Inference in ordered response games with complete information

Andres Aradillas-Lopez
Adam M. Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP37/21

Inference in Ordered Response Games with Complete Information.*

Andres Aradillas-Lopez†
Pennsylvania State University

Adam M. Rosen‡
Duke University and CeMMAP

October 15, 2021

Abstract

We study inference in complete information games with discrete strategy spaces. Unlike binary games, we allow for rich strategy spaces and we only assume that they are ordinal in nature. We derive observable implications of equilibrium play under mild shape restrictions on payoff functions, and we characterize sharp identified sets for model parameters. We propose a novel inference method based on a test statistic that embeds conditional moment inequalities implied by equilibrium behavior. Our statistic has asymptotically pivotal properties that depend on the measure of contact sets, to which our statistic adapts automatically. In the case of two players and strategic substitutes we show that certain payoff parameters are point identified under mild conditions. We embed conventional point estimates for these parameters in our conditional moment inequality test statistic in order to perform inference on the remaining (partially identified) parameters. We apply our method to model the number of stores operated by Lowe's and Home Depot in geographic markets and perform inference on several quantities of economic interest.

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

JEL classification: C01, C31, C35.

*This paper has benefited from feedback from participants at several conferences and seminars. We thank Eleni Aristodemou, Andrew Chesher, Jeremy Fox, Ariel Pakes, Aureo de Paula, Elie Tamer, and especially Francesca Molinari for helpful discussion and suggestions, and we thank the editor, an associate editor, and two anonymous referees for several comments and suggestions that have helped to improve the paper. Xinyue Bei provided excellent research assistance. Adam Rosen gratefully acknowledges financial support from the UK Economic and Social Research Council through grants (RES-589-28-0001, RES-589-28-0002 and ES/P008909/1) 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, from the European Research Council (ERC) through grant ERC-2009-StG-240910-ROMETA, and from a British Academy Mid-Career Fellowship.

†Address: Department of Economics, Pennsylvania State University, University Park, PA 16802, United States. Email: aaradill@psu.edu.

‡Address: Department of Economics, Duke University, 213 Social Sciences Building, Box 90019, United States. Email: adam.rosen@duke.edu.

1 Introduction

Econometric models of static complete information discrete games are well studied, and have been commonly used to model firm entry decisions. Most research has focused on either binary games or games with a very limited action space. This paper aims to contribute to the literature by tackling the problem of inference in games where the action space is discrete but rich and potentially unbounded. Our only assumption regarding the action space is that it is ordinal in nature. This enables the extension of firm entry models to incorporate not just firm presence, but the intensity of market presence, for instance as measured by the number of stores.

Without further restrictions, a game with a rich action space can pose serious challenges for inference due to multiple equilibria and the potential complexity of equilibrium configurations. We solve this by imposing mild shape restrictions on payoff functions which greatly simplify the necessary conditions that must be satisfied in equilibrium. Under these shape restrictions, equilibrium behavior can be described as a simultaneous move ordered-response game. From here, necessary conditions for a pre-specified action profile y to be an equilibrium can be described in a way that does not require the econometrician to find all equilibria, but instead only requires payoff comparisons across adjacent actions. This greatly simplifies the analysis and allows us to deal with cases where the action space is unbounded.

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 payoff-shifter. Our model employs shape restrictions on payoff functions 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 equalities and inequalities that characterize the identified set of agents' payoff functions.

When the number of actions and/or players is sufficiently large, 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 possible set estimates, this may in some cases be infeasible. We thus also characterize outer sets that embed a subset of the full set of moment inequalities. Use of these outer sets can be computationally much easier for estimation and, as shown in our application, can sometimes achieve economically meaningful inference.

We propose a novel inference method that embeds the conditional moment inequalities (CMI) implied by our model. Our statistic has two main attractive properties. First, it automatically adapts to the measure of so-called "contact sets" (the set of realizations of the conditioning variables where the inequalities are binding at a given parameter value). Second, this feature confers it asymptotically pivotal properties, which allows us to use critical values that do not have to be approximated for each parameter value evaluated, thus facilitating its application to models with many parameters and many conditioning variables. Adapting to the measure of the contact sets also means that our statistic is not conservative, in the sense that it is not based on least favorable configurations in which contact sets have the largest possible measure for each parameter value.

Our inferential method shares some conceptual similarities to other CMI criterion function based approaches such as those of Andrews and Shi (2013), Lee, Song, and Whang (2013), Lee, Song, and Whang (2018), Armstrong (2015), Armstrong (2014), Chetverikov (2017), Armstrong and Chan (2016) and Armstrong (2018), but with two important differences. The first difference regards the scaling of moment inequality violations. The procedures cited above use statistics that measure empirical violations of CMIs scaled by their standard errors. The statistic here instead first aggregates these violations, and then scales the aggregate violation.[1] This approach is what allows our procedure to adapt asymptotically to the measure of the contact sets. Regularizing the estimator for the asymptotic variance of our statistic allows us to standardize it in a way that produces asymptotically pivotal properties.

In the case of two players and strategic substitutes we show that a subset of parameters are point identified under mild conditions. Our proposal is to consistently estimate the point identified parameters in a first step by maximum likelihood, and then incorporate these point estimates in our moment inequality statistic to perform inference on the remaining partially identified parameters. Our statistic is particularly well-suited to this case because of its asymptotic properties, in particular the fact that it has a linear representation.[2] Ultimately we construct a Wald-type statistic incorporating the joint asymptotic distribution of the first-step estimator and our moment inequality statistic.

We apply a parametric version of our model to study market presence, as measured by the number of stores, and competition in geographic markets by Lowe's and Home Depot. In addition to presenting point estimates for the point identified parameters and confidence sets for the partially identified ones, we also illustrate how our model can allow us to perform inference on a number of quantities of economic interest, such as the likelihood that particular action profiles are equilibria, and the propensity of the underlying equilibrium selection mechanism to choose any particular outcome among multiple equilibria. Our results reveal several substantive findings

[1] The scale factor employs a truncation sequence to ensure that it is bounded away from zero, similar to that used by Armstrong (2014) when scaling individual moment inequality violations. Armstrong (2014) shows that for a test based on a Kolmogorov-Smirnov statistic this can lead to improvements in estimation rates and local asymptotic power relative to using bounded weights. For the statistic considered here, which is based on aggregate moment inequality violations, truncation by a decreasing sequence ensures that the violation is asymptotically weighted by its inverse standard error, which is used to establish the asymptotic validity of fixed chi-square critical values.

[2] Other recent papers that feature set identification with a point-identified component but with different approaches include Kaido and White (2014), Kline and Tamer (2016), Romano, Shaikh, and Wolf (2014), and Shi and Shum (2015). The first of these focuses primarily on consistent set estimation, with subsampling suggested for inference. The other three papers provide useful and widely applicable approaches for inference based on unconditional moment inequalities, but do not cover inference based on conditional moment inequalities with continuous conditioning variables, as encountered here.

that illuminate the nature of the strategic interaction between these firms in our model, as well as some features of the underlying equilibrium selection mechanism.

The paper proceeds as follows. In Section 1.1 we discuss the related literature on econometric models of discrete 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 then introduce our approach for inference on elements of the identified set. In Section 6 we briefly summarize findings from Monte Carlo experiments, the full details of which are provided in an additional Empirical Supplement.[3] In Section 7 we apply our method to model capacity (number of stores) decisions by Lowe's and Home Depot. Some additional results from the empirical application excluded here for brevity are additionally provided in the aforementioned Empirical Supplement. Section 8 concludes. All proofs of our econometric results can be found in an accompanying Econometric Supplement.[4]

1.1 Related Literature and the Contribution of this Paper

Econometric models of discrete action games of complete information fall in the class of simultaneous discrete choice models, which have been studied in several papers, going back at least to Heckman (1978). A foundational paper that explicitly analyzed econometric inference in a discrete game is Bjorn and Vuong (1984), later followed by the seminal work of Bresnahan and Reiss (1990), Bresnahan and Reiss (1991b), and Berry (1992). These models can produce multiple equilibria for certain realizations of payoff shifters, and they may alternatively fail to produce any (pure strategy) equilibria for others. Multiplicity and non-existence of equilibria result in *incompleteness* and *incoherence*, respectively (see Tamer (2003) and Berry and Reiss (2006)). While an incoherent game-theoretic model may require a different approach, incomplete models can be "completed" by introducing an equilibrium selection theory. For a more thorough treatment of these issues and the related literature, we refer to Chesher and Rosen (2020). A review of the econometric analysis of static games can be found in Aradillas-López (2020).

The idea of relying only on equilibrium conditions without completing a model through a possibly ad-hoc equilibrium selection theory is a common feature of the recent literature. A partial list of examples includes Tamer (2003), Ishii (2005), Ciliberto and Tamer (2009), Andrews and Jia-Barwick (2010), and Pakes, Porter, Ho, and Ishii (2015). Aradillas-López and Tamer (2008) show how weaker restrictions than Nash Equilibrium, in particular rationalizability and iterated deletion of dominated strategies, can be used to set identify the parameters of discrete games. The

[3] Available at `http://www.personal.psu.edu/aza12/ordered-game-empirical-supplement.pdf`
[4] Available at `http://www.personal.psu.edu/aza12/ordered-game-econometric-supplement.pdf`.

econometric challenges presented by games with multiple equilibria have helped produce significant contributions to the general problem of inference in partially identifying models. 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 (2020) build on concepts in both of these papers as well as Chesher and Rosen (2017) to compare identified sets obtained from alternative approaches to deal with incompleteness and in particular incoherence in simultaneous discrete outcome models.

Most existing work has focused on either binary-choice games, or games with a very limited action space. As a consequence, a number of interesting real-world applications fall outside the scope of prior analyses. Our paper aims to contribute to the literature by considering games where the action space is rich and potentially unbounded (or treated as unbounded by the researcher) but *ordinal* in nature. We describe mild shape restrictions for payoff functions that effectively turn the game into a simultaneous ordered-response model where necessary equilibrium conditions involve only comparisons across adjacent actions and thus do not necessitate looking at (or even knowing) the entire action space.

Under these shape restrictions, equilibrium conditions comprise conditional moment inequalities (CMIs) involving payoff comparisons across adjacent actions only. Our paper additionally contributes to the literature on inference methods with CMIs by proposing a novel approach based on a test statistic that embeds all the information of the CMIs, and which adapts asymptotically to the measure of the "contact set" of values of the conditioning variables at which the CMIs bind. This allows us to bypass the need to pretest for the slackness of CMIs while also avoiding the conservative features of tests that use critical values based on a least favorable configuration in which all CMIs are binding. Our approach is also applicable when a subset of parameters are point-identified and can be estimated with a root-n consistent estimator, as illustrated in our application.

Also related to our multiple-entry empirical application 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 (2015). 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.[5]

[5]Of the papers in this literature, only Ellickson, Houghton, and Timmins (2013) and Nishida (2015) also allow an ordered but non-binary within-market action space. Nishida (2015), in similar manner to Jia (2008), employs an equi-

Some other existing papers specifically consider alternative models of ordered response with endogeneity. Davis (2006) considers a simultaneous equations model with a game-theoretic foundation, employing enough additional structure on equilibrium selection so as to complete the model and achieve point-identification. Ishii (2005) studies ATM networks, using 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 and Smolinski (2012) provide set identification results for a single equation ordered response model with endogenous regressors and instrumental variables. Chesher, Rosen, and Siddique (2019) use such models to perform inference on counterfactuals and marginal effects of continuous endogenous variables, in comparison to results obtained by complete triangular models. Here we exploit the structure provided by the simultaneous (rather than single equation) model. Aradillas-López (2011) and Aradillas-López and Gandhi (2016) also consider simultaneous models of ordered response. In contrast to this paper, Aradillas-López (2011) focuses on nonparametric estimation of bounds on Nash outcome probabilities, and Aradillas-López and Gandhi (2016) on a model with *incomplete* information. The parametric structure imposed here allows us to conduct inference on economic quantities of interest and perform counterfactual experiments that are beyond the scope of those papers.

2 The Model

Our model consists of J players $\mathcal{J} = \{1,...,J\}$ who each simultaneously choose an action Y_j from the ordered action space $\mathcal{Y}_j = \left\{s_j^0, s_j^1, \ldots, s_j^{M_j}\right\}$, where $s_j^\ell < s_j^{\ell+1}$ $\forall$ ℓ. The action space $\mathcal{Y}_j$ can be unbounded (above and/or below) and we only require that it be ordinal in nature.[6] Each set $\mathcal{Y}_j$ is discrete but everything that follows can be extended to cases where it is continuous. We will denote a generic element of $\mathcal{Y}_j$ as y_j and we define

$$y_j^- = \begin{cases} s_j^{\ell-1} \text{ if } y_j = s_j^\ell \text{ with } \ell > 0 \\ s_j^0 - 1 \text{ if } y_j = s_j^0 \end{cases} \qquad y_j^+ = \begin{cases} s_j^{\ell+1} \text{ if } y_j = s_j^\ell \text{ with } \ell < M_j \\ s_j^{M_j} + 1 \text{ if } y_j = s_j^{M_j} \end{cases} \tag{1}$$

librium 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.

[6]While the actions may be of cardinal significance in some applications, we will only exploit their ordinal nature.

That is, y_j^- and y_j^+ are the adjacent actions to y_j.[7] We define Y_j^- and Y_j^+ exactly as in (1) for the actions chosen by j. We denote $Y \equiv \left(Y_1,...,Y_J\right)'$ as the action profile chosen by all J players, and for any player $j \in \mathcal{J}$ we adopt the common convention that Y_{-j} denotes the vector of actions chosen by j's rivals, $Y_{-j} \equiv \left(Y_1,...,Y_{j-1},Y_{j+1},...,Y_J\right)'$. As shorthand we sometimes write $\left(Y_j,Y_{-j}\right)$ 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. 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 $(\Omega,\mathcal{F},\mathbb{P}_0)$, where X denotes the composite vector of observable payoff shifters X_j, $j \in J$, without repetition of any common components. We use P to denote the corresponding 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) with distribution P. The random sampling assumption guarantees identification of P.[8]

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

Restriction SR (Shape Restrictions): Payoff functions $\left(\pi_1,...,\pi_J\right)$ 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:
(i) Payoffs are *strictly concave* in y_j:

$$\forall y_j \in \mathcal{Y}_j, \quad \pi_j\left((y_j^+,y_{-j}),x_j,u_j\right) - \pi_j\left((y_j,y_{-j}),x_j,u_j\right) < \pi_j\left((y_j,y_{-j}),x_j,u_j\right) - \pi_j\left((y_j^-,y_{-j}),x_j,u_j\right),$$

where $\pi_j((y_j^-,\cdot),\cdot) = -\infty$ if $y_j = s_j^0$ and $\pi_j((y_j^+,\cdot),\cdot) = -\infty$ if $y_j = s_j^{M_j}$.
(ii) $\forall \left(y_{-j},x_j\right) \in \mathcal{Y}_{-j} \times \mathcal{X}$, $\pi_j\left(\left(y_j,y_{-j}\right),x_j,u_j\right)$ has strictly *increasing differences* in $\left(y_j,u_j\right)$, that is if $u_j' > u_j$ and $y_j' > y_j$, then

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

[7] Note that we have decreed $y_j^- = s_j^0 - 1$ and $y_j^+ = s_j^{M_j} + 1$ for the lower and upper bounds of $\mathcal{Y}_j$, respectively. Thus, these "adjacent" actions fall outside of $\mathcal{Y}_j$. This is done without loss of generality since (as we will see below), players' payoffs are assumed to be $-\infty$ for any action outside of $\mathcal{Y}_j$.

[8] 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 P is identified, for which random sampling is a sufficient, but not necessary, condition.

Restriction SR(i) imposes that marginal payoffs are decreasing in each player's own action y_j. It also implies that, for any fixed rival pure strategy profile y_{-j}, agent j's best response is unique with probability one. Restriction SR(ii) 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 identification analysis by guaranteeing the existence of intervals for u_j within which any y_j maximizes payoffs for any fixed $\left(y_{-j}, x_j\right)$.

We study models in which the distribution of unobserved heterogeneity is restricted to be independent of payoff shifters. This restriction can be relaxed, see e.g. Kline (2015), 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 $\mathcal{P}_U$. ■

A *structure* comprises a collection of payoff functions $\left(\pi_1, ..., \pi_J\right) \in \mathbf{\Pi}$ satisfying Restrictions SR and I, and a distribution of unobserved heterogeneity $P_U \in \mathcal{P}_U$. Identification analysis aims to deduce which structures $(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U$, and what relevant functionals of (π, P_U), are compatible with P. The collections $\mathbf{\Pi}$ and $\mathcal{P}_U$ may each be parametrically, semiparametrically, or nonparametrically specified. If they are both parametrically specified, then $\mathbf{\Pi}$ and $\mathcal{P}_U$ may be indexed by a finite dimensional parameter vector, say θ. Then each θ in some given parameter space Θ represents a unique structure (π, P_U), and identification analysis reduces to deducing which $\theta \in \Theta$ have associated structures (π, P_U) compatible with P.

3 Equilibrium Behavior and Observable Implications

We assume that players have complete information and thus know the realizations of all payoff shifters (X, U) when they choose their actions.[9] We focus attention on Pure Strategy Nash Equilibrium (PSNE) as our solution concept, although other solution concepts can be used with our inference approach. For example, 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). In the working paper Aradillas-Lopez and Rosen (2014) we further describe an alternative behavioral model that nests Nash equilibrium (in either pure or mixed strategies) as a special case but that allows for incorrect beliefs, and we outline how our inference approach can then be applied.[10]

[9]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 (2016) and the references therein.

[10]One motivation for using alternative solution concepts is the possibility of non-existence of PSNE. However, in games in which all actions are strategic complements, or in 2 player games where actions are either strategic substitutes or complements, such as that used in our application, 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 a reformulation used by Vives (1999,

We now formalize the restriction to PSNE behavior. Define each player j's best response correspondence as

$$\boldsymbol{y}_j^*\left(y_{-j}, x_j, u_j\right) \equiv \underset{y_j \in \mathcal{Y}_j}{\arg\max} \quad \pi_j\left(\left(y_j, y_{-j}\right), x_j, u_j\right), \tag{2}$$

which delivers the set of payoff maximizing alternatives y_j for player j as a function of $\left(y_{-j}, x_j, u_j\right)$.
Restriction PSNE (Pure Strategy Nash Equilibrium): With probability one, for all $j \in \mathcal{J}$, $Y_j \in \boldsymbol{y}_j^*\left(Y_{-j}, X_j, U_j\right)$.

Strict concavity of each player j's payoff in her own action under Restriction SR(i) in fact 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. We therefore adopt a minor abuse of notation and write $Y_j = \boldsymbol{y}_j^*\left(Y_{-j}, X_j, U_j\right)$. Concavity also provides a further simplification of the conditions for PSNE.

Lemma 1 *Suppose Restriction SR(i) holds. Then Restriction PSNE holds if and only if with probability one, for all* $j \in \mathcal{J}$, $\pi_j\left(Y, X_j, U_j\right) \geq \max\left\{\pi_j\left(\left(Y_j^+, Y_{-j}\right), X_j, U_j\right), \pi_j\left(\left(Y_j^-, Y_{-j}\right), X_j, U_j\right)\right\}$*, where, as before, we define* $\pi_j\left(\left(Y_j^-, Y_{-j}\right), X_j, U_j\right) = -\infty$ *if* $Y_j = s_j^0$ *and* $\pi_j\left(\left(Y_j^+, Y_{-j}\right), X_j, U_j\right) = -\infty$ *if* $Y_j = s_j^{M_j}$.

The proof of Lemma 1 is simple and thus omitted. That Restriction PSNE implies the inequality in the lemma is immediate. The other direction follows from noting that if the inequality in Lemma 1 holds then violation of (2) would contradict strict concavity of $\pi_j\left(\left(y_j, Y_{-j}\right), X_j, U_j\right)$ in y_j.

We now characterize the identified set of structures (π, P_U). Define

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

as the incremental payoff of action Y_j relative to Y_j^- for any $\left(Y_{-j}, X, U_j\right)$. From Restriction SR (ii), $\Delta\pi_j\left(Y, X, U_j\right)$ is strictly increasing in U_j and thus invertible. Combining this with Lemma 1 allows us to deduce that for each player j there exists, for each $\left(y_{-j}, x\right)$, an increasing sequence of non-overlapping thresholds, $\left\{u_j^*\left(y_j, y_{-j}, x\right) : y_j = s_j^0, \ldots, s_j^{M_j}, s_j^{M_j} + 1\right\}$, with $u_j^*(s_j^{M_j} + 1, y_{-j}, x) = -u_j^*\left(s_j^0, y_{-j}, x\right) = \infty$, such that[11]

$$\boldsymbol{y}_j^*\left(y_{-j}, x_j, u_j\right) = y_j \Leftrightarrow u_j^*\left(y_j, y_{-j}, x\right) < u_j \leq u_j^*\left(y_j^+, y_{-j}, x\right). \tag{3}$$

That is, given $\left(y_{-j}, x\right)$, each player j's best response $\boldsymbol{y}_j^*\left(y_{-j}, x_j, u_j\right)$ is uniquely determined by within which of the non-overlapping intervals $\left(u_j^*\left(y_j, y_{-j}, x\right), u_j^*\left(y_j^+, y_{-j}, x\right)\right]$ unobservable U_j falls. It fol-

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.

[11] Recall from (1) that we have decreed $y_j^+ = s_j^{M_j} + 1$ if $y_j = s_j^{M_j}$.

lows that with probability one

$$U \in \mathcal{R}_\pi(Y,X) \equiv \underset{j\in\mathcal{J}}{\times}\left(u_j^*\left(Y_j, Y_{-j}, X\right), u_j^*\left(Y_j^+, Y_{-j}, X\right)\right]. \tag{4}$$

In other words, Y is an equilibrium precisely if U belongs to the rectangle $\mathcal{R}_\pi(Y,X)$.

The equilibrium implication (4) is the key implication that, when combined with previous set identification results in the literature – e.g. those of Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), and Chesher and Rosen (2017) – delivers the identified set for (π, P_U). Define for any set $\tilde{\mathcal{Y}} \subseteq \mathcal{Y}$ and all $x \in \mathcal{X}$, $\mathcal{R}_\pi\left(\tilde{\mathcal{Y}}, x\right) \equiv \bigcup_{y\in\tilde{\mathcal{Y}}} \mathcal{R}_\pi(y,x)$, 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 SR, I, and PSNE hold. Then the identified set of structures is*

$$\mathcal{S}^* = \left\{(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \forall \mathcal{U} \in \mathsf{R}^\cup(x),\ P_U(\mathcal{U}) \geq P\left(\mathcal{R}_\pi(Y,X) \subseteq \mathcal{U} | X = x\right)\ a.e.\ x \in \mathcal{X}\right\}, \tag{5}$$

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

$$\mathsf{R}^\cup(x) \equiv \left\{\mathcal{U} \subseteq \mathbb{R}^J : \begin{array}{c} \mathcal{U} = \mathcal{R}_\pi\left(\tilde{\mathcal{Y}}, x\right) \text{ for some } \tilde{\mathcal{Y}} \subseteq \mathcal{Y} \text{ such that } \forall \text{ nonempty } \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2 \subseteq \mathcal{Y} \text{ with} \\ \tilde{\mathcal{Y}}_1 \cup \tilde{\mathcal{Y}}_2 = \tilde{\mathcal{Y}} \text{ and } \tilde{\mathcal{Y}}_1 \cap \tilde{\mathcal{Y}}_2 = \emptyset,\ P_U\left(\mathcal{R}_\pi\left(\tilde{\mathcal{Y}}_1, x\right) \cap \mathcal{R}_\pi\left(\tilde{\mathcal{Y}}_2, x\right)\right) > 0 \end{array}\right\}. \tag{6}$$

Equivalently, the identified set of structures can be written

$$\mathcal{S}^* = \left\{\begin{array}{c} (\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \\ \forall \mathcal{C} \in 2^{\mathcal{Y}},\ P_U\left(\exists y \in \mathcal{C} : y \in \mathrm{PSNE}(\pi, X, U) | X = x\right) \geq P\left(Y \in \mathcal{C} | X = x\right)\ a.e.\ x \in \mathcal{X} \end{array}\right\}, \tag{7}$$

where $\mathrm{PSNE}(\pi, X, U)$ *denotes the set of PSNE when the payoff functions are π for the given* (X, U).

Characterization (5) follows from application of Chesher and Rosen (2017, Theorems 2-3), while (7) can be obtained by application of either one of Galichon and Henry (2011, Theorem 1) or Beresteanu, Molchanov, and Molinari (2011, Theorem D.2). The former expresses $\mathcal{S}^*$ as those (π, P_U) such that $P_U(\mathcal{U}) \geq P\left(\mathcal{R}_\pi(Y,X) \subseteq \mathcal{U} | X = x\right)$, a.e $x \in \mathcal{X}$, over the collection of sets $\mathcal{U} \in \mathsf{R}^\cup(x)$.[12] The collection $\mathsf{R}^\cup(x)$ is a collection of *core-determining sets*, as defined by Galichon and Henry (2011, Theorem 1), shown to be sufficient by Chesher and Rosen (2017, Theorem 3) to imply (5) for all closed $\mathcal{U} \subseteq \mathbb{R}^J$, thus ensuring sharpness. Applying those results here, and in particular using

[12]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 trivially when $\tilde{\mathcal{Y}} = \{y\}$.

the implication that Y is a PSNE if and only if (4) holds, we see that the core-determining sets in this model comprise unions of rectangles in $\mathbb{R}^2$. The proof of Theorem 1 is omitted; equivalence of the two representations follows from application of Chesher and Rosen (2017, Theorem 4).

The identified set $\mathcal{S}^*$ expressed in (5) may comprise a rather large number of conditional moment inequalities, namely as many as belong to $\mathsf{R}^{\cup}(x)$, for each x. More inequalities will in general produce smaller identified sets, but the use of a very large number of inequalities may pose a practical challenge, 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 (5) applied to 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* $\mathsf{U}(x) : \mathcal{X} \to 2^{\mathcal{U}}$*. Then* $\mathcal{S}^*(\mathsf{U})$ *is an outer set for* $\mathcal{S}^*$*, in that* $\mathcal{S}^* \subseteq \mathcal{S}^*(\mathsf{U})$*, where*

$$\mathcal{S}^*(\mathsf{U}) = \{(\pi, P_U) \in \mathbf{\Pi} \times \mathcal{P}_U : \forall \mathcal{U} \in \mathsf{U}(x),\ P_U(\mathcal{U}) \geq P(\mathcal{R}_\pi(Y,X) \subseteq \mathcal{U} | X = x) \text{ a.e. } x \in \mathcal{X}\}. \tag{8}$$

The Corollary follows directly from Theorem 1 because the inequality

$$P_U(\mathcal{U}) \geq P(\mathcal{R}_\pi(Y,X) \subseteq \mathcal{U} | X = x) \tag{9}$$

for all $\mathcal{U} \in \mathsf{R}^{\cup}(x)$ and almost every $x \in \mathcal{X}$ implies that this must hold for all $\mathcal{U} \in \overline{\mathsf{R}^{\cup}}(x)$, and further for all $\mathcal{U} \in \mathsf{U}(x)$. Furthermore, 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). Although $\mathcal{S}^*(\mathsf{U})$ is a larger set than $\mathcal{S}^*$, its reliance on fewer inequalities can lead to significant computational gains for bound estimation and inference. 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 could potentially 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.[13] Typically, checking the imposed inequality restrictions involves searching over a multi-dimensional parameter space, so the computational advantage can be substantial.

The difference between the size of the outer set $\mathcal{S}^*(\mathsf{U})$ and the identified set may or may not be large. For a given collection of conditional moment inequalities defining $\mathcal{S}^*(\mathsf{U})$, this will depend on the particular distribution of (Y, X) at hand, and is thus an empirical question. In the two-player parametric model studied in the following Section, and used in the application of Section 7, we show that a particular $\mathsf{U}(\cdot)$ is sufficient to point identify all but three of the model parameters.

[13]This will be valid a valid procedure if the researcher can ensure that the confidence sets are constructed such that that one based on the first outer set contains the one based on the second set incorporating further restrictions with probability one.

4 A Two-Player Game of Strategic Substitutes

In this section we introduce a parametric specification satisfying Restriction SR for a two-player game with $\mathcal{J} = \{1, 2\}$, which we use in our empirical application. We continue to maintain Restrictions I and PSNE. 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 noted in Section 3.

4.1 A Parametric Specification

For each $j \in \mathcal{J}$ we specify the action space $\mathcal{Y}_j = \left\{s_j^0, s_j^1, \ldots, s_j^{M_j}\right\} = \{0, 1, ..., M_j\}$ and payoffs

$$\pi_j\left(Y, X_j, U_j\right) = Y_j \times \left(\delta + X_j\beta - \Delta_j Y_{-j} - \eta Y_j + U_j\right), \tag{10}$$

where we impose the restriction that $\eta > 0$, ensuring that Restrictions SR(i) and SR(ii) hold. In this specification the parameters of the players' payoff functions differ only in the interaction parameters (Δ_1, Δ_2), both restricted to be nonnegative so that actions are strategic substitutes. Given (10), each player j's best response function takes the form (3), where

$$u_j^*\left(0, y_{-j}, x_j\right) = -\infty, \quad u_j^*\left(\tilde{y}_j, y_{-j}, x_j\right) = \eta\left(2\tilde{y}_j - 1\right) + \Delta_j y_{-j} - \delta - x_j\beta, \quad \tilde{y}_j = 1, ..., M_j. \tag{11}$$

In addition we restrict the distribution of bivariate unobserved heterogeneity U to the Farlie-Gumbel-Morgenstern (FGM) copula indexed by parameter $\lambda \in [-1, 1]$.[14] Specifically U_1 and U_2 each have the logistic marginal CDF $G(u_j) = \exp(u_j)/(1 + \exp(u_j))$, and their joint CDF is

$$F(u_1, u_2; \lambda) = G(u_1) \cdot G(u_2) \cdot [1 + \lambda (1 - G(u_1))(1 - G(u_2))]. \tag{12}$$

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 7. Note that ρ captures the correlation remaining after controlling for X. Thus with a sufficiently rich set of payoff shifters 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 $F(u_1, u_2; \lambda)$ 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 $\theta_1 \equiv (\alpha, \beta', \lambda)'$ and $\theta_2 = (\eta, \Delta_1, \Delta_2)'$. We show in the following

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

Section that under fairly mild conditions θ_1 is point identified.[15]

4.2 Observable Implications of Pure Strategy Nash Equilibrium

Given the parametric specification, we express the sets $\mathcal{R}_\pi(Y,X)$ described in (4) as $\mathcal{R}_\theta(Y,X)$. It follows from (11) that observed (Y,X) correspond to a PSNE if and only if $U \in \mathcal{R}_\theta(Y,X)$ where

$$\mathcal{R}_\theta(Y,X) \equiv \mathcal{R}_\theta^1 \times \mathcal{R}_\theta^2, \tag{13}$$

$$\mathcal{R}_\theta^j(Y,X) = \{u_j : -\infty < u_1 \leq \eta + \Delta_j Y_{-j} - \delta - X_j\beta\} \quad \text{if } Y_j = 0, \tag{14}$$

$$\mathcal{R}_\theta^j(Y,X) = \{u_j : \eta\left(2Y_j - 1\right) + \Delta_j Y_{-j} - \delta - X_j\beta < u_j \leq \eta\left(2Y_j + 1\right) + \Delta_j Y_{-j} - \delta - X_j\beta\} \text{ if } Y_j > 0. \tag{15}$$

From Theorem 1 we have the inequality

$$P_U(\mathcal{U}) \geq P(\mathcal{R}_\theta(Y,X) \subseteq \mathcal{U} | X = x) \tag{16}$$

for each $\mathcal{U} \in \mathsf{R}^{\cup}(x)$, a.e. $x \in \mathcal{X}$ (see the definition of $\mathsf{R}^{\cup}(x)$ in (6)). However, it is straightforward to see that $Y = (0,0)$ is a PSNE if and only if

$$U \in (-\infty, \alpha - X_1\beta) \times (-\infty, \alpha - X_2\beta). \tag{17}$$

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). This implies that the conditional moment inequality (16) using $\mathcal{U} = (-\infty, \alpha - X_1\beta) \times (-\infty, \alpha - X_2\beta)$ in fact holds with *equality*.[16] Therefore with $\tilde{\beta} \equiv (\alpha, \beta')'$, and $Z_j \equiv \left(1, -X_j\right)$,

$$P(Y = (0,0)|X) = F\left(Z_1\tilde{\beta}, Z_2\tilde{\beta}; \lambda\right), \tag{18}$$

with $F(\cdot,\cdot;\lambda)$ defined in (12). Henceforth, we will group $W \equiv (Y,X)$. The log-likelihood for the event $Y = (0,0)$ and its complement is then

$$\mathcal{L}(\theta_1) = \sum_{i=1}^{n} \ell(\theta_1; w_i), \tag{19}$$

where $\ell(\theta_1; w) \equiv 1[y = (0,0)] \cdot \log F\left(z_1\tilde{\beta}, z_2\tilde{\beta}; \lambda\right) + 1[y \neq (0,0)] \cdot \log\left(1 - F\left(z_1\tilde{\beta}, z_2\tilde{\beta}; \lambda\right)\right)$. The following theorem establishes that under suitable conditions $E[\ell(\theta_1, W)]$ is uniquely maximized at the pop-

[15] Results from Kline (2015) establish sufficient conditions for point identification of (α, β) under alternative distributions of unobserved heterogeneity, e.g. multivariate normal.

[16] Indeed, this equality is implied by the set of inequalities that define the identified set $\mathcal{S}^*$ stated in Theorem 1 because those inequalities imply that (16) holds for both $\mathcal{R}_\theta((0,0),x)$ and its complement. Since these two sets partition $\mathbb{R}^J$, each side of the inequalities (16) applied to them sum to one, and thus both inequalities must hold with equality.

ulation value for θ_1, which we denote θ_1^*. Thus θ_1^* is point identified and consistently estimated by $\widehat{\theta}_1$, the maximizer of (19), at the parametric rate.

Theorem 2 *For each player $j \in \{1,2\}$ let payoffs take the form (10) with $\tilde{\beta} \equiv (\alpha, \beta')'$ belonging to a known compact set $B \subseteq \mathbb{R}^{r-1}$ and $U \perp\!\!\!\perp X$. Let Restriction PSNE hold and assume (i) $\forall j \in \{1,2\}$ there exists no proper linear subspace of the support of $Z_j \equiv \left(1, -X_j\right)$ that contains Z_j wp1, and (ii) for all conformable column vectors $c \neq 0$, either $P\{Z_2 c \leq 0 | Z_1 c < 0\} > 0$ and $P\{Z_1 c < 0\} > 0$ or $P\{Z_2 c \geq 0 | Z_1 c > 0\} > 0$ and $P\{Z_1 c > 0\} > 0$. Then:*

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 θ_1 takes the form $\{(b(\lambda), \lambda) : \lambda \in \Gamma'\}$ for some function $b(\cdot) : \Gamma \to B$ and some $\Gamma' \subseteq \Gamma$.*

2. *Suppose that the CDF of U is restricted to the parametric family $F(\cdot,\cdot;\lambda)$ given in (12) for some $\lambda \in (-1,1)$ and for each player $j = 1,2$, the ℓ^{th} regressor $X_{j,\ell}$ is continuously distributed conditional on $X_{j,-\ell}$ with $\beta_\ell \neq 0$. Accordingly, $\beta_\ell \neq 0$ for all $\beta \in \Theta$. Take any pair of vectors $c_a \neq c_b$ such that $c_{a,\ell} \neq 0$ and $c_{b,\ell} \neq 0$ and denote $Z_1 c_a \equiv u_{1a}$, $Z_1 c_b \equiv u_{1b}$, $Z_2 c_a \equiv u_{2a}$, and $Z_2 c_b \equiv u_{2b}$. Denote $u_a \equiv (u_{1a}, u_{2a})$ and $u_b \equiv (u_{1b}, u_{2b})$. As before, let $G(\cdot)$ denote the logistic cdf and define*

$$
\begin{aligned}
C(u_a, u_b) &\equiv G(u_{1a})G(u_{2a}) - G(u_{1b})G(u_{2b}), \\
D(u_a) &\equiv G(u_{1a})(1 - G(u_{1a}))G(u_{2a})(1 - G(u_{2a})), \\
D(u_b) &\equiv G(u_{1b})(1 - G(u_{1b}))G(u_{2b})(1 - G(u_{2b})).
\end{aligned}
$$

   Let $\Psi(u_a, u_b) \equiv (C(u_a, u_b), D(u_a), D(u_b))$. If there is no proper linear subspace of the support of $\Psi(u_a, u_b)$ that contains $\Psi(u_a, u_b)$ w.p.1. for any such pair $c_a \neq c_b$, then θ_1^* is point identified and uniquely maximizes $E[\ell(\theta_1, W)]$. Moreover,

$$
\sqrt{n}\left(\widehat{\theta}_1 - \theta_1^*\right) \xrightarrow{d} \mathcal{N}\left(0, H_0^{-1}\right), \quad \textit{where} \quad H_0 = E\left[\frac{\partial \ell\left(\theta_1^*; W\right)}{\partial \theta_1} \frac{\partial \ell\left(\theta_1^*; W\right)'}{\partial \theta_1}\right]. \tag{20}
$$

Theorem 2 makes use of two conditions on the variation in X. The first condition is standard, requiring that for each j, $Z_j \equiv \left(1, -X_j\right)$ is contained in no proper linear subspace with probability 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 $Z_j c < 0$ (> 0), $Z_{-j} c$ is nonpositive (nonnegative) with positive probability. This condition guarantees that for any $b \neq \tilde{\beta}^*$ there exists a positive measure set of values for Z such that the indices $z_1 b$ and $z_2 b$ are either *both* larger than or *both* smaller than each of $z_1 \tilde{\beta}^*$ and $z_2 \tilde{\beta}^*$, with at least one of the comparisons holding strictly. Thus, either $F(z_1 \tilde{\beta}^*, z_2 \tilde{\beta}^*; \lambda) > F(z_1 b, z_2 b; \lambda)$ or $F(z_1 \tilde{\beta}^*, z_2 \tilde{\beta}^*; \lambda) < F(z_1 b, z_2 b; \lambda)$. Condition (ii) 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. It can accommodate bounded covariates.

The theorem establishes point identification of θ_1 if the distribution of unobserved heterogeneity is known. If instead the model restricts the distribution of unobserved heterogeneity to a parametric family $\{F_\lambda : \lambda \in \Gamma\}$, there is for each $\lambda \in \Gamma$ a unique $\beta = b(\lambda)$ that maximizes the expected log-likelihood for each $\lambda \in \Gamma$. When F_λ is restricted to the FGM family, the full-rank conditions stated for the functionals described in the theorem are sufficient for 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 partially identified.

5 Inference on the Full Parameter Vector

In order to perform inference on the full parameter vector θ we combine the results of Theorem 2 with conditional moment inequalities of the form (16). Let F_Y and F_X denote the distributions of Y and X, with supports $\mathcal{Y}$ and $\mathcal{X}$, and typical realizations y and x, respectively. Let $W \equiv (Y,X) \sim P$, with support $\mathcal{W}$ and typical realization denoted w. We assume throughout 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. We denote $z \equiv \dim(X^c)$.

Our approach will use moment functions $m_k(Y,y,x;\theta)$ consisting of indicators over classes of sets, indexed by y, x and θ:

$$m_k(Y,y,x;\theta) \equiv 1[\mathcal{R}_\theta(Y,x) \subseteq \mathcal{U}_k(x,y;\theta)] - P_U(\mathcal{U}_k(x,y;\theta);\theta), \tag{21}$$

where $\mathcal{R}_\theta(Y,x)$ is the rectangle defined in (13). For values of (y,x) we will use those values (y_i,x_i) that are observed in the data. Test sets $\mathcal{U}_k(x_i,y_i;\theta)$ – and hence the corresponding moment inequalities – are additionally indexed by both $k = 1,...,K$ and θ.[17] These sets are chosen by the econometrician. For $x \in \mathcal{X}$ let $\varphi_k(y,x;\theta) \equiv E[m_k(Y,y,x;\theta)|X = x]$. Our procedure is based on the implication of our model that $\varphi_k(Y,X;\theta) \leq 0$ almost surely. Since our approach replaces φ_k with a nonparametric estimator, we construct a statistic that tests whether this inequality is satisfied over a subset of $\mathcal{X}$ where our nonparametric estimators have desirable properties uniformly. We describe this next.

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

5.1 Specifying an inference range

Let $\mathcal{X}^*$ be a pre-specified set such that the projection of $\mathcal{X}^*$ onto $\mathcal{X}^c$ (the support of X^c) is contained in the interior of the projection of $\mathcal{X}$ onto $\mathcal{X}^c$. The set $\mathcal{X}^*$ will be our "inference range". It will ensure that our nonparametric estimators have the necessary uniform asymptotic properties. 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 $\mathcal{X}^*$.

5.2 A statistic to test our model's inequalities

Let $\omega : \mathcal{X} \to \mathbb{R}_+$ be a "trimming function", chosen by the econometrician, which is strictly positive over $\mathcal{X}^*$ and zero everywhere else. The trimming function is bounded above by some constant $\overline{\omega} < \infty$. For a given $w \equiv (y, x)$, define $T_k(w;\theta) \equiv \varphi_k(y, x;\theta) \cdot \omega(x) \cdot f_X(x)$, where $f_X(\cdot)$ is the density of x. Let $\mathcal{W}^*$ be the restriction of $\mathcal{W}$ (the joint support of (Y, X)) such that $X^c \in \mathcal{X}^*$. Note that, by design, for a.e $w \in \mathcal{W}^*$, $T_k(w;\theta) \leq 0$ if and only if $\varphi_k(y, x;\theta) \leq 0$. Our population statistic is

$$R(\theta) \equiv E\left[\omega(X) \cdot \sum_{k=1}^{K} (T_k(W;\theta))_+\right], \tag{22}$$

where $(\cdot)_+ \equiv \max\{\cdot, 0\}$, and the expectation is taken with respect to the distribution of W. The function $R(\theta)$ is nonnegative, and positive only for θ that violate the conditional moment inequality $\varphi_k(y, x;\theta) \leq 0$ with positive probability over our inference range for some $k = 1, ..., K$.

We employ an estimator for $R(\theta)$ that uses a kernel estimator for each $T_k(W;\theta)$, $k = 1, ..., K$, denoted $\widehat{T_k}(w;\theta)$. For kernel-weighting we define

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

where $\mathbf{K}^c : \mathbb{R}^z \to \mathbb{R}$ and $h_n \searrow 0$ are a kernel function and bandwidth sequence that obey Restriction I2 below. The estimators $\widehat{T_k}(w_i;\theta)$ are then defined by

$$\widehat{T_k}(w;\theta) \equiv \frac{1}{nh_n^z} \sum_{i=1}^{n} m_k(y_i, y, x;\theta) \cdot \omega(x_i) \cdot \mathbf{K}(x_i - x; h_n). \tag{23}$$

An advantage of using density-weighted conditional moments is that it enables the use of sample averages without having to use density estimators in the denominator. The trimming function ω and the properties of the inference range $\mathcal{X}^*$ help us to avoid "boundary bias" issues for X^c.[18] To establish desirable properties for these estimators we impose in Restriction I1 that the functional

[18] Alternatively we could employ boundary kernels. For simplicity we focus on the use of a fixed inference range assumed to be bounded away from the boundary of the support of X^c.

$T_k(w;\theta)$ be sufficiently smooth in the continuous components of x. Restriction I2 contains our formal requirements for the kernel function and bandwidth sequences.
Restriction I1 (Smoothness): For some $M \geq 2z+1$, $\omega(x)$, $f_X(x)$, and $\varphi_k(y,x;\theta)$ are almost surely M-times continuously differentiable with respect to x^c, with bounded derivatives, uniformly in $(y,x,\theta) \in \mathcal{Y} \times \mathcal{X}^* \times \Theta$, and f_X is bounded above by some constant $\overline{f} < \infty$. ■

Our test statistic replaces the function $(\psi)_+ \equiv \psi \cdot 1[\psi \geq 0]$ with $\psi \cdot 1[\psi \geq -b_n]$ for an appropriately chosen sequence $b_n \searrow 0$. The estimator is thus of the form

$$\widehat{R}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \omega(x_i) \cdot \left(\sum_{k=1}^{K} \widehat{T}_k(w_i;\theta) \cdot 1\left[\widehat{T}_k(w_i;\theta) \geq -b_n\right] \right).$$

We next describe the conditions imposed on the kernel function and tuning parameters.
Restriction I2 (Kernels and bandwidths): K^c is a bias-reducing kernel of order M that is symmetric around zero, with bounded support, exhibits bounded variation, and satisfies $\sup_{v \in \mathbb{R}^z} |K(v)| \leq \overline{K} < \infty$. For some $\epsilon > 0$, the positive bandwidth sequences b_n and h_n satisfy $n^{1/2-\epsilon} h_n^z b_n \to \infty$, $h_n^{-z} b_n n^{\epsilon} \to 0$, and $n^{1/2+\epsilon} b_n^2 \to 0$, where M is large enough such that $n^{1/2+\epsilon} h_n^M \to 0$, as $n \to \infty$. ■

Consider bandwidths of the form $h_n = C_a \cdot n^{-\alpha_a}$ and $b_n = C_b \cdot n^{-\alpha_b}$. If $\alpha_a = \frac{4z+1}{8z\cdot(2z+1)}$ and $\alpha_b = \frac{8z+5}{16\cdot(2z+1)}$, Restriction I2 is satisfied for $M = 2z+1$ and any $\epsilon < 1/(8\cdot(2z+1))$. Alternative methods for bias reduction, such as the type of bandwidth-jacknife procedure studied in Honoré and Powell (2005, Section 3.3), could potentially be employed instead of higher order kernels, but we focus on the latter approach here.

5.2.1 Asymptotic properties of $\widehat{R}(\theta)$

The limiting distribution of our statistic, like others in the literature, is driven by values of observable variables for which moment inequalities are satisfied with *equality*, i.e. the contact set $\{w \in \mathcal{W}^* : \exists k \in \{1,...,K\} \text{ s.t. } T_k(w;\theta) = 0\}$. Several papers have proposed methods to explicitly detect the set of inequalities that are close to binding for the purpose of calibrating critical values. These include generalized moment selection as developed by Andrews and Soares (2010) and Andrews and Shi (2013), adaptive inequality selection as in Chernozhukov, Lee, and Rosen (2013), the refined moment selection procedure developed by Chetverikov (2017), and the use of contact set estimators proposed by Lee, Song, and Whang (2018). All of these procedures use tuning parameters. In this paper the sequence b_n is used to ensure that the sample criterion $\widehat{R}(\theta)$ adapts automatically to the contact set, bypassing the need to estimate it explicitly or employ moment selection for computing critical values. To this end we introduce the following restriction.
Restriction I3 (VC classes of sets): The following are VC classes of sets (see Pakes and Pollard (1989, Definition (2.2)), Kosorok (2008, Section 9.1.1)) for each $k = 1,\ldots,K$: $\mathscr{D}_{1,k} \equiv \{y \in \mathcal{Y}\colon R_\theta(y,x) \subseteq \mathcal{U}_k(x,v;\theta) \text{ for some } (v,x) \in \mathcal{W},\ \theta \in \Theta\}$, $\mathscr{D}_{2,k} \equiv \{\mathcal{U}_k(x,y;\theta) \text{ for some } (y,x) \in \mathcal{W},\ \theta \in \Theta\}$,

and $\mathscr{D}_{3,k} \equiv \{w \in \mathcal{W}\colon\ T_k(w;\theta) < c \ \text{ for some } \theta \in \Theta,\, c \in \mathbb{R}\}$. ■

There are several known criteria that suffice for a class of sets to have the VC property (see, e.g, Pollard (1984, Section II.4), Dudley (1984, Section 9), Kosorok (2008, Section 9.1.1)). In particular, the class of sets of the form $\mathcal{U}_k(x,y;\theta)$ that comprise $\mathscr{D}_{2,k}$ is a collection of rectangles in Euclidean space, which is shown to be a VC class of sets in Pollard (1984, Section II.4). The previously described core-determining sets are unions of such sets. The class of sets $\mathscr{D}_{3,k}$ involves functionals of the conditional distribution of Y given X as well as the marginal distribution of X. The VC property described in Restriction I3 should be viewed as a restriction on the family of distributions in addition to the smoothness conditions described in Restriction I1.

Using properties of VC classes of sets (e.g, Pakes and Pollard (1989, Lemma 2.5)), the following are also VC classes of sets by Restriction I3, $\mathscr{D}_{4,k} \equiv \{w \in \mathcal{W}\colon\ T_k(w;\theta) \geq c \ \text{ for some } \theta \in \Theta,\, c \in \mathbb{R}\}$, $\mathscr{D}_{5,k} \equiv \{w \in \mathcal{W}\colon\ c_1 \leq T_k(w;\theta) < c_2 \ \text{ for some } \theta \in \Theta,\, c_1 < c_2 \text{ in } \mathbb{R}\}$. Restriction I3 provides sufficient conditions for the classes of functions that produce the relevant empirical process in our problem to be *Euclidean* (as defined in Nolan and Pollard (1987, Definition 8) and Pakes and Pollard (1989, Definition 2.7)). The Euclidean property will suffice for these empirical processes to be *manageable* (as defined in Pollard (1990, Definition 7.9)) and will allow us to use the maximal inequality results in Sherman (1994) in our analysis of the asymptotic properties of $\widehat{R}(\theta)$. Let

$$\widetilde{R}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\left(\sum_{k=1}^{K}\widehat{T}_k(w_i;\theta)\cdot 1\{T_k(w_i;\theta)\geq 0\}\right), \tag{24}$$

which is equivalent to $\widehat{R}(\theta)$ but for the replacement of $1\left[\widehat{T}_k(w_i;\theta) \geq -b_n\right]$ with $1\left[T_k(w_i,\theta) \geq 0\right]$. Next, let $\xi_{k,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left|\widehat{T}_k(w_i;\theta)\right|\cdot\left|1[\widehat{T}_k(w_i;\theta)\geq -b_n] - 1\left[T_k(w_i;\theta)\geq 0\right]\right|$, and note that $\left|\widetilde{R}(\theta)-\widehat{R}(\theta)\right| \leq \left|\xi_{k,n}(\theta)\right|$. Using maximal inequality results in Sherman (1994), in the Econometric Supplement we first show that under Restrictions I1-I3,

$$\xi_{k,n}(\theta) \leq \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right)\right)\times\frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot 1\Big[-2b_n \leq T_k(w_i;\theta) < 0\Big] + o_p(n^{-1/2-\Delta}) \quad \forall\, \Delta > 0,$$

As we can see, the probability of the event $-2b_n \leq T_k(W;\theta) < 0$ plays a key role. This probability should go to zero as $b_n \searrow 0$. The following condition determines this.

Restriction I4 (Behavior of $T_k(W;\theta)$ at zero from below): There are constants $\overline{b} > 0$ and $\overline{A} < \infty$ such that, for all $0 < b \leq \overline{b}$ and each $k = 1,\ldots,K$:

$$\sup_{\theta\in\Theta} E\Big[\omega(X)\cdot 1\left[-b \leq T_k(W;\theta) < 0\right]\Big] \leq b\overline{A}. \quad ■$$

Restriction I4 is a mild requirement. It suffices for example that, for each $\theta \in \Theta$ and conditional on $W \in \mathcal{W}^*$, the density of $T_k(W,\theta)$ be bounded by some finite $\overline{A}$ in the range $0 > T_k(W;\theta) \geq -\overline{b}$ for some fixed but arbitrarily small $\overline{b} > 0$. This allows for $P(T_k(W;\theta) = 0) > 0$, and for the distribution of $T_k(W;\theta)$ to have other mass points. Figure 1 provides three illustrations to further clarify which types of discontinuities in the distribution of $T_k(W,\theta)$ are permitted. From here, using Sherman (1994, Main Corollary), we show in the Econometric Supplement that Restrictions I1-I4 yield $\sup_\theta \left|\frac{1}{n}\sum_{i=1}^n 1\{-2b_n \leq T_k(w_i,\theta) < 0\}\right| = O_p(b_n)$, and consequently that $\sup\limits_{\theta} \left|\xi_{k,n}(\theta)\right| = o_p(n^{-1/2-\epsilon})$, where $\epsilon > 0$ is as described in Restriction I2. From here we obtain the following Lemma.

Lemma 2 *Let Restrictions I1-I4 hold. Then,* $\sup_{\theta\in\Theta}\left|\widetilde{R}(\theta) - \widehat{R}(\theta)\right| = o_p(n^{-1/2-\epsilon})$, *where* $\epsilon > 0$ *is as described in Restriction I2.*

The proof of Lemma 2 and all the auxiliary results can be found in the Econometric Supplement.

5.2.2 An asymptotic linear representation result for $\widehat{R}(\theta)$

Let

$$v_k(w_\ell, w_i;\theta,h) \equiv \omega(x_i)\cdot\omega(x_\ell)\cdot m_k(y_\ell,y_i,x_i;\theta)\cdot \mathbf{K}(x_i - x_\ell;h)\cdot 1[T_k(w_i,\theta)\geq 0]. \tag{25}$$

From Lemma 2 and equation (23), we have

$$\widehat{R}(\theta) = \frac{1}{n^2\cdot h_n^z}\sum_{i=1}^n\sum_{\ell=1}^n v_k(w_\ell,w_i;\theta,h_n) + \vartheta_n(\theta), \tag{26}$$

where $\sup\limits_{\theta\in\Theta}|\vartheta_n(\theta)| = o_p(n^{-1/2-\epsilon})$. The first term in (26) is a U-process of order 2. Under Restrictions I1-I4 a linear representation for $\widehat{R}(\theta) - R(\theta)$ follows from (26) if we employ a Hoeffding projection (see Serfling (1980, pages 177-178) or Sherman (1994, equations (6)-(7))) and the results from Sherman (1994). The expressions for the Hoeffding projection and the resulting linear representation can be simplified if we add the following smoothness condition.

Restriction I5 For all $w \in \mathcal{W}^*$ and $\theta\in\Theta$ let $\lambda_{2k}(w;\theta) \equiv E[m_k(Y,y,x;\theta)\cdot 1[T_k(W;\theta)\geq 0]\,|X = x]$ be M-times continuously differentiable with respect to x^c almost surely, with bounded derivatives, where M is as described in Restriction I1. ■

Theorem 3 *Let*

$$g_k(w_i;\theta) \equiv \omega(x_i)\cdot(T_k(w_i;\theta))_+ + \omega(x_i)^2\cdot\lambda_{2k}(w_i;\theta)\cdot f_X(x_i),$$

$$\psi_R(w_i;\theta) \equiv \sum_{k=1}^K\Big(g_k(w_i;\theta) - E[g_k(W;\theta)]\Big)$$

Let Restrictions I1-I5 hold. Then,

$$\widehat{R}(\theta) = R(\theta) + \frac{1}{n}\sum_{i=1}^n \psi_R(w_i;\theta) + \varepsilon_n(\theta), \text{ where } \sup_{\theta\in\Theta}|\varepsilon_n(\theta)| = o_p\left(n^{-1/2-\epsilon}\right),$$

where $\epsilon > 0$ is described in Restriction I2.

The proof of Theorem 3 can be found in the Econometric Supplement. It follows from the Hoeffding projection of the U-process in (26), results from Sherman (1994), and the smoothness conditions in Restriction I5.

5.2.3 Contact sets and the asymptotic properties of $\widehat{R}(\theta)$

Let Θ^I denote the set of parameter values that satisfy our inequalities w.p.1 over our inference range:

$$\Theta^I \equiv \Big\{\theta \in \Theta : P\left[T_k(W;\theta) \leq 0\right] = 1 \ \text{ for } k = 1,\ldots,K\Big\}.$$

For any $\theta \in \Theta$ we define here the contact set as $\{w \in \mathcal{W}^*\colon T_k(w;\theta) \geq 0 \ \text{ for some } k = 1,\ldots,K\}$. Note that if $\theta \in \Theta^I$, the contact set is $\{w \in \mathcal{W}^*\colon T_k(w;\theta) = 0 \ \text{ for some } k = 1,\ldots,K\}$. Partition Θ^I into two subsets depending on the (probability) measure of the contact sets. Let

$$\begin{aligned} \overline{\Theta}^I &\equiv \Big\{\theta \in \Theta^I\colon P\left(T_k(W;\theta) = 0\right) > 0 \ \text{ for some } k = 1,\ldots,K\Big\}, \\ \Theta_0^I &\equiv \Big\{\theta \in \Theta^I\colon P\left[T_k(W;\theta) < 0\right] = 1 \ \text{ for each } k = 1,\ldots,K\Big\}. \end{aligned} \tag{27}$$

The set $\overline{\Theta}^I$ contains all elements of Θ^I whose contact sets have positive measure, while Θ_0^I contains those for which the contact set has measure zero. By construction, $\Theta^I = \overline{\Theta}^I \cup \Theta_0^I$. Consider the influence function ψ_R defined in Theorem 3 and let $\sigma_R^2(\theta) \equiv Var\left(\psi_R(W;\theta)\right)$. Inspection of $\psi_R(\cdot;\theta)$ shows (i) $\forall\theta \in \Theta$: $E\left[\psi_R(W;\theta)\right] = 0$, (ii) $\forall\theta \in \Theta_0^I$: $\psi_R(W;\theta) = 0$ wp1 and therefore $\sigma_R^2(\theta) = 0$, and (iii) $\forall\theta \notin \Theta_0^I$: $\sigma_R^2(\theta) > 0$. Thus, from Theorem 3,

$$\begin{aligned} &(A) \quad \theta \notin \Theta^I \implies R(\theta) > 0, \sigma_R^2(\theta) > 0 \quad \implies \quad n^{1/2}\cdot\widehat{R}(\theta) \xrightarrow{p} +\infty, \\ &(B) \quad \theta \in \overline{\Theta}^I \implies R(\theta) = 0, \sigma_R^2(\theta) > 0 \quad \implies \quad n^{1/2}\cdot\widehat{R}(\theta) \xrightarrow{d} \mathcal{N}\left(0,\sigma_R^2(\theta)\right), \\ &(C) \quad \theta \in \Theta_0^I \implies R(\theta) = 0, \sigma_R^2(\theta) = 0 \quad \implies \quad n^{1/2}\cdot\widehat{R}(\theta) \xrightarrow{p} 0\,. \end{aligned} \tag{28}$$

The asymptotic behavior of $\widehat{R}(\theta)$ automatically adapts to the measure of the contact sets.

5.2.4 The role played by the tuning parameter b_n

Our use of b_n leads to Lemma 2 and (26). To illustrate this, suppose we drop b_n and use instead

$$\ddot{R}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left(\sum_{k=1}^{K}\widehat{T}_k(w_i;\theta)\cdot 1\left[\widehat{T}_k(w_i,\theta) \geq 0\right]\right) = \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\sum_{k=1}^{K}\left(\widehat{T}_k(w_i,\theta)\right)_+$$

This estimator for $R(\theta)$ directly replaces $(T_k(w_i,\theta))_+$ with $(\widehat{T}_k(w_i,\theta))_+$. Note that,

$$1[\widehat{T}_k(w_i,\theta) \geq 0] \neq 1[T_k(w_i,\theta) \geq 0] \iff \begin{cases} (A)\ \widehat{T}_k(w_i,\theta) \geq 0 \ \text{ and } \ T_k(w_i,\theta) < 0, \ \text{ or} \\ (B)\ \widehat{T}_k(w_i,\theta) < 0 \ \text{ and } \ T_k(w_i,\theta) \geq 0 \end{cases}$$

With the use of the sequence b_n, the relevant event is $1[\widehat{T}_k(w_i,\theta) \geq -b_n] \neq 1[T_k(w_i,\theta) \geq 0]$, and

$$1[\widehat{T}_k(w_i,\theta) \geq -b_n] \neq 1[T_k(w_i,\theta) \geq 0] \iff \begin{cases} (C)\ \widehat{T}_k(w_i,\theta) \geq -b_n \ \text{ and } \ T_k(w_i,\theta) < 0, \ \text{ or} \\ (D)\ \widehat{T}_k(w_i,\theta) < -b_n \ \text{ and } \ T_k(w_i,\theta) \geq 0 \end{cases}$$

Under the assumptions leading to Lemma 2, the probabilities of the events (C) and (D) vanish at the appropriate rate. The same type of assumptions would also take care of the probability of the event (A). However, since $T_k(w_i,\theta)$ is allowed to have a point-mass at zero, *the probability of the event (B) does not necessarily vanish*. Consequently, the analogous result to (26) is:

$$\ddot{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \omega(x_i)\cdot \widehat{T}_k(w_i;\theta)\cdot 1\Big[T_k(w_i,\theta) \geq 0,\, \widehat{T}_k(w_i,\theta) \geq 0\Big] + \varsigma_n(\theta),$$

where $\sup_{\theta\in\Theta} |\varsigma_n(\theta)| = o_p(n^{-1/2})$. Thus, if we employ $\ddot{R}(\theta)$ we must study the asymptotic properties of $1[T_k(w_i,\theta) \geq 0,\, \widehat{T}_k(w_i,\theta) \geq 0]$. This is a more complicated problem that generally requires direct estimation of contact sets (see Lee, Song, and Whang (2018)) in a first step, which also requires the use of tuning parameters. Our statistic uses b_n directly in its construction in order to asymptotically adapt to the measure of the contact set for each θ without an explicit first step.

5.3 Constructing an inferential statistic for $\boldsymbol{\theta}$

We combine the linear representation for $\widehat{R}(\theta)$ given by Theorem 3 with the linear representation of the ML estimator described in Theorem 2 for θ_1. Recall from the parametric assumptions in Section 4.2 that $F(u_1,u_2;\lambda) \equiv G(u_1)\cdot G(u_2)\cdot[1+\lambda(1-G(u_1))(1-G(u_2))]$ is the assumed joint CDF of (U_1,U_2), with $G(u) \equiv \exp(u)/(1+\exp(u))$. There, we denoted $\alpha \equiv \eta - \delta$, $\widetilde{\beta} \equiv (\alpha,\beta')'$, $\theta_1 \equiv (\widetilde{\beta}',\lambda)'$ and $Z_j \equiv (1,-X_j)$ for $j = 1,2$. Denote $r \equiv \dim(\theta_1)$.

In (18) we showed that under the PSNE restriction, $P(Y = (0,0)|X) = F\Big(Z_1\widetilde{\beta}, Z_2\widetilde{\beta};\lambda\Big)$. Denote

$$\Theta^* \equiv \Big\{\theta \in \Theta\colon \theta \in \Theta^I \ \wedge \ P(Y = (0,0)|X) = F\Big(Z_1\widetilde{\beta}, Z_2\widetilde{\beta};\lambda\Big)\Big\}. \tag{29}$$

as the set of parameter values satisfying the likelihood equation for the event $Y = (0,0)$ together with the conditional moment inequalities that define Θ_I. Our goal is to perform inference on values of $\theta \in \Theta^*$. Recall from Theorem 2 that θ_1^*, is point-identified and uniquely maximizes

$E[\ell(\theta_1;W)]$.[19] We proposed to consistently estimate θ_1^* by the ML estimator $\widehat{\theta}_1 \equiv (\widehat{\widetilde{\beta}}', \widehat{\lambda})'$, whose asymptotic properties were described in Theorem 2. Let

$$\psi_M(w_i) \equiv -\left(E\left[\frac{\partial^2 \ell(\theta_1^*, W)}{\partial\theta_1 \partial\theta_1'}\right]\right)^{-1} \times \frac{\partial \ell(\theta_1^*, w_i)}{\partial\theta_1}.$$

From Theorem 2, we have $E[\psi_M(W)] = 0$, $Var[\psi_M(W)] = E\left[\frac{\partial\ell(\theta_1^*,W)}{\partial\theta_1}\cdot\frac{\partial\ell(\theta_1^*,W)}{\partial\theta_1}'\right]^{-1} \equiv H_0^{-1}$ and $\widehat{\theta}_1 = \theta_1^* + \frac{1}{n}\sum_{i=1}^n \psi_M(w_i) + \varepsilon_n^M$, where $\|\varepsilon_n^M\| = o_p\left(n^{-1/2}\right)$. We estimate the influence function ψ_M with $\widehat{\psi}_M(w_i) \equiv \widehat{H}_0^{-1}\cdot\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}$, where $\widehat{H}_0 \equiv -\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial^2\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1\partial\theta_1'}\right)$. Our first step will be to impose restrictions that produce precise asymptotic properties for $\widehat{\psi}_M$. Denote $H(t_1,t_2;\lambda) \equiv F(t_1,t_2;\lambda)\cdot(1-F(t_1,t_2;\lambda))$, and let $\phi_1(t_1,t_2;\lambda) \equiv \frac{\partial F(t_1,t_2;\lambda)}{\partial t_1}\cdot H(t_1,t_2;\lambda)^{-1}$, $\phi_2(t_1,t_2;\lambda) \equiv \frac{\partial F(t_1,t_2;\lambda)}{\partial t_2}\cdot H(t_1,t_2;\lambda)^{-1}$, and $\phi_3(t_1,t_2;\lambda) \equiv \frac{\partial F(t_1,t_2;\lambda)}{\partial \lambda}\cdot H(t_1,t_2;\lambda)^{-1}$. We have

$$\frac{\partial\ell(\theta_1;w)}{\partial\theta_1} = \begin{pmatrix} \left(z_1\cdot\phi_1(z_1\widetilde{\beta},z_2\widetilde{\beta};\lambda) + z_2\cdot\phi_2(z_1\widetilde{\beta},z_2\widetilde{\beta};\lambda)\right)\cdot\left(1[y=(0,0)] - F(z_1\widetilde{\beta},z_2\widetilde{\beta};\lambda)\right) \\ \phi_3(z_1\widetilde{\beta},z_2\widetilde{\beta};\lambda)\cdot\left(1[y=(0,0)] - F(z_1\widetilde{\beta},z_2\widetilde{\beta};\lambda)\right) \end{pmatrix}$$

Restriction I6 There exists a $D(w) \geq 0$ such that $E[\|Z_1\|^q\cdot\|Z_2\|^r\cdot D(W)] < \infty$ for all $q,r \in \mathbb{N}\cup\{0\} : 0 \leq q+r \leq 3$, and a neighborhood $\mathcal{N}$ that contains θ_1^* such that the following conditions hold.

(i) There exists a $\underline{d} > 0$ such that, $\left|\det\left(E\left[\frac{\partial^2\ell(\theta_1,W)}{\partial\theta_1\partial\theta_1'}\right]\right)\right| \geq \underline{d}$ for all $\theta_1 \in \mathcal{N}$.

(ii) Let $\delta_{\ell,m}(y,t_1,t_2;\lambda) \equiv \frac{\partial\phi_\ell(t_1,t_2;\lambda)}{\partial t_m}\cdot\left(1[y=(0,0)] - F(t_1,t_2;\lambda)\right) - \phi_\ell(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial t_m}$. For $(\ell,m) \in \{1,2\}\times\{1,2\}$ and $j = 1,2$, we have $\left\|\delta_{\ell,m}(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)\right\| \leq D(W)$, $\left\|\frac{\partial\delta_{\ell,m}(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial t_j}\right\| \leq D(W)$ and $\left\|\frac{\partial\delta_{\ell,m}(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial\lambda}\right\| \leq D(W)$ for all $\theta_1 \in \mathcal{N}$. Every element of $\left[(Z_\ell Z_m')\cdot\delta_{\ell,m}(Y,Z_1\widetilde{\beta}^*,Z_2\widetilde{\beta}^*;\lambda)\right]$ has finite variance.

(iii) Let $\eta_\ell(y,t_1,t_2;\lambda) \equiv \frac{\partial\phi_\ell(t_1,t_2;\lambda)}{\partial\lambda}\cdot\left(1[y=(0,0)] - F(t_1,t_2;\lambda)\right) - \phi_\ell(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial\lambda}$. For $\ell \in \{1,2\}$ and $j = 1,2$, we have $\left\|\eta_\ell(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)\right\| \leq D(W)$, $\left\|\frac{\partial\eta_\ell(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial t_j}\right\| \leq D(W)$ and $\left\|\frac{\partial\eta_\ell(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial\lambda}\right\| \leq D(W)$ for all $\theta_1 \in \mathcal{N}$. Every element of $\left[Z_\ell\cdot\eta_\ell(Y,Z_1\widetilde{\beta}^*,Z_2\widetilde{\beta}^*;\lambda)\right]$ has finite variance.

(iv) Let $\Upsilon(y,t_1,t_2;\lambda) \equiv \frac{\partial\phi_3(t_1,t_2;\lambda)}{\partial\lambda}\cdot\left(1[y=(0,0)] - F(t_1,t_2;\lambda)\right) - \phi_3(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial\lambda}$. For $j \in \{1,2\}$, we have $\left|\Upsilon(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)\right| \leq D(W)$, $\left\|\frac{\partial\Upsilon(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial t_j}\right\| \leq D(W)$ and $\left\|\frac{\partial\Upsilon(Y,Z_1\widetilde{\beta},Z_2\widetilde{\beta};\lambda)}{\partial\lambda}\right\| \leq D(W)$ for all $\theta_1 \in \mathcal{N}$. Every element of $\left[Z_\ell\cdot\Upsilon(Y,Z_1\widetilde{\beta}^*,Z_2\widetilde{\beta}^*;\lambda)\right]$ has finite variance . ■

[19] The remaining parameter subvector in θ is $\theta_2 \equiv (\eta,\Delta_1,\Delta_2)$, which is restricted only by the conditional moment inequalities described previously.

Assuming that $\mathcal{X}$ is bounded would suffice for all the conditions in Restriction I6 to hold for a constant envelope D, but it is not necessary. Combined with Theorem 2, Restriction I6 yields the result, $\widehat{\theta}_1 = \theta_1^* + \frac{1}{n}\sum_{i=1}^n \psi_M(w_i) + \varepsilon_n^M$, where $\left\|\varepsilon_n^M\right\| = O_p\left(\frac{1}{n}\right)$.

Thus, Restriction I6 refines the linear representation from Theorem 2 by establishing that $\left\|\varepsilon_n^M\right\| = O_p\left(\frac{1}{n}\right)$. Restriction I6 also ensures $\left\|\widehat{H}_0^{-1} - H_0^{-1}\right\| = O_p(n^{-1/2})$ and yields precise asymptotic properties for $\widehat{\psi}_M$ which will be relevant below. Combining the asymptotic linear representation of $\widehat{\theta}_1$ with Theorem 3, we have that for any $\theta \in \Theta$,

$$\begin{aligned}
&\underbrace{\widehat{V}(\theta)}_{(r+1)\times 1} \equiv n^{1/2}\begin{pmatrix} \widehat{\theta}_1 - \theta_1 \\ \widehat{R}(\theta) \end{pmatrix} = n^{1/2}\begin{pmatrix} \theta_1^* - \theta_1 \\ R(\theta) \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix} + \begin{pmatrix} \xi_n^M \\ \xi_n^R(\theta) \end{pmatrix}, \\
&\text{where} \quad \left\|\xi_n^M\right\| = O_p(n^{-1/2}), \quad \text{and} \quad \sup_{\theta\in\Theta}\left|\xi_n^R(\theta)\right| = o_p(n^{-\epsilon}),
\end{aligned} \tag{30}$$

where $\epsilon > 0$ is as described in Restriction I2. Let $\Sigma_{MR}(\theta) \equiv E\left[\psi_M(W)\psi_R(W;\theta)\right]$. We have

$$Var\begin{bmatrix} \psi_M(W) \\ \psi_R(W;\theta) \end{bmatrix} = \begin{pmatrix} H_0^{-1} & \Sigma_{MR}(\theta) \\ \Sigma_{MR}(\theta)' & \sigma_R^2(\theta) \end{pmatrix} \equiv \Sigma(\theta).$$

Our estimator for influence function ψ_R defined in Theorem 3 is

$$\widehat{\psi}_R(w_i;\theta) \equiv \sum_{k=1}^K \left(\widehat{g}_k(w_i;\theta) - \widehat{E}[g_k(W;\theta)]\right),$$

where for any $w \in \mathcal{W}$:

$$\widehat{g}_k(w;\theta) = \frac{1}{n}\sum_{j=1}^n \frac{\widetilde{v}_k(w,w_j;\theta,h_n)}{h_n^z}, \quad \text{and} \quad \widehat{E}[g_k(W;\theta)] = \binom{n}{2}^{-1}\sum_{i<j}\frac{\widetilde{v}_k(w_i,w_j;\theta,h_n)}{h_n^z},$$

with $\widetilde{v}_k(w_1,w_2;\theta,h) \equiv v_k(w_1,w_2;\theta,h) + v_k(w_2,w_1;\theta,h)$, with v_k as defined in (25). We estimate $\widehat{\Sigma}_{MR}(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}_M(w_i)\widehat{\psi}_R(w_i;\theta)$ and $\widehat{\sigma}_R^2(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}_R(w_i;\theta)^2$. In Section 3.1 of the Econometric Supplement we show that under the conditions of Theorems 2 and 3, and Restriction I6:

$$\sup_{\theta\in\Theta}\left\|\widehat{\Sigma}_{MR}(\theta) - \Sigma_{MR}(\theta)\right\| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right), \quad \sup_{\theta\in\Theta}\left|\widehat{\sigma}_R^2(\theta) - \sigma_R^2(\theta)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right). \tag{31}$$

Now partition Θ^* into two subsets, $\overline{\Theta}^* \equiv \Theta^* \cup \overline{\Theta}^I$ and $\Theta_0^* \equiv \Theta^* \cup \Theta_0^I$. The set $\overline{\Theta}^*$ contains all the elements in Θ^* for which the contact sets have positive measure, while Θ_0^* contains those for which the contact sets have measure zero. Note that $\psi_R(W;\theta) = 0$ w.p.1 for all $\theta \in \Theta_0^*$, and therefore $\Sigma_{MR}(\theta) = 0$ and $\sigma_R^2(\theta) = 0$ for all $\theta \in \Theta_0^*$. Thus, using $\widehat{V}(\theta)$ to construct a Wald-type statistic with pivotal asymptotic properties requires some form of regularization for our estimator of $\Sigma(\theta)$. The fact that $\widehat{R}(\theta)$ is a scalar facilitates this. Our proposal is to replace $\widehat{\sigma}^2_R(\theta)$ with $\widehat{\Sigma}_R(\theta) \equiv \max\left\{\widehat{\sigma}^2_R(\theta), \kappa_n\right\}$, where $\kappa_n \searrow 0$ is a decreasing sequence of nonnegative constants satisfying Restriction I7 below. Note from (31) that, under the conditions in Theorem 3, $\sup_{\theta\in\Theta}\left|\widehat{\Sigma}_R(\theta) - \sigma^2_R(\theta)\right| = o_p(1)$. Our proposal is to do inference on θ based on the statistic

$$\widehat{Q}(\theta) \equiv \widehat{V}(\theta)'\widehat{\Sigma}(\theta)^{-1}\widehat{V}(\theta), \quad \text{where} \quad \widehat{\Sigma}(\theta) = \begin{pmatrix} \widehat{H}_0^{-1} & \widehat{\Sigma}_{MR}(\theta) \\ \widehat{\Sigma}_{MR}(\theta)' & \widehat{\Sigma}_R(\theta) \end{pmatrix} \tag{32}$$

We construct a confidence set (CS) for θ based on the properties of $\widehat{Q}$ under the conditions in Theorems 2 and 3 as well as the following additional restriction.

Restriction I7

(i) The sequence $\kappa_n \searrow 0$ satisfies $n^{1/2} \cdot h_n^z \cdot \kappa_n \longrightarrow \infty$ and $n^{2\epsilon} \cdot \kappa_n \longrightarrow \infty$ as $n \longrightarrow \infty$, where $\epsilon > 0$ is as described in Restriction I2.

(ii) $E[||\psi_M(W)||^3] < \infty$ and $\sigma_R(\theta) \geq \underline{\sigma} > 0$ for all $\theta \in \Theta\backslash\Theta_0^I$. For each $\theta \in \Theta\backslash\Theta_0^I$ define $\psi_{MR}(W;\theta) \equiv -\Sigma_{MR}(\theta)'H_0\psi_M(W) + \psi_R(W;\theta)$, and let $\sigma^2_{MR}(\theta) \equiv Var(\psi_{MR}(W;\theta))$. Note that

$$\sigma^2_{MR}(\theta) = \sigma^2_R(\theta) - \Sigma_{MR}(\theta)'H_0\Sigma_{MR}(\theta) = \sigma^2_R(\theta) \times \left(1 - \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)' H_0 \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)\right)$$

We assume that $\left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)' H_0 \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right) \leq \overline{d} < 1$ for all $\theta \in \Theta \setminus \Theta_0^I$. Note that this implies that $\sigma^2_{MR}(\theta) \geq \underline{\sigma}^2(1 - \overline{d}) > 0$ for all $\theta \in \Theta \setminus \Theta_0^I$. Finally, we assume that $\sigma_R(\theta)$ and $\Sigma_{MR}(\theta)$ are continuous everywhere on $\Theta \setminus \Theta_0^I$. ■

The conditions in Restriction I7(ii) are aimed at satisfying multivariate Berry-Esseen bounds (see, e.g, Götze (1991) or Raič (2019)). Suppose $U_1, \ldots, U_n$ are i.i.d random vectors in $\mathbb{R}^\ell$ such that $E[U_i] = 0$ and $Var(U_i) = I_\ell$ (the $\ell \times \ell$ identity matrix) Suppose $E\left[||U_i||^3\right] < \infty$, and let $W_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i$. Note that $Var(W_n) = I_\ell$. Let Φ_ℓ denote the standard Gaussian distribution in $\mathbb{R}^\ell$ and for any measurable set $A \subseteq \mathbb{R}^\ell$, let $\Phi_\ell(A) \equiv P(Z \in A)$, where $Z \sim \Phi_\ell$. Theorem 1.1 in Raič (2019) states that, for all measurable convex sets $A \subseteq R^\ell$, we have $\left|P(W_n \in A) - \Phi_\ell(A)\right| \leq \left(42 \cdot \ell^{1/4} + 16\right) \cdot \frac{1}{\sqrt{n}} \cdot E\left[||U_i||^3\right]$. Note that the bound does not depend on A. For any $c > 0$ let $A_c \equiv \{z\colon z'z \leq c\}$. Note that A_c is convex and that $\Phi_\ell(A_c) = F_{\chi^2_\ell}(c)$, where $F_{\chi^2_\ell}$ is the χ^2_ℓ cdf. It follows that $\left|P(W_n'W_n \leq c) - F_{\chi^2_\ell}(c)\right| \leq \left(42 \cdot \ell^{1/4} + 16\right) \cdot \frac{1}{\sqrt{n}} \cdot E\left[||U_i||^3\right] \quad \forall\, c > 0$, and therefore, $\sup_{c>0}\left|P(W_n'W_n \leq c) - F_{\chi^2_\ell}(c)\right| \to 0$.

Berry-Esseen bounds will be relevant in our case for two processes. Let M be an invertible matrix such that $H_0^{-1} = MM'$. Since H_0^{-1} is a positive-definite variance matrix, such M exists, see e.g, in Lehmann (1999, page 306) or Ruud (2000, Lemma 7.6). Let $\overline{\psi}_M(W) \equiv M^{-1}\psi_M(W)$ and note that $E[\overline{\psi}(W)] = 0$ and $Var(\overline{\psi}(W)) = I_r$. The first process is described as follows. Let $S_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}_M(w_i)$ and note that $E[S_n] = 0$ and $Var(S_n) = I_r$. Restriction I7 implies $E[\|\psi_M(W)\|^3] < \infty$. From here, the multivariate Berry-Esseen bounds in Raič (2019, Theorem 1.1) yields

$$\forall\ c > 0 \quad \left|P\left(S_n' S_n \le c\right) - F_{\chi^2_r}(c)\right| \le \left(42 \cdot r^{1/4} + 16\right) \cdot \frac{1}{\sqrt{n}} \cdot E\left[\left\|\overline{\psi}_M(W)\right\|^3\right]. \tag{33}$$

The second relevant process is the following. For each $\theta \in \Theta \setminus \Theta_0^I$, let $\psi_{MR}(W;\theta)$ and $\sigma^2_{MR}(\theta)$ be as described in Restriction I7 and let $\overline{\psi}_{MR}(W;\theta) \equiv \frac{\psi_{MR}(W;\theta)}{\sigma_{MR}(\theta)}$ and $\overline{\psi}(W;\theta) \equiv \left(\overline{\psi}_M(W)' \quad \overline{\psi}_{MR}(W;\theta)\right)'$, and let $T_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}_M(w_i)$. Note that $E[T_n(\theta)] = 0$ and $Var(T_n(\theta)) = I_{r+1}$ for all $\theta \in \Theta \setminus \Theta_0^I$. Furthermore, we have

$$T_n(\theta)' T_n(\theta) = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix}' \Sigma^{-1}(\theta) \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix}$$

for each $\theta \in \Theta \setminus \Theta_0^I$. By Restriction I7, there exists a $\overline{D}_3 < \infty$ such that $E[\left\|\overline{\psi}(W;\theta)\right\|^3] \le \overline{D}_3$ for all $\theta \in \Theta \setminus \Theta_0^I$. Thus the the conditions for the multivariate Berry-Esseen bounds in Raič (2019, Theorem 1.1) apply and we obtain

$$\forall\ \theta \in \Theta \setminus \Theta_0^I, c > 0 \quad \left|P\left(T_n(\theta)' T_n(\theta) \le c\right) - F_{\chi^2_{r+1}}(c)\right| \le \left(42(r+1)^{1/4} + 16\right) \cdot \frac{1}{\sqrt{n}} \cdot \overline{D}_3. \tag{34}$$

Let $\widehat{\sigma}^2_{MR}(\theta) \equiv \widehat{\Sigma}_R(\theta) - \widehat{\Sigma}_{MR}(\theta)' \widehat{H}_0 \widehat{\Sigma}_{MR}(\theta)$. Using partitioned-matrix inverse properties, we have

$$\widehat{\Sigma}(\theta)^{-1} = \begin{pmatrix} \widehat{H}_0 + \frac{\widehat{H}_0 \widehat{\Sigma}_{MR}(\theta)\widehat{\Sigma}_{MR}(\theta)' \widehat{H}_0}{\widehat{\sigma}^2_{MR}(\theta)} & \frac{-\widehat{H}_0 \widehat{\Sigma}_{MR}(\theta)}{\widehat{\sigma}^2_{MR}(\theta)} \\ \frac{-\widehat{\Sigma}_{MR}(\theta)' \widehat{H}_0}{\widehat{\sigma}^2_{MR}(\theta)} & \frac{1}{\widehat{\sigma}^2_{MR}(\theta)} \end{pmatrix}$$

5.3.1 Asymptotic properties of $\widehat{Q}(\theta)$ over Θ_0^*

The set Θ_0^* includes all parameter values in Θ^* for which the contact sets have measure zero. This is where our regularization becomes relevant. Recall that $\Sigma_{MR}(\theta) = 0$ for all $\theta \in \Theta_0^I$, and therefore $\sup_{\theta \in \Theta_0^I} \left\|\widehat{\Sigma}_{MR}(\theta)\right\| = O_p\left(\frac{1}{n^{1/2} \cdot h_n^z}\right)$ from (31). Next, note that $\widehat{\Sigma}_R(\theta) \ge \kappa_n$ for all θ. Therefore,

$$\sup_{\theta \in \Theta_0^I} \left\|\frac{\widehat{\Sigma}_{MR}(\theta)}{\widehat{\Sigma}_R(\theta)}\right\| = O_p\left(\frac{1}{n^{1/2} \cdot h_n^z \cdot \kappa_n}\right) \quad \text{and} \quad \sup_{\theta \in \Theta_0^I} \left|\frac{\left\|\widehat{\Sigma}_{MR}(\theta)\right\| \cdot \left\|\widehat{\Sigma}_{MR}(\theta)\right\|}{\widehat{\Sigma}_R(\theta)}\right| = O_p\left(\frac{1}{n \cdot h_n^{2z} \cdot \kappa_n}\right).$$

Note that $n^{1/2}\cdot h_n^z\cdot\kappa_n\to\infty$ and $n\cdot h_n^{2z}\cdot\kappa_n\to\infty$ by Restrictions I2 and I7. Therefore, for any $\theta\in\Theta_0^I$,

$$\begin{aligned}&\widehat{\Sigma}(\theta)^{-1}=\begin{pmatrix}H_0+\xi_n^{11}(\theta) & \xi_n^{12}(\theta)\\ \xi_n^{12}(\theta)' & \xi_n^{22}(\theta),\end{pmatrix},\quad\text{where}\\ &\sup_{\theta\in\Theta_0^I}\left\|\xi_n^{11}(\theta)\right\|=O_p\left(\frac{1}{nh_n^{2z}\kappa_n}\right),\quad \sup_{\theta\in\Theta_0^I}\left\|\xi_n^{12}(\theta)\right\|=O_p\left(\frac{1}{n^{1/2}h_n^{z}\kappa_n}\right),\quad \sup_{\theta\in\Theta_0^I}\left|\xi_n^{22}(\theta)\right|=O_p\left(\frac{1}{\kappa_n}\right)\end{aligned}\tag{35}$$

Next, note from (30), $\widehat{V}(\theta)=\left(\frac{1}{\sqrt{n}}\psi_M(w_i)'+\xi_n^{M\prime}\quad \xi_n^R(\theta)\right)'\ \forall\ \theta\in\Theta_0^*$, where $\left\|\xi_n^M\right\|=O_p(n^{-1/2})$, and $\sup_{\theta\in\Theta}\left|\xi_n^R(\theta)\right|=o_p(n^{-\epsilon})$, where $\epsilon>0$ is as described in Restriction I2. Since $\Theta_0^*\subseteq\Theta_0^I$, our previous results combined with (35) yield

$$\forall\ \theta\in\Theta_0^*\quad \widehat{Q}(\theta)=S_n'S_n+\rho_n^Q(\theta)\,,\quad\text{where}\ \sup_{\theta\in\Theta_0^*}\left|\rho_n^Q(\theta)\right|=o_p\left(\frac{1}{n^{1/2}\cdot h_n^z\cdot\kappa_n}\right)+o_p\left(\frac{1}{n^{2\epsilon}\cdot\kappa_n}\right)=o_p(1),\tag{36}$$

From here, (33) yields $\sup_{\theta\in\Theta_0^*}\left|P\left(\widehat{Q}(\theta)\le c\right)-F_{\chi_r^2}(c)\right|\to 0$.

Asymptotic properties of $\widehat{Q}(\theta)$ over $\overline{\Theta}^*$

The set $\overline{\Theta}^*$ includes all parameter values in Θ^* for which the contact sets have positive measure. In this case, the results in Theorems 2 and 3, and Restriction I7 yield

$$\widehat{Q}(\theta)=T_n(\theta)'T_n(\theta)+\xi_n^Q(\theta)\ \forall\ \theta\in\overline{\Theta}^*,\quad\text{where}\quad\sup_{\theta\in\Theta\setminus\Theta_0^I}\left|\xi_n^Q(\theta)\right|=o_p(1)\tag{37}$$

(note that $\overline{\Theta}^*\subseteq\Theta\setminus\Theta_0^I$). From here, (34) yields $\sup_{\theta\in\overline{\Theta}^*}\left|P\left(\widehat{Q}(\theta)\le c\right)-F_{\chi_{r+1}^2}(c)\right|\to 0$.

Asymptotic properties of $\widehat{Q}(\theta)$ when $\theta\notin\Theta^*$

There are two relevant cases when $\theta\notin\Theta^*$: (a) $\theta\in\Theta_0^I$ and (b) $\theta\in\Theta\setminus\Theta_0^I$. Case (a) corresponds to all parameter values $\theta\in\Theta_0^I$: $\theta_1\neq\theta_1^*$. In this case (36) is modified to,

$$\begin{aligned}&\widehat{Q}(\theta)=n(\theta_1^*-\theta_1)'\left(H_0+\xi_n^{11}(\theta)\right)(\theta_1^*-\theta_1)+S_n'S_n+\rho_n^Q(\theta),\\ &\text{where}\quad\sup_{\theta\in\Theta_0^I}\left|\xi_n^{11}(\theta)\right|=o_p(1),\quad\sup_{\theta\in\Theta_0^I}\left|\rho_n^Q(\theta)\right|=o_p(1).\end{aligned}\tag{38}$$

Since $\sup_{\theta\in\Theta_0^I}\left|\xi_n^{11}(\theta)\right|=o_p(1)$, we have that wp$\to$ 1, $H_0+\xi_n^{11}(\theta)$ is positive definite uniformly over Θ_0^I. Thus, $\forall\ d\neq 0$, $\inf_{\theta\in\Theta_0^I} d'\left(H_0+\xi_n^{11}(\theta)\right)d>0$ wp$\to$ 1. Therefore $P\left(\widehat{Q}(\theta)<c\right)\to 0$. Next, take

any sequence $\theta_n \in \Theta_0^I$ such that $n(\theta_{n,1} - \theta_1^*)' H_0(\theta_{n,1} - \theta_1^*) \to a > 0$. Then $\lim_{n\to\infty} P\left(\widehat{Q}(\theta_n) < c\right) = F_{\chi_r^2}(c-a)$.

Now consider case (b) in which $\theta \in \Theta \setminus \Theta_0^I$. This includes all parameter values $\theta \equiv (\theta_1', \theta_2')'$ such that $\sigma_R^2(\theta) > 0$ and either $\theta_1 \neq \theta_1^*$ or $R(\theta) \neq 0$ (or both). In this case (37) is modified to,

$$\widehat{Q}(\theta) = n\mu(\theta)'\left(\Sigma(\theta)^{-1} + \vartheta_n(\theta)\right)\mu(\theta) + T_n(\theta)'T_n(\theta) + \xi_n^Q(\theta), \quad \text{where} \quad \mu(\theta) \equiv \left((\theta_1 - \theta_1^*)' \quad R(\theta)\right)',$$
$$\sup_{\theta\in\Theta\setminus\Theta_0^I} |\vartheta_n(\theta)| = o_p(1), \quad \text{and} \quad \sup_{\theta\in\Theta\setminus\Theta_0^I} \left|\xi_n^Q(\theta)\right| = o_p(1) \tag{39}$$

Since $\sup_{\theta\in\Theta\setminus\Theta_0^I} |\vartheta_n(\theta)| = o_p(1)$, wp$\to$ 1, $\Sigma(\theta)^{-1} + \vartheta_n(\theta)$ is positive-definite uniformly over $\Theta \setminus \Theta_0^I$. Thus, wp$\to$ 1, $\inf_{\theta\in\Theta\setminus\Theta_0^I} d'\left(\Sigma(\theta)^{-1} + \vartheta_n(\theta)\right)d > 0 \quad \forall\, d \neq 0$, and therefore, for any $\theta \in \Theta \setminus \Theta_0^I$: $\mu(\theta) \neq 0$, $n\mu(\theta)'\left(\Sigma(\theta)^{-1} + \vartheta_n(\theta)\right)\mu(\theta) \to \infty$ w.p.1. Thus, for any $c > 0$, $P\left(\widehat{Q}(\theta) < c\right) \to 0\ \forall\ \theta \in \Theta \setminus \Theta_0^I$: $\mu(\theta) \neq 0$. Our previous results combined imply, $P\left(\widehat{Q}(\theta) < c\right) \to 0\ \forall\ \theta \notin \Theta^*$ for any $c > 0$. Next, take any sequence θ_n and let $\mu(\theta_n) \equiv \left(\left(\theta_{n,1} - \theta_1^*\right)' \quad R(\theta_n)\right)'$. If θ_n is such that $\left\|\mu(\theta_n)\right\| \geq \delta_n n^{-1/2} D$ for some fixed $D > 0$ and some sequence of positive constants $\delta_n \to \infty$, we have $P\left(\widehat{Q}(\theta_n) < c\right) \to 0$. Next, take any $\theta_n \in \Theta \setminus \Theta_0^I$ such that $n^{1/2} \cdot \mu(\theta_n)'\Sigma(\theta_n)^{-1}\mu(\theta_n) \to b > 0$. Then, $\lim_{n\to\infty} P\left(\widehat{Q}(\theta_n) < c\right) = F_{\chi_{r+1}^2}(c-b)$.

5.4 A confidence set (CS) for $\boldsymbol{\theta}$

The asymptotic properties of $\widehat{Q}(\theta)$ facilitate construction of a CS for θ. Let $\chi_\ell^2(\tau)$ denote the τ^{th} quantile of the $F_{\chi_\ell^2}$ distribution. We propose a CS with asymptotic target coverage probability $1-\alpha$ as

$$\text{CS}_{1-\alpha} = \left\{\theta \in \Theta\colon\ \widehat{Q}(\theta) < \chi_{r+1}^2(1-\alpha)\right\}, \tag{40}$$

The following theorem summarizes the asymptotic properties of our CS, including the power properties of the associated test for $\theta \notin \Theta^*$. Following the terminology of Lee, Song, and Whang (2018, Definition 3) we say that the associated test has nontrivial power for a sequence of local alternatives $\theta_{na} \notin \Theta^*$ if $\lim_{n\to\infty} P(\theta_{na} \in \text{CS}_{1-\alpha}) < 1-\alpha$.

Theorem 4 *Let Restrictions* ***I1*** *-* ***I7*** *hold. Then* $\text{CS}_{1-\alpha}$ *has the following asymptotic properties.*

(i) Uniform asymptotic coverage: $\lim_{n\to\infty} \inf_{\theta\in\Theta^*} P(\theta \in \text{CS}_{1-\alpha}) \geq 1-\alpha$.

(ii) Consistency of the associated test for $\theta \in \Theta^*$*:* $\forall\ \theta \notin \Theta^*\ \lim_{n\to\infty} P(\theta \in \text{CS}_{1-\alpha}) = 0$. *Moreover, let* $\theta_{na} \notin \Theta^*$ *be a sequence of local alternatives and let* $\mu(\theta_{na}) \equiv \left(\left(\theta_{na,1} - \theta_1^*\right)' \quad R(\theta_{na})\right)'$. *For any* θ_{na} *such that* $\left\|\mu(\theta_{na})\right\| \geq \delta_n n^{-1/2} D$ *for some fixed* $D > 0$ *and some sequence of positive constants* $\delta_n \to \infty$*, we have* $\lim_{n\to\infty} P(\theta_{na} \in \text{CS}_{1-\alpha}) = 0$.

(iii) Nontrivial local power of the associated test for $\theta \in \Theta^*$: For any $\theta_{na} \notin \Theta^*$ such that $\theta_{na} \in \Theta \setminus \Theta_0^I$ and $n^{1/2} \cdot \mu(\theta_{na})'\Sigma(\theta_{na})^{-1}\mu(\theta_{na}) \to b > 0$, we have $\lim_{n\to\infty} P(\theta_{na} \in \mathrm{CS}_{1-\alpha}) < 1-\alpha$. For any θ_{na} such that $\theta_{na} \in \Theta_0^I$ and $n(\theta_{na,1} - \theta_1^*)'H_0(\theta_{na,1} - \theta_1^*) \to a > 0$, we have $\lim_{n\to\infty} P(\theta_{na} \in \mathrm{CS}_{1-\alpha}) < 1-\alpha$ if $\chi^2_{r+1}(1-\alpha) - a < \chi^2_r(1-\alpha)$.

The results in the Theorem follow directly from the asymptotic analysis of $\widehat{Q}$ described previously. A key result is (31), which is proven in the Econometric Supplement. The Theorem establishes that the confidence set $\mathrm{CS}_{1-\alpha}$ provides correct ($\geq 1-\alpha$) asymptotic coverage for fixed P uniformly over $\theta \in \Theta^*$.[20] Moreover, the associated test for $\theta \in \Theta^*$ is consistent against all fixed alternatives as well as all local alternatives $\theta_{na} \notin \Theta^*$ for which $n^{1/2}\|\mu(\theta_{na})\| \to \infty$. Importantly, it also has nontrivial power for local alternatives in $\Theta \setminus \Theta_0^I$ as well as certain local alternatives in Θ_0^I. This last result is important since the asymptotic coverage of $CS_{1-\alpha}$ exceeds $1-\alpha$ over Θ_0^I. As is the case with other methods for inference based on conditional moment inequalities. The class of alternatives covered generally depends on both the set of values for which the conditional moment inequalities are violated and the properties of $\mu(\theta_{na})$ and $\Sigma(\theta_{na})$ under the sequence of local alternatives θ_{na}. The adaptive properties of our statistic $\widehat{R}(\theta)$ to the contact sets helps us avoid the conservative features of tests based on the least favorable null dgp, which assume that the CMIs are binding with probability one.

5.4.1 Allowing the measure of the contact sets to become arbitrarily close to zero over $\Theta \setminus \Theta_0^I$

Restriction I7 assumes that $\sigma_R^2(\theta)$ (and therefore, the measure of the contact sets) is bounded away from zero over $\Theta \setminus \Theta_0^I$.[21] We present now a modified version of our procedure that allows for this. We modify Restriction I7 as follows.

Restriction I7'

(i) Replace the sequence $\kappa_n \searrow 0$ with an arbitrarily small but strictly positive constant $\kappa > 0$.

(ii) $E[\|\psi_M(W)\|^3] < \infty$. For each $\theta \in \Theta \setminus \Theta_0^I$ let $\psi_{MR}(W;\theta) \equiv -\Sigma_{MR}(\theta)'H_0\psi_M(W) + \psi_R(W;\theta)$ and let $\sigma^2_{MR}(\theta) \equiv Var(\psi_{MR}(W;\theta))$. There exists a $\overline{D}_3 < \infty$ such that

$$E\left[\left|\frac{\psi_{MR}(W;\theta)}{\sigma_{MR}(\theta)}\right|^3\right] \leq \overline{D}_3 \quad \forall\, \theta \in \Theta \setminus \Theta_0^I.$$

[20] It is worth noting that our $\mathrm{CS}_{1-\alpha}$ can attain good pointwise asymptotic properties, i.e.

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

under weaker regularity conditions than those stated here. In particular, with Restrictions I1, I2, I3 maintained we could replace the remaining restrictions leading to Theorem 3 with any alternative set of assumptions that can produce the asymptotic linear representation result for $\widehat{R}(\theta)$ obtained in Theorem 3.

[21] Recall that, for parameters outside of Θ^I, our definition of contact sets includes the range of values of w for which the CMIs are violated.

Note that $\sigma^2_{MR}(\theta) = \sigma^2_R(\theta) - \Sigma_{MR}(\theta)'H_0\Sigma_{MR}(\theta) = \sigma^2_R(\theta)\times\left(1 - \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)' H_0 \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)\right)$. We assume that $\left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right)' H_0 \left(\frac{\Sigma_{MR}(\theta)}{\sigma_R(\theta)}\right) \leq \overline{d} < 1$ for all $\theta \in \Theta \setminus \Theta^I_0$. Finally, $\sigma_R(\theta)$ and $\Sigma_{MR}(\theta)$ are continuous everywhere over $\Theta \setminus \Theta^I_0$. ■

Restriction I7' allows $\sigma^2_R(\theta)$ (and therefore the measure of the contact sets) to become arbitrarily close to zero over $\Theta \setminus \Theta^I_0$ as long as the integrability condition (existence of third moment) described in part (ii) is satisfied. This will allow us to invoke the multivariate Berry-Esseen bounds we used under Restriction I7. We have also replaced the sequence $\kappa_n \searrow 0$ with a strictly positive (arbitrarily small) pre-specified constant $\kappa > 0$.

The asymptotic properties of $\widehat{Q}(\theta)$ over Θ^*_0 remain the same as uder Restriction I7. To study its properties over $\overline{\Theta}^*$ we begin by noting that the integrability condition in Restriction I7' directly satisfies the conditions needed for (34) and therefore that result is preserved. Next, define $\sigma^2_{MR,\kappa}(\theta) \equiv \max\{\sigma^2_R(\theta), \kappa\} - \Sigma_{MR}(\theta)'H_0\Sigma_{MR}(\theta) = \sigma^2_{MR}(\theta) + \left(\max\{\sigma^2_R(\theta), \kappa\} - \sigma^2_R(\theta)\right)$. For each $\theta \in \Theta \setminus \Theta^I_0$, let

$$\Sigma_\kappa(\theta) \equiv \begin{pmatrix} H_0^{-1} & \Sigma_{MR}(\theta) \\ \Sigma_{MR}(\theta)' & \max\{\sigma^2_R(\theta)\ ,\ \kappa\} \end{pmatrix} \quad \Longrightarrow \quad \Sigma_\kappa(\theta)^{-1} = \begin{pmatrix} H_0 + \frac{H_0\Sigma_{MR}(\theta)\Sigma_{MR}(\theta)'H_0}{\sigma^2_{MR,\kappa}(\theta)} & -\frac{H_0\Sigma_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)} \\ -\frac{\Sigma_{MR}(\theta)'H_0}{\sigma^2_{MR,\kappa}(\theta)} & \frac{1}{\sigma^2_{MR,\kappa}(\theta)} \end{pmatrix}$$

From Restriction I7', we have $\sigma^2_{MR,\kappa}(\theta) \geq \kappa\cdot(1-\overline{d})\ \forall\ \theta \in \Theta \setminus \Theta^I_0$,. The conditions in Theorems 2 and 3 and Restriction I7' yield the following modified version of the asymptotic result in (41),

$$\begin{aligned} &\widehat{Q}(\theta) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right)' \begin{pmatrix} I & 0 \\ 0 & \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) \end{pmatrix} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right) + \xi^{Q_\kappa}_n(\theta) \quad \forall\ \theta \in \Theta \setminus \Theta^I_0, \\ &\text{where} \quad \sup_{\theta\in\Theta\setminus\Theta^I_0} \left|\xi^{Q_\kappa}_n(\theta)\right| = o_p(1). \end{aligned} \tag{41}$$

The integrability condition in Restriction I7 and the multivariate Berry-Esseen bounds in Raič (2019, Theorem 1.1) imply that, $\exists\ \overline{M} < \infty$ such that, for all $c > 0$, and for all $\theta \in \Theta \setminus \Theta^I_0$,

$$\left| P\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right)' \begin{pmatrix} I & 0 \\ 0 & \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) \end{pmatrix} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right) \leq c\right) - P\left(Z' \begin{pmatrix} I & 0 \\ 0 & \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) \end{pmatrix} Z \leq c\right) \right| \leq \frac{\overline{M}}{\sqrt{n}} \tag{42}$$

where $Z \sim \Phi_{r+1}$. Note that the limiting distribution is $P\left(\mathscr{X}_1 + \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right)\cdot \mathscr{X}_2 \leq c\right)$, where $\mathscr{X}_1 \sim \chi^2_r$, $\mathscr{X}_2 \sim \chi^2_1$ and $\mathscr{X}_1 \perp \mathscr{X}_2$. The general distribution of linear combinations of independent, central chi-square random variables has been obtained, e.g, in Moschopoulos and Canada (1984), as an infinite gamma series whose properties depend on the weights of the linear combination. However, note that $\left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) = 1$ for each $\theta \in \Theta \setminus \Theta^I_0$: $\sigma^2_R(\theta) \geq \kappa$ and $0 < \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) < 1$ for each

$\theta \in \Theta \setminus \Theta_0^I$: $\sigma_R^2(\theta) < \kappa$. Thus, in our case, for all $c > 0$:

$$\sup_{\substack{\theta \in \overline{\Theta}^*: \\ \sigma_R^2(\theta) \geq \kappa}} \left| P\left(\widehat{Q}(\theta) \leq c\right) - F_{\chi^2_{r+1}}(c) \right| \to 0, \quad \text{and} \quad \inf_{\substack{\theta \in \overline{\Theta}^*: \\ \sigma_R^2(\theta) < \kappa}} P\left(\widehat{Q}(\theta) \leq c\right) \geq F_{\chi^2_{r+1}}(c)$$

And therefore, $\inf_{\theta \in \overline{\Theta}^*} P\left(\widehat{Q}(\theta) \leq c\right) \geq F_{\chi^2_{r+1}}(c)$.

In order to analyze the power properties of our procedure under Restriction I7', consider the asymptotic properties of $\widehat{Q}(\theta)$ when $\theta \notin \Theta^*$. First, for any $\theta \in \Theta_0^I$ we obtain the same asymptotic features as those we described under Restriction I7. Next, for $\Theta \setminus \Theta_0^I$ we have

$$\widehat{Q}(\theta) = n\mu(\theta)'\left(\Sigma_\kappa(\theta)^{-1} + \vartheta_n^\kappa(\theta)\right)\mu(\theta) + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right)' \begin{pmatrix} I & 0 \\ 0 & \left(\frac{\sigma^2_{MR}(\theta)}{\sigma^2_{MR,\kappa}(\theta)}\right) \end{pmatrix} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right) + \xi_n^{Q_\kappa}(\theta),$$

$$\text{where} \quad \mu(\theta) \equiv \left((\theta_1 - \theta_1^*)' \quad R(\theta)\right)', \quad \sup_{\theta \in \Theta \setminus \Theta_0^I} |\vartheta_n^\kappa(\theta)| = o_p(1), \quad \text{and} \quad \sup_{\theta \in \Theta \setminus \Theta_0^I} \left|\xi_n^{Q_\kappa}(\theta)\right| = o_p(1) \tag{43}$$

Thus, for all $\theta \in \Theta \setminus \Theta_0^I$: $\mu(\theta) \neq 0$, we have $n\mu(\theta)'\left(\Sigma_\kappa(\theta)^{-1} + \vartheta_n^\kappa(\theta)\right)\mu(\theta) \to \infty$ w.p.1., and therefore $P\left(\widehat{Q}(\theta) < c\right) \to 0$. Our previous results combined yield $P\left(\widehat{Q}(\theta) < c\right) \to 0$ for all $\theta \notin \Theta^*$ and any $c > 0$. Now take any θ_n and let $\mu(\theta_n) \equiv \left(\left(\theta_{n,1} - \theta_1^*\right)' \quad R(\theta_n)\right)'$. If θ_n is such that $\left\|\mu(\theta_n)\right\| \geq \delta_n n^{-1/2} D$ for some fixed $D > 0$ and some sequence of positive constants $\delta_n \to \infty$, then $P\left(\widehat{Q}(\theta_n) < c\right) \to 0$ for all $c > 0$. Now consider $\theta_n \in \Theta \setminus \Theta_0^I$ such that $n^{1/2} \cdot \mu(\theta_{na})'\Sigma(\theta_{na})^{-1}\mu(\theta_{na}) \to b > 0$ and $\sigma_R^2(\theta_{na}) \geq \kappa \ \ \forall\ n$. Then, $\lim_{n\to\infty} P\left(\widehat{Q}(\theta_n) < c\right) = F_{\chi^2_{r+1}}(c - b)$. Take $\theta_n \in \Theta \setminus \Theta_0^I$ with $n^{1/2} \cdot \mu(\theta_{na})'\Sigma(\theta_{na})^{-1}\mu(\theta_{na}) \to B > 0$ and $\sigma_R^2(\theta_{na}) < \kappa$ for a sufficiently large n. Suppose $\left(\frac{\sigma^2_{MR}(\theta_n)}{\sigma^2_{MR,\kappa}(\theta_n)}\right) \to a \in (0,1)$. We have $\lim_{n\to\infty} P\left(\widehat{Q}(\theta_n) < c\right) = P\left(\mathscr{X}_1 + a \cdot \mathscr{X}_2 < c - b\right)$, where $\mathscr{X}_1 \sim \chi^2_r$, $\mathscr{X}_2 \sim \chi^2_1$ and $\mathscr{X}_1 \perp \mathscr{X}_2$. The following theorem summarizes the asymptotic properties of our CS under Restriction I7'.

Theorem 5 *Let Restrictions **I1** - **I6** and **I7'** hold. Then* $\text{CS}_{1-\alpha}$ *has the following asymptotic properties.*

(i) Uniform asymptotic coverage: $\lim_{n\to\infty} \inf_{\theta \in \Theta^*} P(\theta \in \text{CS}_{1-\alpha}) \geq 1 - \alpha$.

(ii) Consistency of the associated test for $\theta \in \Theta^*$*:* $\lim_{n\to\infty} P(\theta \in \text{CS}_{1-\alpha}) = 0 \ \forall\ \theta \notin \Theta^*$. *Next consider sequences of local alternatives* θ_{na} *and let* $\mu(\theta_{na}) \equiv \left(\left(\theta_{na,1} - \theta_1^*\right) \quad R(\theta_{na})\right)$. *For any* θ_{na} *such that* $\left\|\mu(\theta_{na})\right\| \geq \delta_n n^{-1/2} c$ *for some fixed* $c > 0$ *and some sequence of positive constants* $\delta_n \to \infty$, *we have* $\lim_{n\to\infty} P(\theta_{na} \in \text{CS}_{1-\alpha}) = 0$

(iii) Nontrivial local power of the associated test for $\theta \in \Theta^*$*: Consider any sequence of local alternatives* θ_{na} *such that* $\theta_{na} \in \Theta \setminus \Theta_0^I$ *and* $n^{1/2} \cdot \mu(\theta_{na})'\Sigma(\theta_{na})^{-1}\mu(\theta_{na}) \to b > 0$. *If* $\sigma_R^2(\theta_{na}) \geq \kappa \ \ \forall\ n$, *then* $\lim_{n\to\infty} P(\theta_{na} \in \text{CS}_{1-\alpha}) < 1 - \alpha$. *Suppose instead that* $\sigma_R^2(\theta_{na}) < \kappa$ *for all n sufficiently large, and* $\left(\frac{\sigma^2_{MR}(\theta_{na})}{\sigma^2_{MR,\kappa}(\theta_{na})}\right) \to a \in (0,1)$. *Then* $\lim_{n\to\infty} P(\theta_{na} \in \text{CS}_{1-\alpha}) = P\left(\mathscr{X}_1 + a \cdot \mathscr{X}_2 < \chi^2_{r+1}(1-\alpha) - b\right)$,

where $\mathscr{X}_1 \sim \chi^2_r$, $\mathscr{X}_2 \sim \chi^2_1$ *and* $\mathscr{X}_1 \perp \mathscr{X}_2$. *Since* $P(\mathscr{X}_1 + a \cdot \mathscr{X}_2 < c) \leq P\left(\mathscr{X}_1 < \frac{c}{2}\right) + P\left(a \cdot \mathscr{X}_2 < \frac{c}{2}\right)$, *a sufficient condition to have* $\lim_{n\to\infty} P(\theta_{na} \in \mathrm{CS}_{1-\alpha}) < 1-\alpha$ *would be if* $\frac{\chi^2_{r+1}(1-\alpha)-b}{2} < \chi^2_r\left(\frac{1-\alpha}{2}\right)$ *and* $\frac{\chi^2_{r+1}(1-\alpha)-b}{2\cdot a} < \chi^2_1\left(\frac{1-\alpha}{2}\right)$. *Finally, for any* $\theta_{na} \in \Theta^I_0$ *such that* $n(\theta_{na,1} - \theta^*_1)' H_0(\theta_{na,1} - \theta^*_1) \to a > 0$, *we will have* $\lim_{n\to\infty} P(\theta_{na} \in \mathrm{CS}_{1-\alpha}) < 1-\alpha$ *if* $\chi_{r+1}(1-\alpha) - a < \chi_r(1-\alpha)$

One again, the results in the theorem follow directly from the asymptotic analysis of $\widehat{Q}$ described previously. Our CS retains correct asymptotic coverage probability even though it is more conservative for certain parameter values compared to the results under Restriction I7. In that case our CS had coverage probability greater than $1-\alpha$ only for parameters in Θ^I_0. This will be true with Restriction I7' also for parameters in $\Theta \setminus \Theta^I_0$ for which $\sigma^2_R(\theta) < \kappa$. Importantly, the underlying test retains the consistency properties of Theorem 4 and it also has nontrivial asymptotic power against local alternatives $\theta_{na} \notin \Theta^*$, including some belonging in Θ^I_0 as well as some in $\Theta \setminus \Theta^I_0$ for which $\sigma^2_R(\theta_{na}) < \kappa$. This is an important result because the asymptotic coverage in those regions is strictly greater than $1-\alpha$. Note crucially that our underlying test has nontrivial asymptotic power against local alternatives where the CMIs are violated with probability approaching zero.[22]

5.4.2 Achieving uniformity over a class of distributions

Our results were illustrated for a fixed distribution P. Going over each of our restrictions as well as the steps of our proofs in the Econometric Supplement shows what would be required to obtain uniform results over a class of distributions $\mathcal{P}$. We sketch the arguments here, leaving the details for future work for the sake of brevity. The smoothness conditions in Restriction I1 would have to hold uniformly over $\mathcal{P}$ (i.e, we would require a common upper bound over $\mathcal{P}$ for the derivatives of the different functionals described there). Next, we would require a common upper bound for $\mathcal{P}$ for the VC dimension of the classes of sets described in Restriction I3. The bound in Restriction I4 would also have to hold for every $P \in \mathcal{P}$. From here, the maximal inequality results in Sherman (1994, Main Corollary) would hold uniformly over $\mathcal{P}$, and so would Lemma 2. If the smoothness condition in Restriction I5 were also assumed to hold for each element in $\mathcal{P}$, the linear representation in Theorem 3 would also hold uniformly over this class of distributions. From here, if we assume that the integrability conditions described Restriction I6 hold uniformly over $\mathcal{P}$ (i.e, for an envelope $D(W)$ whose second moment has a universal upper bound over $\mathcal{P}$), the type of result shown in (35) would hold uniformly over $\mathcal{P}$. From here, if we assume that the third moments and the bounds $\underline{\sigma}$, $\overline{d}$ and $\overline{D}_3$ described in Restrictions I7 and I7' are valid for all elements in $\mathcal{P}$, we would obtain universal Berry-Esseen bounds for $\mathcal{P}$, enabling extension of results in Theorems 4 and 5 to hold uniformly over this class of distributions.

[22] Recall that our definition of contact sets for parameters outside of Θ^I is $\{w \in \mathcal{W}^* \colon\ T_k(w;\theta) \geq 0 \text{ for some } k = 1,\ldots,K\}$, which includes values of w where the CMIs are violated.

5.5 Comparison with other CMI inferential methods

Our test statistic transforms conditional moment inequalities to unconditional ones using an "instrument function" and then integrates over violations of these unconditional moment functions. The use of instrument functions to move from conditional to unconditional moment inequalities is conceptually similar to Andrews and Shi (2013), and integration over these violations is conceptually similar to their Cramer von-Mises (CvM) statistic as well as the L_p-type functionals used by Lee, Song, and Whang (2018). Instead of pre-specifying a space of instrument functions and then aggregating over it (e.g, as Andrews and Shi (2013) and Armstrong (2018)), the sole instrument function we focus on is simply $1[T_k(w;\theta) \geq 0]$, which captures directly whether the CMIs are violated or not and would therefore be a sufficient instrument if the functional $T_k(w;\theta)$ were known. However, $T_k(w;\theta)$ is unknown, and our contribution focuses on describing conditions under which it can be replaced with a kernel-based estimator so that a result like that in Lemma 2 can be obtained. Our assumptions rely on smoothness conditions that are entirely analogous to those used in many existing semiparametric econometric models (e.g, Powell, Stock, and Stoker (1989), Ichimura (1993), Ahn and Manski (1993), Ahn and Powell (1993), and many more).[23] While restrictive, there is nothing in the nature of our problem that makes these smoothness assumptions less credible than in those models. In return, we obtain a statistic with an asymptotically linear representation and pivotal properties, ideally suited to be combined for inference with a root-n consistent estimator for the point-identified parameters, as in our two-player application.

Our statistic also has contact set properties that distinguish it from existing statistics. First, we showed that it adapts asymptotically to the contact sets in place of using generalized moment selection or explicit contact set estimation. Second, its properties allowed us to construct a CS that allows for contact sets to have measure arbitrarily close to zero (Restriction I7'). Like other existing CvM statistics for CMIs, ours shares qualitatively similar local asymptotic power properties relative to sup-norm type tests such as those of Chernozhukov, Lee, and Rosen (2013) or the Kolmogorov-Smirnov test of Armstrong (2015). Thus our test statistic should be expected to perform well from a power standpoint against relatively flat local alternatives, but less well against relatively sharp-peaked alternatives where the violation of conditional moment inequalities occurs on a set of shrinking measure under the sequence of local alternatives. We leave the details of such a comparison for future work.

Finally, having an asymptotic linear representation, our statistic is uniquely suited to be used in problems where a subset of parameters are point-identified and can be consistently estimated with a regular estimator, for instance by maximum likelihood. Our results show how they can be combined to construct a CS for the full parameter vector. Overall, our statistic contributes to the CMI literature by combining all of the following features simultaneously: (i) having asymp-

[23] The regularity condition in Restriction I4 is the only one that is truly unique to our CMI analysis relative to existing semiparametric models.

totically pivotal properties, (ii) automatically adapting to contact sets, thus avoiding the need to pretest for the slackness of the CMIs while also avoiding the conservative features of tests based on the least favorable null DGP, (iii) being readily applied to models where a subset of parameters can be point-identified and estimated through a root-n consistent estimator, (iv) allowing contact sets to have arbitrarily small measure and having nontrivial asymptotic power against local alternatives where the CMIs are violated with probability approaching zero.

6 Monte Carlo experiments

The Empirical Supplement to this paper contains the details and results for several Monte Carlo experiments. Generating data from ordered response games that satisfy our assumptions, our goal is to study the empirical properties of our approach as well as the results from *incorrectly assuming that the game is binary*. We find that our CS has coverage properties in line with our asymptotic predictions for all the experimental designs, which vary according to payoff features such as the degree of concavity, as well as the equilibrium selection rule used by players. Importantly, we find that misspecifying a true ordered-response game as a binary game systematically underestimates payoff functions. A binary game ignores the intensive-margin nature of strategic interaction. With strategic substitutes, a true binary-choice game with the same non-strategic payoff function component as an ordinal game would produce probabilities of entry that are much larger than the ones produced by its ordinal counterpart.[24] Consequently, our experiments find that incorrectly specifying a true ordinal game as a binary one produces estimates for the non-strategic component of payoffs that are systematically biased downwards. Ignoring the intensive margin nature of the ordinal game also underestimates the strategic interaction effect.

7 Application to a Multiple Entry Game between 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 refer to Lowe's as player 1 and Home Depot as player 2. 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.[25] Our sample consists of a cross section of $n = 954$ markets in April

[24] In a binary game, the incentive to enter depends only on whether the opponent entered. In an ordered game like ours, the incentive to enter is different if the opponent opens one store versus more (e.g. five) stores.

[25] 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

2012. Table 1 summarizes features of the observed distribution of outcomes, distinguishing between small and large markets depending on whether market size (population) is above or below the median in our data.

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.

Table 1: Summary of outcomes observed in the data.

All markets			
	Y_1	Y_2	Y_1 vs. Y_2
Average	1.68	1.97	
Median	1	1	$(\% Y_1 > Y_2)$: 33%
75^{th} percentile	2	1	$(\% Y_1 < Y_2)$: 25%
90^{th} percentile	4	5	$(\% Y_1 + Y_2 > 0)$: 74%
95^{th} percentile	7	11	$(\% Y_1 + Y_2 > 0,\ Y_1 = Y_2)$: 16%
99^{th} percentile	17	25	
Total	1,603	1,880	player 1: Lowe's, player 2: Home Depot.

Markets with population below the median			
	Y_1	Y_2	Y_1 vs. Y_2
Average	0.33	0.23	
Median	0	0	$(\% Y_1 > Y_2)$: 28%
75^{th} percentile	1	0	$(\% Y_1 < Y_2)$: 18%
90^{th} percentile	1	1	$(\% Y_1 + Y_2 > 0)$: 51%
95^{th} percentile	1	1	$(\% Y_1 + Y_2 > 0,\ Y_1 = Y_2)$: 5%
99^{th} percentile	1	1	
Total	156	109	player 1: Lowe's, player 2: Home Depot.

Markets with population above the median			
	Y_1	Y_2	Y_1 vs. Y_2
Average	3.03	3.71	
Median	2	1	$(\% Y_1 > Y_2)$: 38%
75^{th} percentile	3	3	$(\% Y_1 < Y_2)$: 31%
90^{th} percentile	7	10	$(\% Y_1 + Y_2 > 0)$: 97%
95^{th} percentile	12	17	$(\% Y_1 + Y_2 > 0,\ Y_1 = Y_2)$: 28%
99^{th} percentile	21	38	
Total	1,447	1,771	player 1: Lowe's, player 2: Home Depot.

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. Table 1 suggests that market size is an important predictor of entry (97% of markets with population above the median have at least one store), and that Lowe's has a larger proportion of its stores (9.7%) in small markets than Home Depot (5.8%). Lowe's also tends to have more stores than Home Depot in small markets and *vice versa*. These features are entirely compatible with the assumption that the direct effect of market size on players' payoffs (i.e, the coefficient of market size) is the same for both, as these features of the data can be owed to the structural equilibrium features of the underlying game, which is why we allow for players' strategic interaction coefficients to differ from each other. Furthermore, even if we assumed that payoff functions are exactly the same for both players, these patterns in the data could be the result of the underlying, unknown equilibrium selection mechanism.

Naturally, entry decisions take place over time. Our justification for approximating this problem as a static game is the assumption that the outcome observed is the realization of a long-run equilibrium.[26] 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.

7.1 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.[27] 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 stores than Lowe's relative to markets where the opposite held. Overall, Home Depot opened more stores than Lowe's in markets that were larger and that had higher earnings per capita. Our methodology allows us to investigate whether this phenomenon is caused by inherent differences in firms' payoff functions or by the strategic-interaction component (including features of the underlying equilibrium selection mechanism).

[26] 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 arguably well-suited to this application.

[27] Payroll per capita is included both as a measure of income and as a labor market indicator 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 results.

7.2 Test sets for the construction of confidence sets

The class of test sets $\mathcal{U}_k(y_i, x_i; \theta)$ we used is as follows. As before let $\mathcal{R}_\theta$ be as defined in (13). Let

$$\mathsf{S}^{I}_{\theta}(x_i) = \Big\{\mathcal{S} \subseteq \mathbb{R}^2 \colon \mathcal{S} = \mathcal{R}_\theta(y, x_i) \text{ such that } y_1 \leq 2,\, y_2 \leq 2\Big\},$$
$$\mathsf{S}^{II}_{\theta}(x_i) = \Big\{\mathcal{S} \subseteq \mathbb{R}^2 \colon \mathcal{S} = \mathcal{R}_\theta(y, x_i) \cup R_\theta(y', x_i) \text{ for some } y \neq y' \text{ such that } y_1 \leq 2 \text{ and } y_2 \leq 2\Big\}.$$

For test sets $\mathcal{U}_k(x_i, y_i; \theta)$ we use each element of the collections $\mathsf{S}^{I}_{\theta}(x_i)$ and $\mathsf{S}^{II}_{\theta}(x_i)$ as well as the set $\mathcal{R}_\theta(y_i, x_i)$. This yields $K = 82$ tests sets.

7.3 Kernels and tuning parameters

Our covariate vector X is comprised of five continuous random variables. We employed a multiplicative kernel $\mathbf{K}(\psi_1, \ldots, \psi_5) = \mathbf{k}(\psi_1)\mathbf{k}(\psi_2)\cdots\mathbf{k}(\psi_5)$, with $\mathbf{k}(u) = \sum_{\ell=1}^{10} c_\ell \cdot (1 - u^2)^{2\ell} \cdot 1\{|u| \leq 30\}$, where $c_1, \ldots, c_{10}$ 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 $z \equiv \dim(X^c) = 5$ and $\epsilon \equiv \frac{9}{10} \cdot \frac{1}{4z(2z+1)}$, $\alpha_h \equiv \frac{1}{4z} - \epsilon$. For each element of X, the bandwidth used was of the form $h_n = c \cdot \widehat{\sigma}(X) \cdot n^{-\alpha_h}$. The order of the kernel and the bandwidth convergence rate were chosen to satisfy Restriction **I2**. The constant c was set at 0.25.[28] 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 region $\mathcal{X}^*$ was set to include our entire sample, so there was no trimming used in our results.

7.4 Results for payoff parameters

Table 2 presents the ML estimates for our point-identified parameters, along with confidence intervals for each payoff parameter. The second column includes ML 95% CIs for our ML estimates and the third column includes projections given by the smallest and largest values of each parameter in our 95% 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

[28] $c = 0.25$ 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.

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 event of no-entry, and not the ordinal value of the outcome.

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 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 benefits for opening new stores in a market.

Table 2: Estimates and Confidence Intervals for each Parameter

	MLE Estimate	MLE 95% CI	Moment-inequalities 95% CI†
Population (100,000)	2.219	$[0.869, 3.568]$	$[1.757, 3.792]$
Payroll per capita ($5 USD)	0.244	$[-0.023, 0.510]$	$[-0.064, 0.667]$
Land Area (1,000 sq miles)	0.180	$[0.027, 0.333]$	$[0.051, 0.409]$
Distance (100 miles)	-0.544	$[-0.929, -0.159]$	$[-0.988, -0.410]$
ρ ($Corr(U_1, U_2)$)	-0.050	$[-0.304, 0.204]$	$[-0.265, 0.302]$
$\delta - \eta$ (Intercept minus concavity coefficient)	-1.309	$[-2.084, -0.534]$	$[-1.961, -0.656]$
δ (Intercept)	N/A	N/A	$[-0.351, 5.463]$
η (Concavity coefficient)	N/A	N/A	$[1.076, 6.533]$
Δ_1 (Effect of Home Depot on Lowe's)	N/A	N/A	$[0, 2.741]$
Δ_2 (Effect of Lowe's on Home Depot)	N/A	N/A	$[0.910, 4.078]$

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

Figure 2 depicts the joint CS for the strategic interaction coefficients, Δ_1 and Δ_2. The parameter space for these coefficients covered the two-dimensional rectangle $[0, 16] \times [0, 16]$. Our results suggest that the strategic effect of Lowes on Home Depot (measured by Δ_2) is stronger than the effect of Home Depot on Lowes (measured by Δ_1). As we can see in the figure, our CS lies almost entirely above the 45-degree line. Our results conclusively excluded the point $\Delta_1 = \Delta_2 = 0$, so we can reject the assertion that no strategic effect is present. The Empirical Supplement includes graphical inspections of joint CS for pairs of parameters and they did not reveal any holes; however we are not sure about the robustness of this feature for our entire CS given its dimension.

7.5 Inference on Other Quantities of Economic Interest: Likelihood of Equilibria and Features of the Underlying Selection Mechanism

An exercise that is sometimes overlooked in empirical work on partially identified games is to go beyond the basic payoff parameters and use the results to explore other features of the game. These include, for example, the likelihood that a given outcome y is an equilibrium conditional on observables, or the propensity of the underlying equilibrium selection mechanism to choose y whenever it is an equilibrium. In many instances these features can be more interesting than the payoff parameters themselves, and the results can be quite informative.

7.5.1 Overview

The functionals considered are functions of θ, say $g(\theta)$, that map from Θ to some set $\mathcal{G} \subseteq \mathbb{R}$. In the cases we study, g is either a known function or it is of the form $g(\theta) \equiv \Gamma(E[\lambda(\theta;X,Y)])$ for some functions λ and Γ. In this case, $\widehat{g}(\theta)$ can be estimated by replacing $E[\lambda(\theta;X,Y)]$ above with its sample analog. By standard arguments a $1-\alpha$ confidence interval for $g(\theta)$ is then given by $\widehat{g}(\theta) \pm n^{-1/2}\Phi^{-1}(1-\alpha/2)\widehat{\sigma}(\theta)$, where $\widehat{\sigma}(\theta)$ consistently estimates the standard deviation of $n^{1/2}(\widehat{g}(\theta)-g(\theta))$. If θ_0 were known it could be plugged into the expression above to obtain the desired confidence interval for $g(\theta_0)$. Since θ_0 is unknown we use a Bonferroni bound together with $CS_{1-\alpha}$ to construct an asympotically valid $1-2\alpha$ confidence interval for $g(\theta_0)$,

$$CI(g(\theta_0)) \equiv \left[\min_{\theta\in CS_{1-\alpha}} \widehat{g}(\theta) - n^{-1/2}\Phi^{-1}(1-\alpha/2)\widehat{\sigma}(\theta), \quad \max_{\theta\in CS_{1-\alpha}} \widehat{g}(\theta) + n^{-1/2}\Phi^{-1}(1-\alpha/2)\widehat{\sigma}(\theta)\right]. \quad (44)$$

In cases where g is known, we construct $1-\alpha$ confidence intervals by simply taking projections of $CS_{1-\alpha}$ as $CI(g(\theta_0)) \equiv \{g(\theta) : \theta \in CS_{1-\alpha}\}$.

7.5.2 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), 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

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

when $\theta = \theta_0$, where $P_{\mathcal{E}}(y',y|x)$ denotes the conditional probability that both y' and y are equilibria given $X = x$. For the sake of illustration, Table 3 presents a 95% CI for $P_{\mathcal{E}}(y'|y,x)$ 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. Note that the lower bound in our CI was zero in each case since it is impossible to reject the likelihood that the outcome observed was the unique PSNE.

The aggregate probability that the outcome y is an equilibrium, denoted $P_{\mathcal{E}}(y)$, is given by $P_{\mathcal{E}}(y) = E[P_{\mathcal{E}}(y|Y,X)]$. For $\theta = \theta_0$, a consistent estimator for $P_{\mathcal{E}}(y)$ is given by

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

Table 4 presents the 0.90 ($\alpha = 0.05$) 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 observed outcome $(2,2)$ in CBSA 11100 (Amarillo, TX).

y	95% CI for $P_{\mathcal{E}}(y|Y_i,X_i)$	y	95% CI for $P_{\mathcal{E}}(y|Y_i,X_i)$
(0, 4)	[0, 0.9981]	(3, 1)	[0, 0.9510]
(6, 0)	[0, 0.9976]	(4, 0)	[0, 0.9388]
(4, 1)	[0, 0.9971]	(3, 0)	[0, 0.2666]
(0, 3)	[0, 0.9856]	(5, 1)	[0, 0.1001]
(5, 0)	[0, 0.9730]	(0, 5)	[0, 0.0524]
(1, 3)	[0, 0.9622]	(7, 0)	[0, 0.0114]

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

y	90% CI for $P_{\mathcal{E}}(y)$ ($\alpha = 0.05$)	Observed frequency for y	y	90% CI for $P_{\mathcal{E}}(y)$ ($\alpha = 0.05$)	Observed frequency for y
(0, 0)	[0.2415, 0.2847]	0.2631	(2, 0)	[0.0083, 0.1200]	0.0146
(1, 0)	[0.1973, 0.3001]	0.2023	(3, 1)	[0.0078, 0.0399]	0.0136
(1, 1)	[0.1224, 0.1566]	0.1257	(2, 2)	[0.0065, 0.0310]	0.0136
(0, 1)	[0.1081, 0.2552]	0.1205	(3, 2)	[0.0040, 0.0276]	0.0094
(2, 1)	[0.0398, 0.0720]	0.0461	(2, 3)	[0.0038, 0.0224]	0.0094
(1, 2)	[0.0120, 0.0691]	0.0199	(3, 3)	[0.0045, 0.0177]	0.0083

7.5.3 Propensity of Equilibrium Selection

Our model makes no assumptions 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, $P_{\mathcal{M}}(y) \equiv \frac{P(Y=y)}{P_{\mathcal{E}}(y)}$. Recall that $(0,0)$ cannot coexist with any other PSNE 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 is 1, since it is always possible that the outcome in question is the unique equilibrium. Thus, only the lower bounds of our CIs on the selection probabilities are informative.

Table 5: Propensity $P_{\mathcal{M}}(y)$ to select outcome y when it is a PSNE.

y	90% CI for $P_{\mathcal{M}}(y)$ ($\alpha = 0.05$)	Observed frequency for y	y	90% CI for $P_{\mathcal{M}}(y)$ ($\alpha = 0.05$)	Observed frequency for y
(1, 1)	[0.8884, 1]	0.1257	(2, 2)	[0.4503, 1]	0.0136
(1, 0)	[0.6896, 1]	0.2023	(3, 1)	[0.2850, 1]	0.0136
(2, 1)	[0.6350, 1]	0.0461	(3, 2)	[0.2381, 1]	0.0094
(0, 1)	[0.3932, 1]	0.1205	(2, 3)	[0.2322, 1]	0.0094
(3, 3)	[0.5772, 1]	0.0083	(1, 2)	[0.1920, 1]	0.0199

We can also make direct comparisons between the propensity of equilibrium selection for specific outcomes. The Empirical Supplement includes a number of such comparisons where we learn, among other things, that comparing outcomes where only one store is opened, there is a higher selection propensity for Lowe's to have the only store than for Home Depot, and that the selection propensity for equilibria in which both firms operate one store is higher than those in which only one firm does.

7.5.4 Counterfactual equilibrium selection mechanisms

As we have shown, 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. With this information at hand we can do counterfactual analysis based on pre-specified equilibrium selection mechanisms. The goal would be to find out which of these hypothetical selection mechanisms would produce outcomes that more closely resemble our data.

In the Empirical Supplement we explore this by looking at four hypothetical equilibrium selection mechanisms. The first two favor each firm individually, the third one favors entry by both firms, maximizing the total number of stores, while the fourth one favors symmetric outcomes. Our results there find evidence in favor of an underlying selection mechanism that favors choosing symmetric outcomes whenever they exist as equilibria.

7.6 Estimation as a binary entry game

In the Empirical Supplement we also re-estimated the game as a binary entry/no-entry game and our results showed the same patterns we uncovered in our Monte Carlo experiments when a true ordinal game is misspecified as a binary choice game. First, our payoff-function estimates shifted downwards relative to the ordered-response results. Second, our nonparametric estimates for $Pr\left(d_i = 1|Y_j, X\right)$, where $d_i = 1\left[Y_i \geq 1\right]$, proved to be monotonically decreasing in Y_j for both players, which is the pattern we observed in our experiments when the true underlying game is ordinal as opposed to binary. While not definitive and not grounded on a formal specification test, these findings are consistent with the features of a true ordered-response game according to the Monte Carlo results also included in the Empirical Supplement.

7.7 Other counterfactual exercises

The Empirical Supplement also includes other counterfactual experiments allowed by our methodology. For example, we compare the outcomes observed with a counterfactual scenario where firms cooperate and maximize their combined payoffs. We find, for example, that the expected number of total stores under cooperation would increase from 1 to more than 2 in at least 85 markets. We also conduct a counterfactual experiment where one of these competitors exits the industry and we find that the number of markets without a store would increase from 251 in the data to as many as 465 if Home Depot became a monopolist.

8 Conclusion

Econometric inference for discrete games with complete information has largely focused on games with very limited action spaces, with the binary case being the most prominent. One reason behind this is that a rich action space can significantly complicate equilibrium analysis if this requires payoff comparisons across the entire action space. This paper contributes to the literature by studying games with rich, possibly unbounded action spaces. Assuming that actions are ordinal in nature, we describe shape restrictions on payoff functions that turn the game effectively into a simultaneous ordered response model. This, in turn, greatly simplifies the econometric analysis as necessary conditions for Nash equilibria involve only adjacent actions and therefore do not require finding all equilibria or analyzing (or even knowing) the entire action space.

Focusing on the case of a two-player game of strategic substitutes, we showed that assuming pure strategy Nash equilibrium (PSNE) behavior can point-identify a subset of parameters, which stands in contrast with two-player binary choice games where PSNE behavior can point-identify all the parameters in the model. We developed a novel inference approach for all model parameters that combines the point identified parameters (estimated by MLE) with the conditional moment inequalities implied by PSNE behavior. Inference is based on a statistic with pivotal asymptotic properties, which facilitates its use in games with a rich collection of parameters and conditioning variables.

We applied our methodology to a multiple entry game between Lowe's and Home Depot. Our results revealed interesting features about the strategic interaction between these firms. Beyond results pertaining to payoff parameters, we showed that our framework also allows inference on other economic quantitites of interest, such as probability distributions of equilibrium outcomes, the propensity of the underlying (and unspecified) selection mechanism to choose certain outcomes, and other counterfactuals. Monte Carlo experiments and our empirical results also suggest that misspecifying a true ordinal game as a binary game can lead to systematic bias of payoff parameters as well as strategic interaction coefficients. This bias arises from ignoring the intensive margin of competitive effects. This suggests the importance of developing nonparametric specification tests that can help us know the true nature of the action space in these models.

References

Ahn, H. and C. Manski (1993). Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of Econometrics 56*, 291–321.

Ahn, H. and J. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*, 3–29.

Andrews, D. and P. Jia-Barwick (2010). Inference for parameters defined by moment inequalities. *Econometrica 80*, 2805–2826.

Andrews, D. W. K. and X. Shi (2013). Inference for parameters defined by conditional moment inequalities. *Econometrica 81*(2), 609–666.

Andrews, D. W. K. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Aradillas-López, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics 157*(2), 409–431.

Aradillas-López, A. (2011). Nonparametric probability bounds for nash equilibrium actions in a simultaneous discrete game. *Quantitative Economics 2*, 135–171.

Aradillas-López, A. (2020). The econometrics of static games. *Annual Review of Economics 12*, 135–165.

Aradillas-López, A. and A. Gandhi (2016). Estimation of games with ordered actions: An application to chain-store entry. *Quantitative Economics 7*(1), 727–780.

Aradillas-López, A., A. Gandhi, and D. Quint (2013). Identification and inference in ascending auctions with correlated private values. *Econometrica 81*, 489–534.

Aradillas-Lopez, A. and A. M. Rosen (2014). Inference in ordered response games with complete information. CEMMAP working paper CWP36/14.

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. (2014). Weighted ks statistics for inference on conditional moment inequalities. *Journal of Econometrics 181*(2), 92–116.

Armstrong, T. B. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics 186*(1), 51–65.

Armstrong, T. B. (2018). On the choice of test statistic for conditional moment inequalities. *Journal of Econometrics 203*(1), 241–255.

Armstrong, T. B. and H. P. Chan (2016). Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics 194*(1), 24–43.

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 M. Armstrong and P. Rob (Eds.), *The Handbook of Industrial Organization*, Volume 3, pp. 1845–1886. North-Holland.

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.

Bresnahan, T. F. and P. J. Reiss (1991a). Empirical models of discrete games. *Journal of Econometrics 48*(1-2), 57–81.

Bresnahan, T. F. and P. J. Reiss (1991b). Entry and competition in concetrated markets. *Journal of Political Economy 99*(5), 977–1009.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds, estimation and inference. *Econometrica 81*(2), 667–737.

Chesher, A. and A. Rosen (2017). Generalized instrumental variable models. *Econometrica 85*(3), 959–989.

Chesher, A. and A. Rosen (2020). Econometric modeling of interdependent discrete choice with applications to market structure. CeMMAP working paper CWP25/20.

Chesher, A., A. M. Rosen, and Z. Siddique (2019). Estimating endogenous effects on ordinal outcomes. CeMMAP Working Paper 66/19.

Chesher, A. and K. Smolinski (2012). IV models of ordered choice. *Journal of Econometrics 166*, 33–48.

Chetverikov, D. (2017). Adaptive test of conditional moment inequalities. *Econometric Theory 34*(1), 186–227.

Ciliberto, F. and E. Tamer (2009, November). 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.

Dudley, R. (1984). A course on empirical processes. *Ecole d'Été de Probabilitiés de Saint-Flour, XII-1982. Lecture Notes in Math 1097*, 1–142.

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.

Götze, F. (1991). On the rate of convergence in the multivariate clt. *The Annals of Probability 19*, 724–739.

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.

Holmes, T. J. (2011). The diffusion of wal-mart and economies of density. *Econometrica 79*(1), 253–302.

Honoré, B. and J. Powell (2005). Pairwise difference estimators for nonlinear models. In D. Andrews and J. Stock (Eds.), *Identification and Inference for Econometric Models*, pp. 520–553. Cambridge.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics 58*, 71–120.

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.

Kaido, H. and H. White (2014). A two-stage procedure for partially identified models. *Journal of Econometrics 182*(1), 5–13.

Kline, B. (2015). Identification of complete information games. *Journal of Econometrics 189*, 117–131.

Kline, B. and E. Tamer (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics 7*, 329–366.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

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

Lee, S., K. Song, and Y.-J. Whang (2018). Testing for a general class of functional inequalities. *Econometric Theory* (34), 1018–1064.

Lehmann, E. (1999). *Elements of Large-Sample Theory*. Springer.

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.

Moschopoulos, P. and W. Canada (1984). The distribution function of a linear combination of chi-squares. *Computers and Mathematics with Applications 10*, 383–386.

Nishida, M. (2015). Estimating a model of strategic network choice: The convenience-store industry in okinawa. *Marketing Science 34*, 20–38.

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 (2015). Moment inequalities and their application. *Econometrica 83*(1), 315–334.

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

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.

Powell, J., J. Stock, and T. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*, 1403–1430.

Raič, M. (2019). A multivariate berry-esseen theorem with explicit constants. *Bernoulli 25*, 2824–2853.

Romano, J., A. Shaikh, and M. Wolf (2014, September). A practical two-step method for testing moment inequalities. *Econometrica 82*(5), 1979–2002.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. New York, NY.

Sherman, R. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *Annals of Statistics 22*, 439–459.

Shi, X. and M. Shum (2015). Simple two-stage inference for a class of partially identified models. *Econometric Theory 31*, 493–520.

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

Tamer, E. (2003, January). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–167.

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

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

A Figures

Figure 1: Illustration of Restriction I4

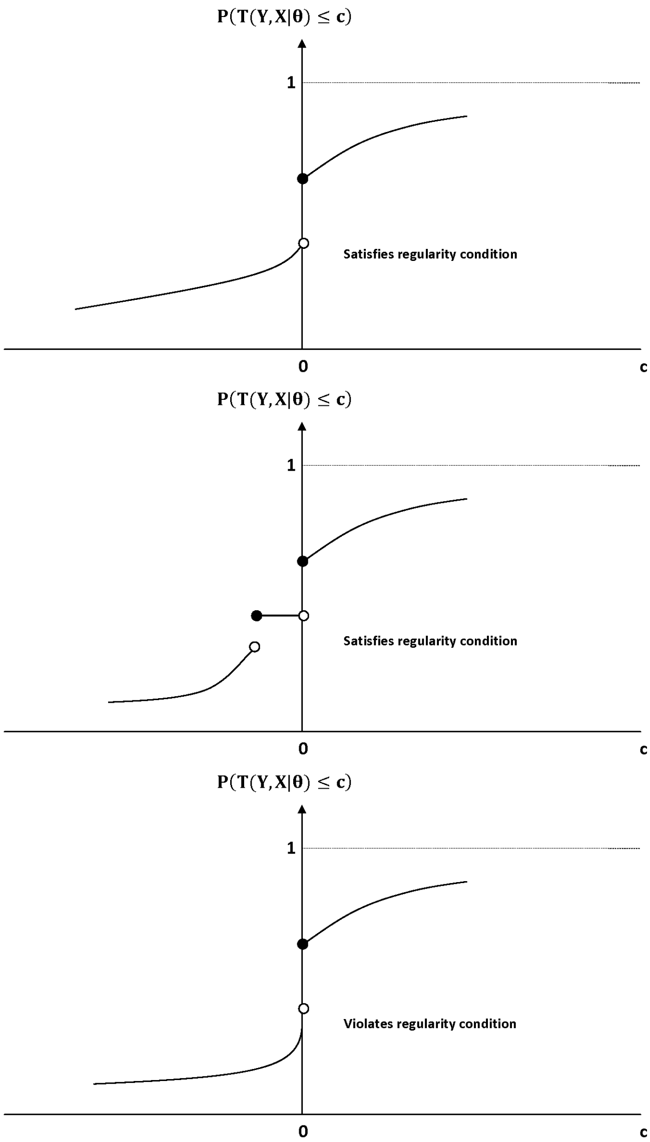

Figure 2: Joint 95% confidence region for strategic interaction coefficients