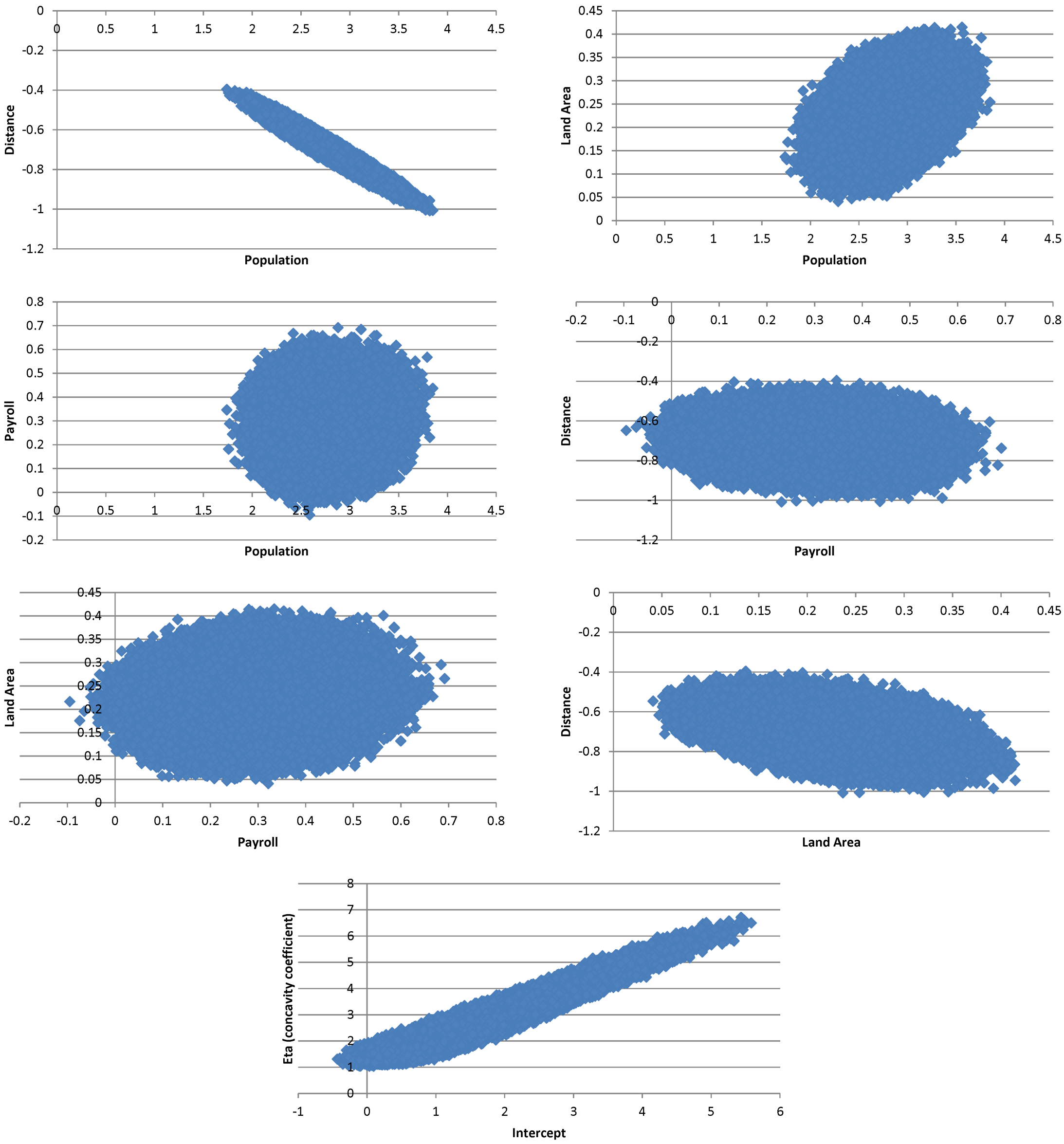


Figure 4: Joint 95% confidence region for strategic interaction coefficients

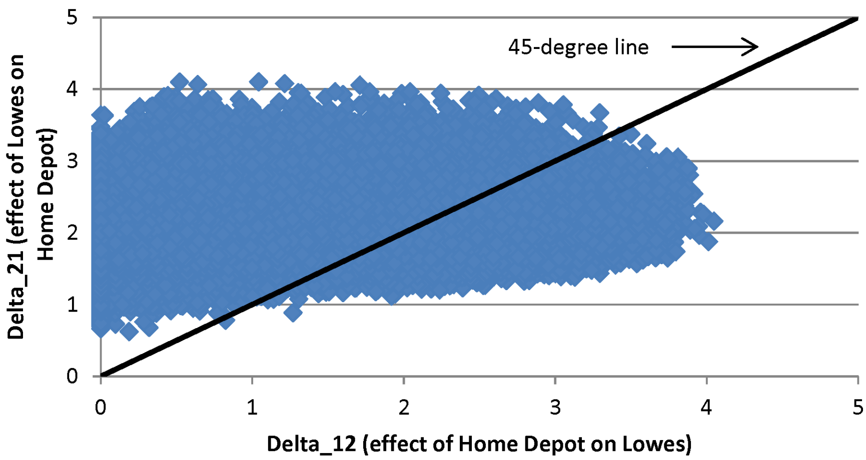


Figure 5: Joint 95% confidence region for strategic interaction coefficients and slope parameters

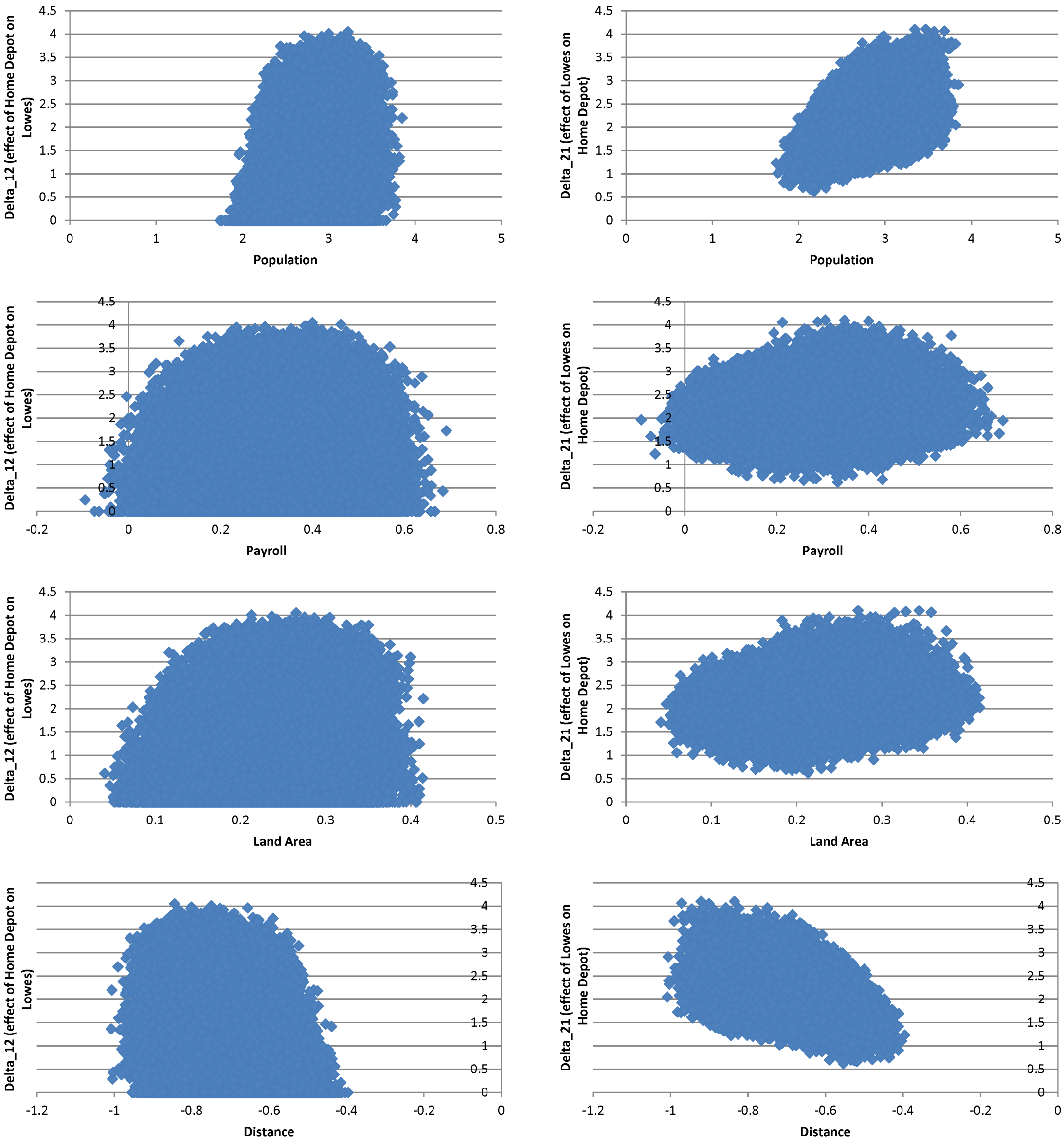
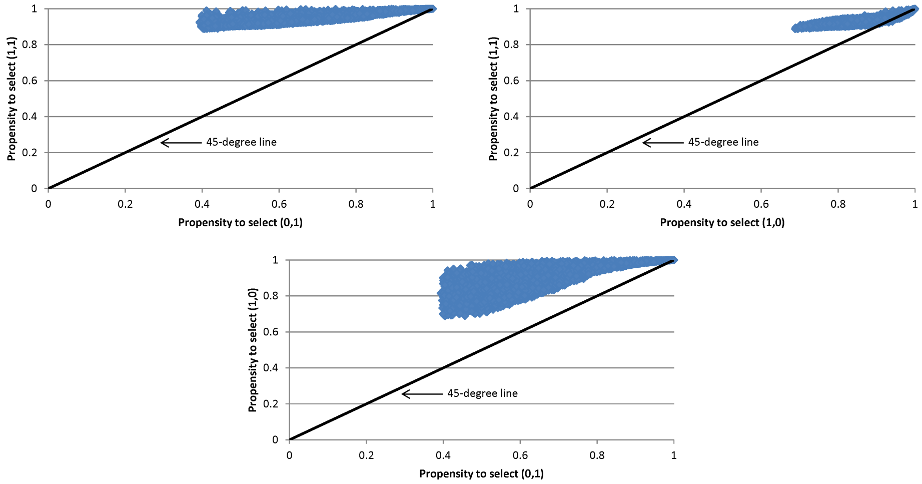
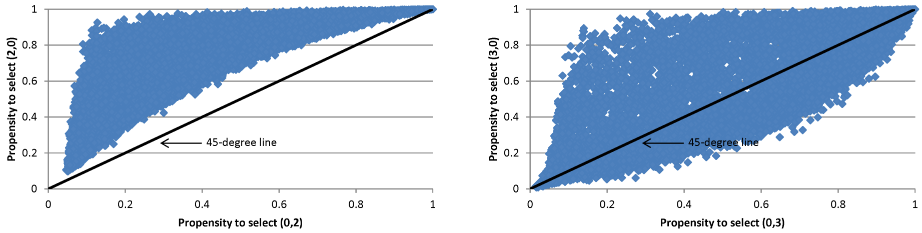


Figure 6: Confidence sets for estimated propensities of equilibrium selection

(A) Equilibria with at most one store by each firm



(B) Equilibria with a monopolist opening multiple stores



(C) Equilibria where both firms enter with the same number of stores