Estimation of Games with Ordered Actions: An Application to Chain-Store Entry

Andres Aradillas-Lopez and Amit Gandhi
Pennsylvania State University and University of Wisconsin-Madison
December 3, 2015

Abstract

We study the estimation of static games where players are allowed to have ordered actions, such as the number of stores to enter into a market. Assuming that payoff functions satisfy general shape restrictions, we show that equilibrium of the game implies a covariance restriction between each player’s action and a component of the player’s payoff function that we call the “strategic index”. The strategic index captures the direction of strategic interaction (i.e., patterns of substitutability or complementarity) as well as the relative effects of opponents’ decisions on players’ payoffs. The covariance restriction we derive is robust to the presence of multiple equilibria, and provides a basis for identification and estimation of the strategic index. We introduce an econometric method for inference in our model that exploits the information in moment inequalities in a computationally simple way. We analyze its properties through Monte Carlo experiments and then apply our approach to study entry behavior by chain stores where there is both an intensive margin of entry (how many stores to open in a market) as well as the usual extensive margin of entry (whether to enter a market or not). Using data from retail pharmacies we find evidence of asymmetries in strategic effects among firms in the industry, which has implications for merger policy. We also find that business stealing effects are less pronounced in larger markets, which helps explain the large positive correlation in entry behavior observed in the data.

Keywords: Static games, multiple equilibria, partial identification, conditional moment inequalities, entry decisions.

1 Introduction

The econometric analysis and applications of static games has been an increasingly active area of research in the recent past. A partial list of papers would include Bjorn and Vuong (1984), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Berry (1992), Tamer (2003), Seim (2006), Davis (2006), Berry and Tamer (2006), Pesendorfer and Schmidt-Dengler (2008), Sweeting (2009), Aradillas-Lopez (2010), Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), Bajari, Hong, Kreiner, and Nekipelov (2009), Bajari, Hong, and Ryan (2005), Ciliberto and Tamer (2009), Kline and Tamer (2010), Gowrisankaran and Krainer (2011), Aradillas-Lopez (2011), De Paula and Tang (2012), Lewbel and Tang (2012) and Grieco (2012). Most of the existing econometric work on static games has been characterized by at least one of two features: (i) a full parametrization of payoff functions with fairly limited forms of strategic effects (e.g., constant strategic effects), and (ii) a limited strategy space, with binary choice games being the most common example. One of the major difficulties with using richer models of strategic interaction in empirical work is that the multiplicity of equilibria can complicate the use of methods which require computing the equilibria in the game. Furthermore, even inferential approaches that rely solely on necessary conditions in equilibrium could also become impractical because characterizing such conditions can be difficult if the game has a rich strategy space.
In this paper we study static games with a rich, possibly unbounded strategy space that is only required to be ordered in nature (and can be discrete or continuous). Players’ payoffs are left nonparametrically specified except for a component that summarizes the strategic interaction effect. This “strategic index” captures the direction of strategic interaction (i.e., patterns of substitutability or complementarity) as well as the relative effects of opponents’ strategies on players’ payoffs as well as the potentially continuous variation in these effects with observable covariates (i.e., market size, demographics, etc) in an empirically flexible way. Instead of fully parameterizing payoff functions we only impose weak shape restrictions on payoffs that are motivated by economic theory. Our main result is showing that these shape restrictions alone are sufficient for doing inference on the strategic index in a way that is fully robust to the presence of multiple equilibria.

The key idea in the paper is that we exploit the multiplicity of equilibria as a source of identifying power for estimating strategic interactions. We treat multiple equilibria as a source of unobserved heterogeneity in the model—the equilibrium being played in a market is ex-ante unknown to the econometrician. We show that regardless of the distribution of this unobservable (which is not identified), the model predicts a conditional covariance between each player’s action and the strategic index she faces. We use this moment inequality as a basis for estimating the parameters underlying each player’s strategic index. Using multiple equilibria as a tool for identification has been done previously, for example, in De Paula and Tang (2012) and Sweeting (2009).

Our identification strategy is most closely related to De Paula and Tang (2012), who also exploit the identifying power of multiple equilibria in a static game context. De Paula and Tang (2012) focus on binary choice games under a symmetry assumption whereby each player places equal weight on the individual strategies of each of his opponents. Their goal is to identify the direction (i.e, the sign) of strategic interaction conditional on observable (to the econometrician) payoff shifters, which must be treated as categorical in their econometric implementation. Instead of only predicting the sign of strategic interaction in a binary choice game for a known strategic index, we are interested in estimating the parameters of the strategic index itself. We can thus address the empirical question of estimating the magnitude of the relative effects of opponents’ decisions on each player’s payoff and the variation of these effects with market observables within a rich action space. The model we study is thus considerably more general. First, we go beyond binary choice games to a much richer strategy space where actions sets are only ordered but can be quite large (possibly unbounded). Second, rather than assume symmetry in the strategic interaction effects in payoff functions, we allow players to be affected differently by the individual strategies of each opponent. Indeed, such richness of strategy spaces and strategic interactions causes a major growth in the multiplicity of equilibria, which adds to the identifying power of the model in a way that can be quite substantial for applied work (see more below). Third, we allow for these strategic interactions to vary with continuously distributed observed payoff shifters, such as market size, which reveals important information about the nature of competition and further enhances the identifying power of the model.

Our model’s testable implications takes the form of a sign restriction on a conditional covariance. By the definition of a covariance, this restriction can be expressed as an inequality involving a nonlinear transformation of conditional moments. Among existing methods for inference with conditional moment inequalities, those that avoid the use of nonparametrically estimated conditional moments and rely instead on spaces of “instrument functions” (Andrews and Shi (2011a, 2011b), Armstrong (2011a, 2011b)) are not directly applicable to our case since they are not designed to handle in general nonlinear transformations of a collection of conditional moments. In general a problem like ours requires the use of plug-in nonparametric estimators for the conditional moments involved. Along these lines, the methodology proposed in Chernozhukov, Lee, and Rosen (2011) could potentially be adapted
and applied to our problem. Its implementation would require the computation of a supremum of a particular test statistic over the target testing range of the conditioning variables. However, when these include a large number of elements with rich support, approximating this supremum with a reasonable degree of precision would pose a computational challenge. This is the case of our empirical application, where the vector of conditioning covariates includes eight continuously distributed elements. To be able to conduct inference in a setting like ours we propose an inferential approach based on a particular type of one-sided expectation\(^1\) whose construction uses plug-in nonparametric estimators. Unlike existing methods which also rely on one-sided \(L^p\) functionals in related problems (Lee, Song, and Whang (2013)), our approach is not based on a least-favorable configuration and is therefore less conservative when used to construct confidence sets. By design, our method is computationally easy to implement even in the presence of a rich model with multiple conditioning covariates with continuous support. We describe our approach in the main body of the paper and we establish its asymptotic properties in the econometric appendix.

We apply our approach to study the pattern of entry by the three major national drug store chains (CVS, Walgreens, and Rite-Aid) competing in local geographic markets. Our model allows us to study both the extensive-margin decision of whether to enter a market or not, as well as the intensive-margin decision of how many establishments to open in a market. Most papers (see e.g., Bresnahan and Reiss (1991b), Berry (1992), Seim (2006), Ciliberto and Tamer (2009)) have modeled entry exclusively as an extensive margin binary decision and have therefore have abstracted away from the intensive margin. Some exceptions to this include Davis (2006) and Gowrisankaran and Krainer (2011) but these papers rely on very strong parametric assumptions and equilibrium selection restrictions.\(^2\)

Our application shows that this intensive margin reveals many important features of competition that is obscured by the extensive margin alone. In particular, we find important evidence of asymmetries in the competition among these players which suggest that the least anticompetitive takeover of Rite-Aid by one of the competitors (a policy currently under consideration) would be CVS rather than Walgreens. We also find that evidence that the strength of strategic interactions diminishes with market size, which plays a central role in explaining the large positive correlation of entry behavior found in the data.

The rest of the paper proceeds as follows. Section 2 describes our general assumptions along with the resulting properties of our model. The observable implications that result from our model are studied in Section 3. Section 4 describes our econometric inferential procedure in semiparametric models and characterizes its asymptotic properties. Section 5 analyzes the properties of our procedure through Monte Carlo experiments. Section 6 applies our approach to entry decisions in the U.S drug store industry, modeling entry strategies as involving not only a binary choice of entry but also a capacity (number of stores) choice. Section 7 concludes. All proofs are included in the appendix.

\section{A static game with a rich strategy space}

We now present a nonparametric game with incomplete information and derive its testable implications. Our model non-parametrically generalizes three main features of existing models. First, we allow for a rich action space which includes binary choice games as a special case. This expands the scope of real world problems that can be studied through our approach. Second, we place no restrictions on the dimension and the “magnitude” of private

\(^1\)A one-sided expectation refers to an expectation of the form \(E[\max\{Z, 0\}]\).

\(^2\)Aradillas-Lopez (2011) also focuses on rich strategy spaces but the goal there is to answer a different question than the one posed here.
information nor the manner in which private information shifts the payoff function. Third, we isolate a fundamental feature of the game which aggregates the effect that rivals’ strategies have on a player’s own payoff. However, instead of imposing a full functional form on payoffs we only place general restrictions regarding the way this index enters a player’s payoffs. These restrictions formalize the idea that a larger value of the strategic index, by definition, decreases a player’s marginal payoff from increasing its own action. Our main questions would then include how the strategic index changes with the actions of players’ rivals (which would determine patterns of strategic substitutability or complementarity as well as the relative impact of rivals’ strategies on a given player), as well as how these features depend on observable characteristics of the environment. In the context of entry models the strategic index would capture the competition effect, summarizing how a firm’s marginal payoff from increasing its presence in a market is affected by the entry decisions of others. It can also help us learn how these features change from one market to another given the observable market characteristics available to the researcher.

2.1 Players and actions
We have \( p = 1, \ldots, P \) players (\( -p \) denotes the collection of all players except \( p \)), each \( p \) has a real-valued decision variable \( Y^p \), which is either binary (i.e, \( Y^p \in \{0, 1\} \)) or (if it can take on more than two values), it is ordinal in nature, with \( Y^p \in A^p \). The strategy space \( A^p \) can be unbounded, it can be discrete or continuous (or it can consist of the union of discrete and continuous sets in \( \mathbb{R} \)), and its ordered elements do not have to be evenly spaced. In fact, our identification results do not require that the econometrician know the exact structure of \( A^p \). The only restriction is that it must possess a natural order. We let \( A^{-p} = \prod_{q \neq p} A^q \) denote the action space of \( p \)'s opponents. We use lower case \( y^p \) to denote a potential action (in \( A^p \)) for \( p \) and \( y^{-p} \equiv (y^q)_{q \neq p} \) to denote a potential action profile (in \( A^{-p} \)) for \( p \)'s opponents. We use upper case letters \( (Y^p \text{ and } Y^{-p} \equiv (Y^q)_{q \neq p}) \) to denote the actions (profiles of actions) actually chosen by players. The game is simultaneous.

2.2 Payoff functions
Each player \( p \) has a payoff function that indicates the (von Neumann-Morgenstern) utility associated with their choices. The payoff for \( p \) if \( Y^p = y^p \) and \( Y^{-p} = y^{-p} \) is given by

\[
\nu^p(y^p, y^{-p}; \xi^p),
\]

(1)

\( \xi^p \) denotes \( p \)'s payoff shifters (other than opponents’ choices). For convenience and in accordance with the boundaries of \( A^p \), for any \( y^{-p} \in A^{-p} \) we decree \( \nu^p(y^p, \cdot; \cdot) = -\infty \) for any \( y^p \notin A^p \). We will partition \( p \)'s payoff shifters as

\[
\xi^p = (X, \varepsilon^p),
\]

where \( X \) is observed by the econometrician and \( \varepsilon^p \) is not. The dimension of \( \varepsilon^p \) is left unspecified and we allow \( \varepsilon^p \) and \( X \) to be correlated in an arbitrary way. We will not make assumptions here about the direction in which payoffs shift in response to particular elements of \( X \). Furthermore we will not assume the existence of player-specific observable payoff shifters. Throughout, \( X \) will denote the collection of all covariates observable to the researcher.

2.2.1 Basic restrictions on payoff functions
We assume that payoff functions can be expressed in the following way.
Assumption 1. (Generic expression of payoff functions)
\( \nu^p \) can be expressed as follows,
\[
\nu^p(y^p, y^{-p}; \xi^p) = \nu^{p.a}(y^p; \xi^p) - \nu^{p,b}(y^p; \xi^p) \cdot \eta^p(y^{-p}; X),
\]  
(2)
where \( \nu^{p,b} \) and \( \eta^p \) are real-valued functions or “indices” whose product captures the entire strategic effect of \( p \)'s opponents on his payoff function.

The key feature about \( \eta^p \) is that it depends on \( \xi^p \) solely through \( X \). While strategic interaction effects are allowed to depend on unobservable components of payoff shifters, this dependence must be fully captured by \( \nu^{p,b} \).

Expected payoff functions and Assumption 1

We assume Bayesian Nash equilibrium (BNE) behavior here. As a result, we can focus on beliefs for \( p \) that can be expressed as probability functions defined over \( A^{-p} \). For any set of beliefs \( \sigma^{-p} : A^{-p} \to [0, 1] \), the associated expected utility for \( p \) of choosing \( Y^p = y^p \) is
\[
\bar{\nu}^p_{\sigma}(y^p; \xi^p) = \sum_{y^{-p} \in A^{-p}} \sigma^{-p}(y^{-p}) \cdot \nu^p(y^p, y^{-p}; \xi^p)
\]
\[
= \nu^{p.a}(y^p; \xi^p) - \nu^{p,b}(y^p; \xi^p) \cdot \bar{\eta}^p_{\sigma}(X),
\]  
where
\[
\bar{\eta}^p_{\sigma}(X) = \sum_{y^{-p} \in A^{-p}} \sigma^{-p}(y^{-p}) \cdot \eta^p(y^{-p}; X).
\]
A key feature of \( p \)'s beliefs is that they do not depend on \( p \)'s own action. This independence is the defining feature of Nash equilibrium as opposed, e.g., to correlated equilibrium.

Our model will normalize the “strategic meaning” of the index \( \eta^p(y^{-p}; X) \) by assuming that \( \nu^{p,b}(\cdot; \xi^p) \) is nondecreasing w.p.1. This in turn will imply that the marginal gain for \( p \) of increasing his own strategy is nonincreasing in the expected value of the strategic index \( \eta^p \).

Assumption 2. (Marginal benefit of \( Y^p \) is nonincreasing in \( \eta^p \)) With probability one in \( \xi^p \), the function \( \nu^{p,b}(\cdot; \xi^p) \) is nondecreasing over \( A^p \). That is, for any \( v > u \in A^p \) we have \( \nu^{p,b}(v; \xi^p) \geq \nu^{p,b}(u; \xi^p) \) w.p.1.

Take any pair of actions \( v > u \in A^p \). Take any pair of beliefs \( \sigma^{-p} \) and \( \sigma^{-p'} \). Then,
\[
[\bar{\nu}^p_{\sigma}(v; \xi^p) - \bar{\nu}^p_{\sigma}(u; \xi^p)] - [\bar{\nu}^p_{\sigma'}(v; \xi^p) - \bar{\nu}^p_{\sigma'}(u; \xi^p)] = [\bar{\eta}^p_{\sigma}(X) - \bar{\eta}^p_{\sigma'}(X)] \cdot [\nu^{p,b}(v; \xi^p) - \nu^{p,b}(u; \xi^p)].
\]
Therefore by Assumption 2,
\[
\bar{\eta}^p_{\sigma}(X) \geq \bar{\eta}^p_{\sigma'}(X) \implies \bar{\nu}^p_{\sigma}(v; \xi^p) - \bar{\nu}^p_{\sigma}(u; \xi^p) \leq \bar{\nu}^p_{\sigma'}(v; \xi^p) - \bar{\nu}^p_{\sigma'}(u; \xi^p) \quad \forall u < v \in A^p
\]  
(4)
The “shape” restriction described in Assumption 2 will be the key to our identification results. It is reminiscent of conditions found in the supermodular game literature (more precisely, it amounts to a supermodularity property for \( -\nu^{p,b} \); see Topkis (1998) and Vives (1999)) but our setup extends beyond supermodularity since it allows for very general patterns of pairwise complementarity or substitutability. In this paper we will not make any assumptions\(^3\) regarding how payoffs shift with specific elements in \( \xi^p \).

\(^3\)If economic theory provides ex-ante information about how payoffs should shift with some specific elements in \( \xi^p \), this information could potentially be used in order to refine the results that follow.
Observe that given Assumption 2, \( Y^q \) is a strategic substitute (complement) for \( Y^p \) if \( \eta^p(y^{-p}; \xi^p) \) is increasing (decreasing) in \( y^q \). Cournot competition (where firms compete in quantities with each other) is a classic case of a game of strategic substitutes. In that case \( \eta^p(y^{-p}; X) \) would be increasing in each element of \( y^{-p} \). Conversely, if an increase in player \( q \)'s action \( Y^q \) lowers \( \eta^p \), then by Assumption 2 it increases the marginal gain to player \( p \) from increasing its actions and thus \( Y^q \) would be a strategic complement for \( Y^p \). Bertrand competition (where firms compete in prices with each other) is a classic case of a game of strategic complements. Note that Assumption 2 allows for any pattern of pairwise complementarity or substitutability between players’ strategies. Whether player \( q \)'s strategy is a complement or a substitute for player \( p \)'s will be determined by whether the index \( \eta^p \) is decreasing or increasing in \( y^q \).

### 2.3 Example: A structural model of imperfect competition

It is useful to contextualize our setup within a well-known structural economic model. Consider a model of Cournot competition between \( P \) firms with differentiated products. To avoid confusion with our notation (where we have used ‘\( p \)’ to denote each player and \( P \) as the total number of players) let us use script typeface letters to denote prices \( \mathcal{P} \) and quantities \( \mathcal{Q} \). Suppose the model is described by a linear demand system where

\[
\mathcal{Q}^p = \sum_{q=1}^{P} d^{p,q}(\xi^p) \cdot \mathcal{P}^q + f^p(\xi^p), \quad \text{for } p = 1, \ldots, P.
\]

Suppose \( \sum_{q=1}^{P} d^{p,q}(\xi^p) \neq 0 \) w.p.1 (an assumption grounded on economic theory). Define \( \zeta^p(\xi^p) \equiv f^p(\xi^p)/\sum_{q=1}^{P} d^{p,q}(\xi^p) \). Our assumptions will imply restrictions on the structure of the coefficients \( d^{p,q}(\xi^p) \). Specifically, suppose we can express \( d^{p,q}(\xi^p) = \phi^p(\varepsilon^p) \cdot a^{p,q}(X) \). The demand system can be expressed as

\[
\mathcal{Q}^p = \phi^p(\varepsilon^p) \cdot \sum_{q=1}^{P} a^{p,q}(X) \cdot \left( \mathcal{P}^q + \zeta^p(\xi^p) \right), \quad \text{for } p = 1, \ldots, P.
\]

Let \( A(X) \) denote a \( P \times P \) matrix where \( [A(X)]_{p,q} = a^{p,q}(X) \) and let \( D(\phi(\varepsilon)) \) denote a \( P \times P \) diagonal matrix where \( [D(\phi(\varepsilon))]_{p,p} = \phi^p(\varepsilon^p) \). By our above assumption the last matrix is invertible w.p.1. Suppose this is also true for \( A(X) \) and denote \( [A(X)^{-1}]_{p,q} \equiv b^{p,q}(X) \). Then inverse demands are of the form

\[
\mathcal{P}^p = \frac{1}{\phi^p(\varepsilon^p)} \sum_{q=1}^{P} b^{p,q}(X) \cdot \mathcal{Q}^q - \zeta^p(\xi^p), \quad \text{for } p = 1, \ldots, P.
\]

Denote firm \( p \)'s cost function as \( C^p(\mathcal{Q}^p; \xi^p) \), which can be entirely unrestricted (e.g. it can include a fixed cost and need not have to display increasing marginal costs). Profit functions are of the form,

\[
\pi^p(\mathcal{Q}^p; \mathcal{Q}^{-p}; \xi) = \left( \frac{1}{\phi^p(\varepsilon^p)} \sum_{q=1}^{P} b^{p,q}(X) \cdot \mathcal{Q}^q - \zeta^p(\xi^p) \right) \cdot \mathcal{Q}^p - C^p(\mathcal{Q}^p; \xi^p).
\]
In a Cournot model firms compete in quantities, so \( Y^p = Q^p \). This model fits our representation of payoffs (profits) in (2). We have \( \nu^p(y^p; y^{-p}; \xi) = \nu^{p,a}(y^p; \xi) - \nu^{p,b}(y^p; \xi) \cdot \eta^p(y^{-p}; X) \), where

\[
\begin{align*}
\nu^{p,a}(y^p; \xi) &= \left( \frac{1}{\phi^p(\varepsilon^p)} b^{p,a}(X) \cdot y^p - \xi^p(\xi^p) \right) \cdot y^p - C^p(y^p; \xi^p), \\
\nu^{p,b}(y^p; \xi) &= \frac{y^p}{\phi^p(\varepsilon^p)}.
\end{align*}
\]

In order to satisfy Assumption 2 it suffices that the function \( \phi^p(\varepsilon^p) \) be of constant sign. Given our structural model, it is natural to assume that \( \phi^p(\varepsilon^p) \geq 0 \) w.p.1. \( \phi^p(\varepsilon^p) > 0 \) w.p.1. given our invertibility assumptions. In this case the strategic index would be

\[
\eta^p(y^{-p}; \xi) = - \sum_{q \neq p} b^{p,q}(X) \cdot y^q
\]

the \( q^{th} \) good will be a substitute for the \( p^{th} \) good if \( b^{p,q}(X) \leq 0 \). Otherwise it will be a complement. Note that, since \( [A(X)^{-1}]_{p,q} = b^{p,q}(X) \), the strategic indices \( \eta^p \) allows us to recover \( A(X) \), a key structural component of the model..

Suppose instead that we have a log-linear system of demand,

\[
\log (Q^p) = \sum_{q=1}^{P} d^{p,q}(\xi^p) \cdot \log (P^q) + f^p(\xi^p), \quad \text{for } p = 1, \ldots, P.
\]

Now the coefficients \( d^{p,q}(\xi^p) \) directly measure elasticities of demand. In this case our assumptions imply a different set of restrictions. We now need \( d^{p,q}(\xi^p) = d^{p,q}(X) \) (privately observed shocks \( \varepsilon^p \) should now be excluded from these elasticities). Suppose \( \sum_{q=1}^{P} d^{p,q}(X) \neq 0 \) w.p.1 for each \( p \) (a reasonable assumption given the homogeneity properties of demand). Define \( \lambda^p(\xi^p) = f^p(\xi^p) / \sum_{q=1}^{P} d^{p,q}(X) \). Then the demand system can be re-written as

\[
\log (Q^p) = \sum_{q=1}^{P} d^{p,q}(X) \cdot \left( \log (P^q) + \lambda^p(\xi^p) \right), \quad \text{for } p = 1, \ldots, P.
\]

Let us maintain that the \( P \times P \) matrix \( D(X) \) where \( [D(X)]_{p,q} = d^{p,q}(X) \) is invertible w.p.1 and denote \( [D(X)^{-1}]_{p,q} \equiv r^{p,q}(X) \). Inverting the demand system we obtain the following inverse demands,

\[
P^p = e^{-\lambda^p(\xi^p)} \cdot \prod_{q=1}^{P} (Q^q)^{r^{p,q}(X)}, \quad \text{for } p = 1, \ldots, P.
\]

Profit functions are now of the form

\[
\pi^p (Q^p, Q^{-p}; \xi^p) = e^{-\lambda^p(\xi^p)} \cdot (Q^p)^{r^{p,p}(X)+1} \cdot \prod_{q \neq p} (Q^q)^{r^{p,q}(X)} - C^p (Q^p; \xi^p).
\]

Define \( \nu^{p,a}(y^p; \xi^p) = -C^p(y^p; \xi^p) \). For \( \nu^{p,b} \) and \( \eta^p \) we can proceed as follows. Satisfying the condition in Assumption 2 depends on the sign of \( r^{p,p}(X) + 1 \). It is easy to see that our payoff representation in (2) and the
condition in Assumption 2 will be satisfied if we define
\[
\nu_{p,b}(y^p; \xi^p) = e^{-\lambda_p^p(\xi^p)} \cdot \left( \mathbb{I} \{ r^{p,p}(X) \geq -1 \} - \mathbb{I} \{ r^{p,p}(X) < -1 \} \right) \cdot (y^p)_{r^{p,p}(X)},
\]
\[
\eta^p(y^{-p}; X) = \left( \mathbb{I} \{ r^{p,p}(X) \geq -1 \} - \mathbb{I} \{ r^{p,p}(X) < -1 \} \right) \cdot \prod_{q \neq p} (y^q)_{r^{p,q}(X)}.
\]
Suppose \( r^{p,p}(X) \geq -1 \). Then the \( q^{th} \) good is a substitute for the \( p^{th} \) good if \( r^{p,q}(X) > 0 \) and it is a complement otherwise. If \( r^{p,p}(X) < -1 \) then this holds with the reverse the inequalities. Once again the index \( \eta^p(y^{-p}; X) \) has a structural interpretation as it contains information about the relative price elasticities in the demand system.

Using the demand systems described above we could also study competition in prices instead of quantities. In that case our assumptions would place restrictions on firms’ cost functions while allowing more flexibility in the specification of demand functions compared to the Cournot case (which in placed no restrictions on firms’ cost functions as we showed above).

### 2.4 Strategic interaction features captured by the index \( \eta^p \)

Given our payoff representation, the overall scale of the strategic effect would be absorbed into the term \( \nu^{p,b} \).

While the index \( \eta^p \) would not capture the overall scale of strategic interaction it would nevertheless summarize the following key features of strategic interaction in the model,

(i) The directional patterns of strategic interaction between any subset of players: This is captured by the direction in which the strategic indices move in response to rivals’ actions.

(ii) The relative magnitude of the effects of strategic interaction between one player and each one of his opponents: This is captured by the relative magnitude in which the strategic indices shift in response to each rival’s action.

As we illustrated in the previous section, different conjectures involving these strategic features can be incorporated directly into the structure of \( \eta^p \).

### 2.5 Information and behavior

An incomplete information setting allows us to focus on pure strategy equilibria in which players strictly best respond to each other in equilibrium. Such an equilibrium restriction is natural for empirical work because equilibria can then be interpreted as a steady state outcome, which mixed strategy equilibria does not allow.\(^4\) The empirical appeal of pure strategy equilibria is the key motivation for Harsanyi’s well known purification theorem (Harsanyi (1973)). His result showed that when there exist (potentially small) private information about own payoffs in a normal form game, then this ensures that all equilibria generically take this pure strategy form.\(^5\) He modeled private information shocks to be idiosyncratic and hence independent across players. We follow in this approach and assume that the private payoff shocks to firms are independent conditional on publicly observable payoff shifters, but we analyze what happens when this assumption is violated in our Monte Carlo experiment section.

---

\(^4\)Mixed strategies force one to question why it is the case that when a player is indifferent among several strategies, he or she mixes over these strategies in exactly such a way that makes the other player indifferent. For a further discussion see Morris (2008).\(^5\)The second part of his result, the so called “approachability” party, showed that the set of pure strategy equilibria in the perturbed private information game is arbitrarily close to the set of all mixed strategy equilibria of the corresponding unperturbed complete information game.
Assumption 3. (Independent private shocks)

$X$ is perfectly observed by all players, but $\varepsilon^p$ is only privately observed by $p$. We assume that each $\varepsilon^p$ is independent of $\varepsilon^{-p}$ conditional on $X$. The true distribution of $(X, (\varepsilon^p)^P_{p=1})$ is common knowledge among the players, as are the functional forms of payoff functions $(\nu^p)^P_{p=1}$. Thus, the only source of incomplete information for $p$ is the realization of $\varepsilon^{-p}$.

The dimension of $\varepsilon^p$ is left unspecified and we allow $\varepsilon^p$ and $X$ to be correlated in an arbitrary way. A special case of Assumption 3 is one where some player $p$ possesses no private information and $\xi^p = X$ (recall that $\xi^p = (X, \varepsilon^p)$).

Thus, a game of complete information would be a special case of our setting as long as $\xi^p = X$ for each $p$. In this case the only source of unobserved heterogeneity for the econometrician would be the equilibrium selection mechanism. Assumption 3 is strong, and for this reason in our Monte Carlo section we study the properties of our procedure when there is in fact correlation in private shocks. We find there that, as long as the correlation is relatively minor conditional on $X$, the properties of our inferential method will remain relatively unchanged.

### 2.5.1 Bayesian Nash equilibrium (BNE) behavior

We maintain that the outcome observed is the result of a BNE of the underlying game. Given the independent-private shock restriction in Assumption 3, any BNE can be characterized as a collection of conditional (on $X$) probability functions $\{\sigma^p(\cdot|X) : \mathcal{A}^p \rightarrow [0, 1]\}^P_{p=1} \equiv \sigma_*(X)$ with corresponding expected utility functions

$$
\mathcal{P}^p_{\sigma_*}(\cdot; \xi^p) = \sum_{y^{-p} \in \mathcal{A}^{-p}} \sigma_*^{-p}(y^{-p}|X) \cdot \nu^p(\cdot, y^{-p}; \xi^p),
$$

where, for each $y^{-p} \equiv (y^q)^{q \neq p} \in \mathcal{A}^{-p}$ and $y^p \in \mathcal{A}^p$,

$$
\sigma_*^{-p}(y^{-p}|X) = \prod_{q \neq p} \sigma_*^q(y^q|X) \quad \text{and} \quad \sigma_*^p(y^p|X) > 0 \quad \text{only if} \quad y^p \in \arg\max_{y^p \in \mathcal{A}^p} \mathcal{P}^p_{\sigma_*}(y; \xi^p).
$$

**Assumption 4.** The outcome observed is the realization of a BNE. That is,

$$
Y^p \in \arg\max_{y^p \in \mathcal{A}^p} \mathcal{P}^p_{\sigma_*}(y; \xi^p) \text{ for some BNE } \sigma_*(X).
$$

For a given realization of payoff shifters, multiple BNE may exist and we leave the underlying selection mechanism $S$ unspecified except for the assumption that it always picks a BNE $\sigma_*(X)$ such that the resulting expected payoff function $\mathcal{P}^p_{\sigma_*}(\cdot; \xi^p)$ has a unique optimal choice.

Assuming pure-strategy play in games of incomplete information is not a very restrictive assumption. Recall the above discussion that Harsanyi’s purification theorem ensures that the restriction to pure strategy equilibria with unique best responses is generically without loss of generality in (finite) incomplete information games. In more general games of incomplete information where player types are conditionally independent—the type of setting we assume here—Milgrom and Weber (1985) show that every mixed strategy equilibrium has a nearby “purification” pure strategy such that the distribution of players’ observed behavior and expected payoffs are identical. The setup described in Assumptions 1-4 encompasses many existing static models of incomplete information as special cases. The two key restrictions are either binary or otherwise ordinal nature of the strategy space, and conditional

---

Note that Assumption (4) implicitly imposes an additional restriction on payoff functions. Namely, the existence of equilibria where each player has a unique best-response. Sufficient conditions can be made precise in the context of specific structural models (see our examples in Section 2.3).
independence of players’ private information. Examples include, among others, Seim (2006), the binary-choice (i.e., two-time period) model in Sweeting (2009), the general setup in De Paula and Tang (2012), binary-choice or ordinal-choice versions of the incomplete information games studied in Bajari, Hong, Kreiner, and Nekipelov (2009), and of the Quantal Response Equilibrium (QRE) model proposed in McKelvey and Palfrey (1995) and studied further in Haile, Hortacsu, and Kosenok (2008). Our setup can also encompass complete-information games under the restriction \( \xi^p = X \) in which case it would encompass, for example, the binary-choice games in Bresnahan and Reiss (1990) and Tamer (2003). Beyond the existing literature, our assumptions can handle models under unprecedentedly general assumptions regarding players’ payoffs which include discrete or continuous strategy spaces. While our results may retain some of their validity if there is relatively small correlation in players’ private shocks (see our experimental results in Section 5), static games where the strategy space does not have a natural order (e.g., truly multinomial-choice games) are entirely outside the scope of our paper.

3 Implications of Assumptions 1-4

3.1 Properties of players’ best-responses

As we stated above, we focus on equilibrium beliefs that yield a unique optimal choice to players. For any such set of beliefs our payoff shape restrictions imply a monotonicity property between optimal actions and the expected value of the strategic index induced by each player’s beliefs. We describe this next.

Result 1. Let \( \sigma_p \) and \( \sigma'_p \) denote any pair of beliefs that produce unique expected-payoff maximizing choices for \( p \) given the realization of \( \xi^p \), and let \( y_p^p(\xi^p) \) and \( y_p'^p(\xi^p) \) denote the corresponding optimal choices. If Assumptions 1-2 hold, then w.p.1 we have,

\[
\text{If } \pi^p_\sigma(X) \geq \pi^p_{\sigma'}(X), \text{ then } \mathbb{I}\{y_p^p(\xi^p) \leq y^p\} \geq \mathbb{I}\{y_p'^p(\xi^p) \leq y^p\} \quad \forall y^p \in A^p.
\]

Proof: In Appendix A. \( \Box \)

3.2 Main result

Let \( \sigma_{s_j} \) and \( \sigma_{s_k} \) denote any pair of existing BNE that the selection mechanism \( S \) could choose with positive probability. By Result 1, w.p.1 we must have,

\[
\text{If } \pi^p_{\sigma_{s_j}}(X) \geq \pi^p_{\sigma_{s_k}}(X), \text{ then } \mathbb{I}\{y^p_{\sigma_{s_j}}(\xi^p) \leq y^p\} \geq \mathbb{I}\{y^p_{\sigma_{s_k}}(\xi^p) \leq y^p\} \quad \forall y^p \in A^p.
\]

Our main result will follow from here and the independence condition in Assumption 3.

Theorem 1. Let \( y^p \) be given. If Assumptions 1-4 hold, then w.p.1 in \( X \) we have

\[
E[\mathbb{I}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p};X)\mid X] \geq E[\mathbb{I}\{Y^p \leq y^p\}\mid X] \cdot E[\eta^p(Y^{-p};X)\mid X] \quad \forall y^p. \tag{5}
\]

Proof: In Appendix A. \( \Box \)

If the underlying game has a unique equilibrium w.p.1 –or more generally if it has a degenerate equilibrium selection mechanism– we would have \( Y^p \perp Y^{-p}\mid X \) and therefore any measurable function \( g(Y^{-p};X) \) should satisfy
Theorem 1 as an equality. Therefore Theorem 1 provides identification power for $\eta^p$ only if the underlying game has multiple equilibria for at least a subset of realizations of payoff shifters and if players randomize across such equilibria.

**Remark 1.** De Paula and Tang (2012) derived the conditions in Theorem 1 for the case of binary choice games. Their result relies essentially on the same conditions as Assumptions 1 and 3, along with Nash equilibrium behavior (Assumption 4 in our case). Our primary contribution is introducing Assumption 2 and showing that it is sufficient to extend the results from binary choice games to general ordered choices. In addition, while De Paula and Tang (2012) impose the symmetry restriction that each player cares equally about the decisions of every opponent, we explicitly allow for asymmetric effects; and while their econometric approach presupposes that $X$ is discrete, ours will allow for continuous covariates. This will allow us to study more carefully the properties of the strategic interaction index $\eta^p$ as a function of $X$ rather than focusing only on the direction (the sign) of strategic interaction effects as in De Paula and Tang (2012).

**Remark 2.** Two important results follow from Theorem 1.

(i) **No-strategic interaction.** Our assumptions cannot rule out that there is no strategic interaction effect at all, since $\eta^p(Y^{-p}; X) = g(X)$ would satisfy Theorem 1 as an equality (a fact that is also true in De Paula and Tang (2012)). The value of Theorem 1 is its ability to help us test many different conjectures about strategic interaction under the assumption that some interaction effect exists.

(ii) **Rejecting unique equilibria.** As we pointed out above, if the underlying game has a unique equilibrium w.p.1—or more generally if it has a degenerate equilibrium selection mechanism—then any measurable function $g(Y^{-p}; X)$ should satisfy Theorem 1 as an equality. Therefore, if we maintain the assumptions in our model, the existence of such a function $g(Y^{-p}; X)$ that violates the result in Theorem 1 would immediately reject the notion that the game has a unique equilibrium w.p.1. In particular, this would reject the assertion that there is no strategic interaction in the model.

(iii) **Rejecting our model.** Under the assumptions of our model there must exist a function $\eta^p(Y^{-p}; X)$ that satisfies the result in Theorem 1. Thus, ruling out the existence of such a function would immediately reject our model. Thus, our set of assumptions is falsifiable and could be tested nonparametrically.

The rest of our paper will be devoted to using Theorem 1 to do inference on the strategic index $\eta^p$ in a context where this index is assumed to belong to a parametric family of functions while leaving every other aspect of the model nonparametric.

## 4 Inference of Strategic Interactions in a Semiparametric Model

The result in Theorem 1 does not rely on a parametric specification for the strategic interaction index $\eta^p$. If we assume symmetry of interaction effects so that players care equally about the actions of each opponent, the strategic index could simply be $\eta^p(Y^{-p}; X(\theta^p)) = \sum_{q \neq p} \pm Y_q$, where the sign $\pm$ would indicate whether actions are substitutes or complements and this in turn would be identified through the restriction in Theorem 1. The model then would remain fully nonparametric throughout. If we want to allow for asymmetries and more complexity in the interaction effects we need a more flexible characterization of the strategic index. We will focus now on the case
where the strategic interaction index $\eta^p$ is assumed to belong to a parametric family of functions of the form

$$\eta^p(Y^{-p}; X|\theta^p),$$

with all other elements of the model left nonparametrically specified. In the examples of Section 2.3, this could be done by specifying a parametrization for the matrix $A(X)$ in the case of linear demands, and for the matrix $D(X)$ in the log-linear case. All other components of the structural models would be left unspecified in both cases. Let $\theta = (\theta^p)_{p=1}^P$ and let $\Theta$ denote the parameter space. The true value of $\theta$ will be denoted by $\theta_0$. For given $y^p$, $x$ and $\theta^p$ define

$$F_{Y^p}(y^p|x) = E_{Y^p|X}[\mathbb{I}\{Y^p \leq y^p\}|X = x],$$

$$\lambda^p(x; \theta^p) = E_{Y^{-p}|X}[\eta^p(Y^{-p}; x|\theta^p)|X = x],$$

$$\mu^p(y^p|x; \theta^p) = E_{Y|X}[\mathbb{I}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; x|\theta^p)|X = x],$$

$$\tau^p(y^p|x; \theta^p) = E_{Y^p}(y^p|x) \cdot \lambda^p(x; \theta^p) - \mu^p(y^p|x; \theta^p).$$

Theorem 1 predicts that for each $p$,

$$\tau^p(y^p|X; \theta^p_0) \leq 0 \text{ w.p.1 in } X \forall y^p \in A^p.$$

The econometrician is not required to know the exact structure of $A^p$. Since $\text{Supp}(Y^p) \subseteq A^p$, it is natural to focus on testing the above inequality over $y^p \in \text{Supp}(Y^p)$. For this reason, we choose to test whether the inequality holds over $\text{Supp}(Y^p, X)$. Therefore, our inferential approach is based on the fact that our model predicts,

$$Pr(\tau^p(Y^p|X; \theta^p_0) \leq 0) = 1. \quad (6)$$

We will propose an inferential method based on the restriction in (6) and refer to the identified set $\Theta^I$ as the collection of parameter values that satisfy (6). That is,

$$\Theta^I = \{\theta \in \Theta : Pr(\tau^p(Y^p|X; \theta^p) \leq 0) = 1 \forall p = 1, \ldots, P\}. \quad (7)$$

Note that the restriction in (6) involves inequalities of nonlinear transformations of conditional moments (the conditional covariance involves the product of two conditional expectations). Developing methods for inference with conditional moment inequalities has been an area of active research in the recent past. There are generically speaking two types of methods. The first type avoids having to estimate the conditional expectations involved and relies instead on instrument functions. Examples of this approach include Armstrong (2011a, 2011b) and Andrews and Shi (2011a, 2011b). Suppose $m(W; \theta)$ is a vector of known functions such that $E[m(W; \theta_0)|X] \leq 0$ w.p.1. Let $\mathcal{G}$ be a space of measurable, nonnegative functions of $X$. Then the previous inequality implies that we must have $E[m(W; \theta_0) \cdot g(X)] \leq 0$ for all $g \in \mathcal{G}$. Thus, for a given choice of $\mathcal{G}$ the conditional moment inequality implies unconditional moment inequality restrictions everywhere on $\mathcal{G}$. Cramer von Mises or Kolmogorov-Smirnov test-statistics can be constructed from here. This approach has the great advantage of not having to rely on smoothness assumptions about the conditional moments. However, it is not applicable here since our problem involves a nonlinear transformation of conditional moments and therefore it cannot be written as $E[m(W; \theta_0)|X] \leq 0$ for a known function $m(\cdot)$.
The second type of approach relies on plug-in estimators of the conditional moments involved. Most of the existing work in this area has been devoted to testing nonparametric restrictions rather than doing inference on a finite dimensional parameter. One notable exception is Chernozhukov, Lee, and Rosen (2011). Based on their approach, we would test whether $\theta_p$ satisfies our restrictions for player $p$ over a range $(y_p, x) \in W$ by using a test-statistic of the form

$$\hat{\nu}_p(\theta_p) = \inf_{(y_p, x) \in W} \left[ (-\hat{\tau}_p(y_p|\theta_p)) + \hat{k}(\alpha) \cdot \hat{\sigma}_p(y_p|\theta_p) \right],$$

where $\hat{\sigma}_p$ is an estimator of the standard error of $\hat{\tau}_p$ and $\hat{k}(\alpha)$ is a critical value based on the $\alpha^\text{th}$ quantile of a particular process. We would reject the inequalities for $\theta_p$ if $\hat{\nu}_p(\theta_p) < 0$ and fail to reject them otherwise.

While this method works in principle, in practice being able to compute the statistic with precision can be a computational challenge when $X$ includes a large number of covariates with rich support. This will be the case in our empirical application where $X$ includes 8 such covariates. In this case it is not clear how to do a grid search in eight dimensions in order to compute the test-statistic (and approximate the critical value) with a reasonable degree of precision, especially if the parametrization of our strategic index $\eta_p$ is such that $\tau_p(y|\theta_p)$ is nonseparable in $\theta_p$. In such cases the critical value $\hat{k}(\alpha)$ would also depend on $\theta_p$ further complicating its use for the construction of a confidence set.

Since the instrument-function approach does not apply to our setting and since procedures that rely on computing the supremum over $X$ of a semiparametric test-statistic can pose significant computational challenges when $X$ is large (as in our empirical example), we propose a different approach. Our method will be based on an unconditional mean-zero restriction implied by our inequalities. We describe it next.

### 4.1 Expressing our inequalities using unconditional mean-zero restrictions

For a given $\theta_p$ consider the following one-sided expectation,

$$T_p(\theta_p) = E_{Y_p,X} \left[ \max \{ \tau_p(Y_p|X; \theta_p), 0 \} \right]$$

Note that $T_p(\theta_p) \geq 0$ for any $\theta_p$. For a given $\theta = (\theta_p)_{p=1}^P$ let

$$T(\theta) = \sum_{p=1}^P T_p(\theta_p).$$

Note that $T(\theta) \geq 0 \forall \theta$ and $T(\theta) = 0$ if and only if $\theta \in \Theta^I$. Therefore we can re-express the identified set as

$$\Theta^I = \{ \theta \in \Theta : T(\theta) = 0 \}$$

Our method will rely on nonparametric plug-in estimators, focusing on the expectations defined above taken over an inference range where our estimators satisfy uniform asymptotic properties. Let $\mathcal{X} \subset \text{Supp}(X)$ denote a pre-specified set such that

$$x^c \in \text{int} \left( \text{Supp} \left( X^c|X^d = x^d \right) \right) \quad \forall (x^c, x^d) \in \mathcal{X}.$$
We maintain the assumption that \( f_X(x) \geq \ell > 0 \) for all \( x \in \mathcal{X} \). Let \( \mathbb{I}_X(x) \) denote a “trimming” function such that \( \mathbb{I}_X(x) = 0 \) if \( x \notin \mathcal{X} \) and \( \mathbb{I}_X(x) > 0 \) otherwise. Let

\[
T^p_X(\theta^p) = E_{Y^p \cdot X} \left[ \max \{ \tau^p(Y^p|X; \theta^p), 0 \} \cdot \mathbb{I}_X(X) \right], \quad T_X(\theta) = \sum_{p=1}^P T^p_X(\theta^p). \tag{8}
\]

The inference range \( \mathcal{X} \) will be assumed to be such that the nonparametric estimators involved in our construction have appropriate asymptotic properties uniformly over it. Given our choice of \( \mathcal{X} \), we focus attention of the following superset of the identified set \( \Theta^I \)

\[
\Theta^I_{\mathcal{X}} = \{ \theta \in \Theta : T^p_X(\theta^p) = 0 \text{ for } p = 1, \ldots, P \}.
\]

Note that \( \Theta^I \subseteq \Theta^I_{\mathcal{X}} \). Also note that choosing a very limited inference range \( \mathcal{X} \) may result in a loss of identification power if we preclude realizations of \( X \) that lead to multiple equilibria (see Remark 2). Under some conditions (e.g., compactness and density uniformly bounded away from zero) we could allow for the inference range \( \mathcal{X} \) to correspond to the entire support of \( X \), or we can allow \( \mathcal{X} \) to grow with the sample size and cover the entire support of \( X \) asymptotically.

### 4.2 Summary of econometric methodology

The details of our econometric methodology are in the econometric appendix B but we provide a summary here. Our basic setting is one where the researcher observes an iid sample \((Y^p_i)_{i=1}^n, X_i \) produced by a model satisfying our assumptions. We replace the objects in (B-1) with estimators of the form

\[
\hat{T}^p_X(\theta^p) = \frac{1}{n} \sum_{i=1}^n \hat{\tau}^p(Y^p_i|X_i; \theta^p) \cdot \mathbb{I}_X(X_i), \quad \hat{T}_X(\theta) = \sum_{p=1}^P \hat{T}^p_X(\theta^p)
\]

where \( b_n \to 0 \) is a nonnegative sequence going to zero at an appropriate rate. The use of \( b_n \) will allow us to deal with the “kink” of the \( \max \{ 0, z \} \) function at \( z = 0 \) while producing asymptotically pivotal properties. To construct \( \hat{\tau}^p \) we use kernel-based estimators.

In the econometric appendix we describe conditions under which

\[
\hat{T}_X(\theta) = T_X(\theta) + \frac{1}{n} \sum_{i=1}^n \psi(Y_i, X_i; \theta) + \varepsilon_n(\theta),
\]

where \( \sup_{\theta \in \Theta} |\varepsilon_n(\theta)| = O_p \left( n^{-1/2-\epsilon} \right) \) for some \( \epsilon > 0 \).

The “influence function” \( \psi \) can be expressed as

\[
\psi(Y_i, X_i; \theta) = \sum_{p=1}^P \left( \max \{ \tau^p(Y^p_i|X_i; \theta^p), 0 \} \cdot \mathbb{I}_X(X_i) - T^p_X(\theta^p) \right) + \sum_{p=1}^P \psi^p_U(Y_i, X_i; \theta^p).
\]

\( \psi^p_U \) is the leading term in the Hoeffding decomposition of a U-statistic and it is a function of conditional expectations (projections) and is therefore identified. The function \( \psi(Y_i, X_i; \theta) \) is identified and has two key properties:

(i) \( E[\psi(Y_i, X_i; \theta)] = 0 \quad \forall \theta \in \Theta \).
(ii) Let
\[ \Theta_{X} = \{ \theta \in \Theta : \tau^p(Y^p|X;\theta^p) < 0 \text{ w.p.1 over } X, \forall p = 1, \ldots, P. \} \]
Then \( \psi(Y_i, X_i; \theta) = 0 \text{ w.p.1, } \forall \theta \in \Theta_{X}. \)

\( \Theta_{X} \) is the collection of parameter values that satisfy our inequalities as strict inequalities w.p.1 over our inference range. Let \( \sigma^2(\theta) = \text{Var}(\psi(Y_i, X_i; \theta)). \) Based on the properties outlined above, we have
\[ \sqrt{n} \hat{T}_X(\theta) = \sqrt{n} T_X(\theta) + V_n(\theta) + \xi_n(\theta), \]
where \( V_n(\theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta)) \) and \( \sup_{\theta \in \Theta} |\xi_n(\theta)| = o_p(n^{-\epsilon}) \) for some \( \epsilon > 0. \) Given these features, our statistic will be of the form
\[ \hat{t}_n(\theta) = \frac{\sqrt{n} \hat{T}_X(\theta)}{\max \{ \kappa_n, \hat{\sigma}(\theta) \}}, \]
where \( \hat{\sigma}^2(\theta) \) is an estimator of \( \sigma^2(\theta) \) and \( \kappa_n \) is a sequence converging to zero at a sufficiently slow rate (it must satisfy \( \kappa_n \cdot n^\epsilon \longrightarrow \infty \) for any \( \epsilon > 0. \) Recall from our results described above that \( \sup_{\theta \in \Theta_{X}} |\sqrt{n} \cdot \hat{t}_n(\theta)| = O_p \left( n^{-1/2-\epsilon} \right) \) for some \( \epsilon > 0. \) The use of \( \kappa_n \) allows our statistic to satisfy \( \sup_{\theta \in \Theta_{X}} |\sqrt{n} \cdot \hat{t}_n(\theta)| = o_p(1). \)

For a desired coverage probability \( 1 - \alpha, \) our CS for \( \theta_0 \) is of the form
\[ CS_n(1 - \alpha) = \{ \theta \in \Theta: \hat{t}_n(\theta) \leq c_{1-\alpha} \}, \]
where \( c_{1-\alpha} \) is the Standard Normal critical value for \( 1 - \alpha. \) By the features outlined above our CS will have correct pointwise coverage properties. Namely,
\[ \inf_{\theta \in \Theta} \liminf_{n \rightarrow \infty} P \left( \theta \in CS_n(1 - \alpha) \right) \geq 1 - \alpha. \]
Suppose we generalize our basic setting and assume that \( \{(Y_i^p)_{p=1}^n, X_i \} : 1 \leq i \leq n \} \) is a triangular array which is row-wise iid with distribution \( F_n \in \mathcal{F}. \) In order for our CS to possess correct coverage properties uniformly over \( (\mathcal{F}, \Theta) \) we need to equip \( \mathcal{F} \) with integrability conditions such that:

(i) A Central Limit Theorem for triangular arrays holds for
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Y_i, X_i; \theta_n, F_n) \]
for any sequence \( F_n \in \mathcal{F} \) and \( \theta_n \in \Theta \setminus \Theta_{X}(F_n). \)

(ii) The necessary Laws of Large Numbers for triangular arrays hold to ensure that \( |\hat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n)| = o_p(1) \)
over any sequence \( F_n \in \mathcal{F} \) and \( \theta_n \in \Theta. \)

We describe such conditions in the appendix. If they hold, then
\[ \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \frac{1}{P_{F_\Theta} \left( \theta \in CS_n(1 - \alpha) \right)} \geq 1 - \alpha \]
In the econometric appendix we also study the power properties of our approach. Unlike methods which rely on
one-sided $L^p$–functionals (e.g., Lee, Song, and Whang (2011)) our approach is not guided by a least favorable configuration. In such settings test-statistics are normalized by looking at the largest possible variance that would still be consistent with the inequalities. In our context this would amount to using a test-statistic of the form

$$\tilde{t}_n(\theta) = \frac{\sqrt{n}\hat{T}_X(\theta)}{\hat{\Omega}(\theta)},$$

where $\hat{\Omega}(\theta)$ is the estimator of $\hat{\sigma}(\theta)$ that would result if the inequalities were binding a.s. To construct it we would replace each indicator function $1\{\tau^p(Y^p|X; \theta^p) \geq 0\}$ with 1. By breaking away from least-favorable configurations our procedure is, by construction, less conservative. The cost is having to introduce the tuning parameter $\kappa_n$. By design, our methodology is computationally simple to implement even in the presence of a rich parametrization and a large collection of conditioning covariates $X$. This computational simplicity also enables us to study the sensitivity of our results to various choices of the tuning (bandwidth) parameters involved. Computing the confidence set for different values of these parameters is a computationally costless exercise.

### 4.3 On a nonparametric treatment of $\eta^p$

As we stated previously, under assumptions such as symmetry of interaction effects, the strategic index $\eta^p(Y^{-p}; X)$ can be characterized simply as an aggregate function of $Y^{-p}$ without the need for any parametrization. In a more general setting, the computational simplicity of our approach can allow us to handle a very rich and flexible parameterization of the strategic index. However, the use of a parametric approximation may shrink the identified set for $\eta^p$ in ways that could be difficult to predict (see Ponomareva and Tamer (2011) for a discussion of related issues). To avoid these issues, a fully nonparametric treatment for $\eta^p$ can be considered, and a sieves approximation (see Chen (2007)) seems appropriate since it allows us to impose shape or sign restrictions in addition to the conditional moment inequalities from Theorem 1. Due to its complexity, a complete characterization of its asymptotic properties is beyond the scope of this paper and is the current focus of ongoing work on a separate, more general paper on the subject of sieves-based nonparametric inference with conditional moment inequalities.

### 5 Monte Carlo Experiments

In this section we apply our inferential approach to a Monte Carlo design described as follows. We consider a model of imperfect competition between three firms.

#### 5.1 Demand system

We consider a model of imperfect competition with differentiated products. The system of (inverse) demand functions is of the form

$$\mathcal{P}^1 = \zeta^1 \cdot X_a - (\lambda^1 + \delta^1 \cdot X_b) \cdot Y^1 - (\beta^12 + \gamma^12 \cdot X_b) \cdot Y^2 - (\beta^13 + \gamma^13 \cdot X_b) \cdot Y^3,$$

$$\mathcal{P}^2 = \zeta^2 \cdot X_a - (\lambda^2 + \delta^2 \cdot X_b) \cdot Y^2 - (\beta^21 + \gamma^21 \cdot X_b) \cdot Y^1 - (\beta^23 + \gamma^23 \cdot X_b) \cdot Y^3,$$

$$\mathcal{P}^3 = \zeta^3 \cdot X_a - (\lambda^3 + \delta^3 \cdot X_b) \cdot Y^3 - (\beta^31 + \gamma^31 \cdot X_b) \cdot Y^1 - (\beta^32 + \gamma^32 \cdot X_b) \cdot Y^2.$$
$Y^p$ refers to the quantity produced by firm $p$ and $X_a$ and $X_b$ are demand shifters assumed to be observed by the econometrician and the firms. The rest are parameters which for the purposes of our experiment were chosen to be

$$
\zeta^1 = \zeta^2 = \zeta^3 = 2, \quad \lambda^1 = \lambda^2 = \lambda^3 = 0, \quad \delta^1 = \delta^2 = \delta^3 = 1, \\
\beta^{12} = \beta^{21} = 0, \quad \beta^{13} = \beta^{31} = \beta^{23} = \beta^{32} = 1, \\
\gamma^{12} = \gamma^{21} = \gamma^{23} = \gamma^{31} = 1, \quad \gamma^{13} = \gamma^{31} = -5.
$$

The demand shifters $X_a$ and $X_b$ are independently drawn from the following distributions,

$$X_a = \exp \{ Z_a \}, \text{ where } Z_a \sim \mathcal{N}(0, 1), \quad \text{and} \quad X_b \sim U[0, 1].$$

### 5.1.1 Strategy space

The strategy space is given by

$$\mathcal{Y}^p = \{0, 1, 2, \ldots, 10\}.$$  

Using our previous notation this means $M^p = 10$ for each $p$.

### 5.2 Cost functions

Cost functions are of the form

$$C^p(Y^p) = F^p \cdot \mathbb{1}(Y^p > 0) + \mu^p \cdot (X_{mc} + \varsigma^p) \cdot Y^p.$$  

$X_{mc}$, $F^p$ and $\varsigma^p$ are random cost shifters and $\mu^p$ is a parameter. Here $X_{mc}$ is observed by the econometrician, but both $F^p$ and $\varsigma^p$ are only privately observed by firm $p$. The marginal cost parameter $\mu^p$ was set to $\mu^p = 1/10$ for each $p$. The cost shifters $F^p$ and $\varsigma^p$ are independently drawn from the following distributions,

$$F^p \sim U[1, 2], \quad \varsigma^p \sim U[0, 1].$$  

These private shocks are independent across $p = 1, 2, 3$. The observable cost shifter $X_{mc}$ is drawn from a $U[0, 1]$ distribution, independent of $(F^p, \varsigma^p)^{p=1,2,3}$. 

---

17
5.3 Payoff functions

Firms compete in quantities produced. Given our previous description of demand and costs, firms’ profit (payoff) functions are given by

\[
\begin{align*}
\nu^1(y^1, y^2, y^3; X, \varepsilon^1) &= 2X_a - X_b \cdot (y^1)^2 - F^1 \cdot \mathbb{1} \{ y^1 > 0 \} - \frac{1}{10} \cdot (X_{mc} + \xi^{1}) \cdot y^1 - \frac{X_b \cdot y^2 + (1 - 5X_b) \cdot y^3}{\eta^2(y^2, y^3; X)} \cdot y^1, \\
\nu^2(y^2, y^1, y^3; X, \varepsilon^2) &= 2X_a - X_b \cdot (y^2)^2 - F^2 \cdot \mathbb{1} \{ y^2 > 0 \} - \frac{1}{10} \cdot (X_{mc} + \xi^{2}) \cdot y^2 - \frac{X_b \cdot y^1 + (1 + X_b) \cdot y^3}{\eta^2(y^1, y^3; X)} \cdot y^2, \tag{9} \\
\nu^3(y^3, y^1, y^2; X, \varepsilon^3) &= 2X_a - X_b \cdot (y^3)^2 - F^3 \cdot \mathbb{1} \{ y^3 > 0 \} - \frac{1}{10} \cdot (X_{mc} + \xi^{3}) \cdot y^3 - \frac{(1 - 5X_b) \cdot y^1 + (1 + X_b) \cdot y^2}{\eta^3(y^1, y^2; X)} \cdot y^3.
\end{align*}
\]

The game is played simultaneously and, in accordance with the assumptions in previous sections, the outcome is a Bayesian-Nash equilibrium induced by degenerate beliefs.

5.3.1 Strategic indices, substitutability and complementarity

It is easy to verify that the payoff functions described in (9) satisfy our Assumption 1. The strategic indices are given by

\[
\begin{align*}
\eta^1(y^2, y^3; X) &= X_b \cdot y^2 + (1 - 5X_b) \cdot y^3, \\
\eta^2(y^1, y^3; X) &= X_b \cdot y^1 + (1 + X_b) \cdot y^3, \\
\eta^3(y^1, y^2; X) &= (1 - 5X_b) \cdot y^1 + (1 + X_b) \cdot y^2. \tag{10}
\end{align*}
\]

The patterns of substitutability/complementarity that emerge from (10) can be summarized as follows,

(i) \((Y^1, Y^3)\) are strategic complements whenever \(X_b > \frac{1}{5}\), which occurs with probability 80% since \(X_b \sim U[0, 1]\).

(ii) \((Y^1, Y^2)\) and \((Y^2, Y^3)\) are always strategic substitutes.

5.4 Equilibrium selection rule

The nature of the strategy space (discrete and bounded at \(M^p = 10\)) induces the existence of multiple equilibria. Whenever multiple equilibria exist we impose the following equilibrium selection rule.

(i) An equilibrium selection device \(\xi\) is randomly drawn from a \([0, 1]\) distribution. This draw is independent from all payoff shifters in the model.

(ii) If \(\xi < \frac{1}{5}\), the equilibrium is selected completely at random from the existing equilibria.

(iii) If \(\frac{1}{5} \leq \xi < \frac{2}{5}\), the equilibrium selected is the one that yields the largest combined profits for firms 1 and 3.

(iv) If \(\xi \geq \frac{2}{5}\), the equilibrium selected is the one that yields the largest profits for firm 2.
As we remarked in the paragraph immediately following Theorem 1, the identification power of our procedure requires a nondegenerate equilibrium selected mechanism. The one described above is a particular instance of a nondegenerate selection mechanism.

In summary, the researcher observes \((Y_1, Y_2, Y_3)\). In addition, the collection of covariates observed by the researcher is

\[ X \equiv (X_a, X_b, X_{mc}), \]

and the unobserved payoff shifters are

\[ \varepsilon^p \equiv (F^p, \varsigma^p), \quad p = 1, 2, 3. \]

Also in adherence to the assumptions of our model, \(\varepsilon^p\) is only privately observed by \(p\).

### 5.5 Summary of equilibrium features of the experimental data

The existence of multiple equilibria was prevalent within our designs. Table 1 summarizes the number of equilibria observed in 500,000 simulations.

<table>
<thead>
<tr>
<th>% of games with one BNE</th>
<th>% of games with two BNE</th>
<th>% of games with three BNE</th>
<th>% of games with four or more BNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.7%</td>
<td>63.5%</td>
<td>7.7%</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

Results from 500,000 simulations.

Our design also allows us to corroborate the negative association, conditional on \(X\), between \(Y^p\) and the strategic index \(\eta^p(Y^{\neg p}; X)\) which is at the center of our inference and identification results. Table 2 illustrates this important feature.

<table>
<thead>
<tr>
<th>Firm 1</th>
<th>Firm 2</th>
<th>Firm 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.716</td>
<td>-0.384</td>
<td>-0.743</td>
</tr>
</tbody>
</table>

Results from 500,000 simulations.

As a quick gauge of the identification power of our main result, we can compare the correlations in Table 2 with the ones that would result from a function other than the strategic index \(\eta^p\). For example, take firm 1 and consider the (incorrect) index \(g^1(Y^{-1}; X) = Y^2 + Y^3\). For this function, the average value of \(\text{Corr} (Y^1, g^1(Y^{-1}; X) | X)\) is 0.32, which would lead us to reject \(g^1\) as the strategic index \(\eta^1\).

### 5.6 Inference in Monte Carlo experiments

Next we describe the econometric analysis of our Monte Carlo data.
5.6.1 Parametric family for strategic interaction indices

We focus on a parameterization that is compatible with our true strategic indices, described in (10). Specifically we focus on a parametric family of the following form,

\[ \eta^p(y^q, y^r; X) = (\beta^{pq} + \gamma^{pq} \cdot X_b) \cdot y^q + (\beta^{pr} + \gamma^{pr} \cdot X_b) \cdot y^r, \quad (p, q, r) \in \{1, 2, 3\}. \]

Let us group the parameters of the strategic indices as

\[ \theta^p \equiv (\beta^{pq}, \gamma^{pq}) \quad \text{and} \quad \theta \equiv (\theta^1, \theta^2, \theta^3). \]

Their true values are given by

\[ \theta^1_0 = (0, 1, 1, -5), \quad \theta^2_0 = (0, 1, 1, 1), \quad \theta^3_0 = (1, -5, 1, 1), \quad \theta_0 \equiv (\theta^1_0, \theta^2_0, \theta^3_0). \]  

Parameter space

Some of our inference experiments (and choice of tuning parameters) involved searches over a parameter space \( \Theta \). The parameter space used throughout was the Cartesian product given by the interval \([0, 4]\) for each \(\beta^{pq}\), the interval \([-10, 0]\) for \(\gamma^{13}\) and \(\gamma^{31}\), and the interval \([0, 4]\) for every other \(\gamma^{pq}\).

5.6.2 Kernels and bandwidths

In order to study their finite-sample properties, we employed kernels and bandwidths identical to those that we will use in our empirical application in Section 6. These are described in exact detail in Appendix B.7. As we describe there\(^7\), the kernel employed is of order \(M = 18\), which is of higher order than needed given that the experimental data includes three continuously distributed observable covariates \(X \equiv (X_a, X_b, X_{mc})\). We opted to utilize this specific kernel because it is the one employed in our empirical application in Section 6, and we want to study its finite-sample properties in the context of our Monte Carlo experiments. Our target inference range included the entire data.

5.6.3 Inference exercises

Next we describe the specific features of our confidence sets (CS) that were studied in our experiments.

(A) Inclusion of the true parameter value \(\theta_0\) in our CS

Our first exercise is to evaluate the ability of our procedure to include the true parameter value \(\theta_0\) in the CS. Given that our Monte Carlo experiments are designed to satisfy the assumptions underlying our econometric procedure, we know that asymptotically \(\theta_0\) will be included in our CS with probability at least \(1 - \alpha\), where the latter represents our target coverage probability. Our first goal is to determine the ability of our procedure to accomplish this in finite samples. To this end we generated 1,000 simulations of the Monte Carlo design described above for sample sizes \(n = 500, n = 1,000, n = 1,500\) and \(n = 2,000\) (1,000 simulations in each case) and computed the proportion of samples for which \(\theta_0\) was accepted into our CS. The results are shown in Table 3.

---

\(^7\)Some of the tuning parameters described in Appendix B.7 involve searches over the parameter space \(\Theta\), which was described above for our Monte Carlo experiments.
Exclusion of false parameter values from our CS

Our next exercise is aimed at studying the power of our econometric approach of rejecting false conjectures about strategic interaction. Suppose we maintain the (incorrect) assumption that the goods produced by the three firms are always strategic substitutes. Under this maintained assumption we test the power of our approach of rejecting the following three conjectures,

(i) **Symmetric and constant strategic interaction effects.** Our first false parameter value is aimed at testing the conjecture that strategic effects are symmetric and constant across markets and across players. Under the maintained assumption of strategic substitutes, this is equivalent to testing whether the parameter value belongs in our CS,

\[ \theta^p_a \equiv (\beta^p_{pq} a, \gamma^p_{pq} a, \beta^p_{pr} a, \gamma^p_{pr} a) = (1, 0, 1, 0), \quad \text{with} \quad \theta_a \equiv (\theta^1_a, \theta^2_a, \theta^3_a). \]

(ii) **Symmetric interaction effects, equal to zero if \( X_b = 0 \).** Our next false tests the conjecture that strategic effects are once again symmetric across players, but they depend on \( X_b \) in a way such that if \( X_b = 0 \) then there is no strategic effect. Under the maintained assumption of strategic substitutes, this is equivalent to testing whether the parameter value belongs in our CS,

\[ \theta^p_b \equiv (\beta^p_{pq} b, \gamma^p_{pq} b, \beta^p_{pr} b, \gamma^p_{pr} b) = (0, 1, 0, 1), \quad \text{with} \quad \theta_b \equiv (\theta^1_b, \theta^2_b, \theta^3_b). \]

(iii) **Symmetric interaction effects, with \( \beta^pq = \gamma^pq \).** As a third case we aim to test the conjecture that strategic effects are symmetric across players and they satisfy \( \beta_{pq} = \gamma_{pq} \) for each \( p \) and \( q \). Under the maintained assumption of strategic substitutes, this amounts to testing whether the parameter value belongs in our CS,

\[ \theta^p_c \equiv (\beta^p_{pq} c, \gamma^p_{pq} c, \beta^p_{pr} c, \gamma^p_{pr} c) = (1, 1, 1, 1), \quad \text{with} \quad \theta_c \equiv (\theta^1_c, \theta^2_c, \theta^3_c). \]

Asymptotically each one of these parameter values will be excluded from our CS with probability one. Our goal here is to study the ability of our econometric procedure to reject these false conjectures by excluding the corresponding parameter values from our CS in finite samples. To this end we generated 1,000 simulations of the Monte Carlo design described above for sample sizes \( n = 500, n = 1,000, n = 1,500 \) and \( n = 2,000 \) (1,000 simulations in each case) and computed the proportion of samples for which \( \theta_a, \theta_b \) and \( \theta_c \) were rejected from our CS. The results are summarized in Table 4.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Target coverage: 95% ((c_{1-\alpha} = 1.645))</th>
<th>Target coverage: 99% ((c_{1-\alpha} = 2.33))</th>
<th>(95^{th}) percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 500)</td>
<td>99.9%</td>
<td>100%</td>
<td>1.074</td>
<td>1.344</td>
</tr>
<tr>
<td>(n = 1,000)</td>
<td>99%</td>
<td>100%</td>
<td>1.291</td>
<td>2.163</td>
</tr>
<tr>
<td>(n = 1,500)</td>
<td>98%</td>
<td>100%</td>
<td>1.485</td>
<td>2.162</td>
</tr>
<tr>
<td>(n = 2,000)</td>
<td>94.9%</td>
<td>99.9%</td>
<td>1.647</td>
<td>2.352</td>
</tr>
</tbody>
</table>
Table 4: Observed frequency with which false parameter values were EXCLUDED from our CS

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Target coverage: 95% ( (c_{1-\alpha} = 1.645) )</th>
<th>Target coverage: 99% ( (c_{1-\alpha} = 2.33) )</th>
<th>95th percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>54.3%</td>
<td>45.2%</td>
<td>7.324</td>
<td>11.811</td>
</tr>
<tr>
<td>1000</td>
<td>66.3%</td>
<td>57.7%</td>
<td>16.537</td>
<td>15.598</td>
</tr>
<tr>
<td>1500</td>
<td>77.6%</td>
<td>70.0%</td>
<td>13.977</td>
<td>21.931</td>
</tr>
<tr>
<td>2000</td>
<td>81.4%</td>
<td>76.4%</td>
<td>17.410</td>
<td>22.769</td>
</tr>
</tbody>
</table>

Observed frequency with which \( \theta_b \) was EXCLUDED from our CS

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Target coverage: 95% ( (c_{1-\alpha} = 1.645) )</th>
<th>Target coverage: 99% ( (c_{1-\alpha} = 2.33) )</th>
<th>95th percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>57.5%</td>
<td>48.3%</td>
<td>7.746</td>
<td>11.498</td>
</tr>
<tr>
<td>1000</td>
<td>69.2%</td>
<td>61.1%</td>
<td>11.652</td>
<td>15.088</td>
</tr>
<tr>
<td>1500</td>
<td>79.7%</td>
<td>73.3%</td>
<td>15.113</td>
<td>19.838</td>
</tr>
<tr>
<td>2000</td>
<td>84.6%</td>
<td>79.1%</td>
<td>18.199</td>
<td>22.386</td>
</tr>
</tbody>
</table>

Observed frequency with which \( \theta_c \) was EXCLUDED from our CS

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Target coverage: 95% ( (c_{1-\alpha} = 1.645) )</th>
<th>Target coverage: 99% ( (c_{1-\alpha} = 2.33) )</th>
<th>95th percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>54.6%</td>
<td>46.2%</td>
<td>7.488</td>
<td>9.431</td>
</tr>
<tr>
<td>1000</td>
<td>67.4%</td>
<td>59.7%</td>
<td>11.056</td>
<td>15.718</td>
</tr>
<tr>
<td>1500</td>
<td>78.8%</td>
<td>71.7%</td>
<td>14.653</td>
<td>20.546</td>
</tr>
<tr>
<td>2000</td>
<td>83.0%</td>
<td>77.3%</td>
<td>17.774</td>
<td>20.966</td>
</tr>
</tbody>
</table>

(C) Exploring other features and conjectures of our CS

Suppose we maintain the following (correct) conjectures about \( \theta_0 \).

(i) \( \beta_{pq} = \beta_{qp} \) and \( \gamma_{pq} = \gamma_{qp} \) for each \( p, q \). We impose this restriction on the parameter space \( \Theta \).

(ii) \( (Y^1, Y^2) \) and \( (Y^2, Y^3) \) are always strategic substitutes. This can be captured by imposing the following restrictions on \( \Theta \),

\[
\beta^{12} \geq 0, \gamma^{12} \geq 0, \beta^{23} \geq 0, \gamma^{23} \geq 0, \quad \text{with} \quad |\beta^{12}| + |\gamma^{12}| > 0, |\beta^{23}| + |\gamma^{23}| > 0.
\]

(iii) \( (Y^1, Y^3) \) are strategic substitutes for “small” values of \( X_b \) but they are strategic complements for “large” values of \( X_b \). However, we do not know the corresponding threshold for \( X_b \) that induces complementarity. This can be captured by imposing the following restrictions on \( \Theta \),

\[
\beta^{13} > 0, \quad \gamma^{13} < 0.
\]

Given these maintained assumptions, suppose we want to study the following two strategic interaction features:

- The relative effects of firms 1 and 3 on firm 2.
- How large does \( X_b \) have to be to induce strategic complementarities between firms 1 and 3?
Both features will be analyzed by estimating a CS after imposing the restrictions on $\Theta$ described above. Our construction involved a grid search over two million points in $\Theta$. The results described below represent the results for a randomly drawn sample generated according to our previous description. Our target coverage probability was 95%.

Relative effects of firms 1 and 3 on firm 2

We have maintained that both $Y^1$ and $Y^3$ are always strategic substitutes for $Y^2$. We want to explore which one of the two rivals of firm 2 has a larger strategic effect on firm 2’s payoff. In this sense, we want to study which one of firm 2’s rivals is a closer competitor to firm 2. Given our parameterization of the strategic indices, this question involves a comparison between $\beta^{21} + \gamma^{21} \cdot X_b$ and $\beta^{23} + \gamma^{23} \cdot X_b$. In order to aggregate $X_b$, we will compare

$$E_{X_b} [\beta^{21} + \gamma^{21} \cdot X_b] = \beta^{21} + \gamma^{21} \cdot \frac{1}{2}$$

and

$$E_{X_b} [\beta^{23} + \gamma^{23} \cdot X_b] = \beta^{23} + \gamma^{23} \cdot \frac{1}{2}.$$

Note that the true values of these quantities are

$$E_{X_b} [\beta^{21} + \gamma^{21} \cdot X_b] = \frac{1}{2} \quad \text{and} \quad E_{X_b} [\beta^{23} + \gamma^{23} \cdot X_b] = \frac{3}{2}$$

and therefore firm 3 is the true closest competitor to firm 2. Figure 1 shows that our CS overwhelmingly reflects this key property. It also illustrates how our CS contains the true value of these parameters.

Figure 1: Relative effects of firms 1 and 3 on firm 2. 95% joint CS for $E_{X_b} [\beta^{21} + \gamma^{21} \cdot X_b]$ and $E_{X_b} [\beta^{23} + \gamma^{23} \cdot X_b]$.

\[8\] Recall that $X_b \sim U [0, 1]$. 

23
Strategic complementarities between firms 1 and 3

We have maintained the conjecture that \( Y_1 \) and \( Y_3 \) are strategic complements for “large” values of \( X_b \). Let \( x^* \) denote the value such that \( Y_1 \) and \( Y_3 \) are strategic complements whenever \( X_b > x^* \) and substitutes otherwise. This threshold is given by

\[
x^* = -\frac{\beta_1^{13}}{\gamma_{13}}.
\]

Note that \( x^* = 0.20 \) given the parameters of our Monte Carlo design. We can use our 95% CS to construct a corresponding confidence interval (CI) for \( x^* \). This is given by,

\[
95\% \text{ CI for } x^* : [0.003, 0.384].
\]

The true value \( x^* = 0.2 \) is practically the midpoint of our CI.

5.7 Performance under violations to our assumptions

Our last goal in this section is to investigate the extent to which the properties of our CS break down when some of our key assumptions are violated. Specifically we want to study what happens when two key conditions are violated:

(i) Violations to Assumption 3 introducing correlation in players’ private shocks.

(ii) Violations to Assumption 1. Specifically, to the assumption that the strategic index \( \eta^p \) can be expressed as a function solely of observable payoff shifters \( X \).

To modify our design in a way that violates both assumptions, the demand system is now given by

\[
\begin{align*}
P_1 &= \zeta_1 \cdot X_a - (\lambda_1 + \delta_1 \cdot X_b) \cdot Y_1 - (\beta_1^{12} + \gamma_1^{12} \cdot X_b + \rho \cdot \zeta) \cdot Y^2 - (\beta_1^{13} + \gamma_1^{13} \cdot X_b + \rho \cdot \zeta) \cdot Y^3, \\
P_2 &= \zeta_2 \cdot X_a - (\lambda_2 + \delta_2 \cdot X_b) \cdot Y_2 - (\beta_2^{11} + \gamma_2^{11} \cdot X_b + \rho \cdot \zeta) \cdot Y^1 - (\beta_2^{23} + \gamma_2^{23} \cdot X_b + \rho \cdot \zeta) \cdot Y^3, \\
P_3 &= \zeta_3 \cdot X_a - (\lambda_3 + \delta_3 \cdot X_b) \cdot Y_3 - (\beta_3^{11} + \gamma_3^{11} \cdot X_b + \rho \cdot \zeta) \cdot Y^1 - (\beta_3^{23} + \gamma_3^{23} \cdot X_b + \rho \cdot \zeta) \cdot Y^2.
\end{align*}
\]

Where \( \zeta \) is unobserved by the econometrician but perfectly observed by all three firms and \( \rho \) is a parameter that measures the importance of \( \zeta \) as a payoff shifter. Since the latter is a common component of players’ private shocks, \( \rho \) provides also a measure of the correlation between players’ private shocks. We generate \( \zeta \sim U [0, 1] \), independent of all other covariates in the model. With this to the demand system, it is no longer possible to express the strategic index \( \eta^p \) as a function only of observables. There is also correlation in payoff shocks unobserved by the econometrician, violating Assumption 3. As a result of these violations, the main result in Theorem 1 is no longer valid. For finite samples, our a-priori conjectures are the following,

(a) The asymptotic predictions of our approach should retain some of their validity for small values of \( \rho \) (i.e., small correlation between private shocks).

(b) For increasingly larger values of \( \rho \) (i.e., larger correlation between players’ private shocks), our approach should provide evidence that the model is incorrect, leading to a smaller (and eventually empty) CS. In particular, for larger values of \( \rho \) our CS should exclude not only incorrect values of \( \theta \) but also its true value, \( \theta_0 \).
To investigate the validity of our conjectures we repeat two of the exercises done previously in Tables 3 and 4. We generated 1,000 samples of size \( n = 2,000 \) and we tried different values of \( \rho \). For each one we computed the frequency with which our CS included \( \theta_0 \) and excluded \( \theta_b \) (as defined above). According to our conjectures, our approach should still lead us to reject the fake value \( \theta_b \) and, for increasingly larger values of \( \rho \), it should also lead us to reject \( \theta_0 \). Our results are summarized in Tables 5 and 6 and are directly comparable to those in Table 3 and in the second panel in Table 4.

Table 5: Observed frequency with which \( \theta_0 \) was INCLUDED in our CS when Assumptions 1 and 3 are violated

<table>
<thead>
<tr>
<th>Value of ( \rho )</th>
<th>Target coverage: 95% ( (c_{1-a} = 1.645) )</th>
<th>Target coverage: 99% ( (c_{1-a} = 2.33) )</th>
<th>95% percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = 0.25 )</td>
<td>95.2%</td>
<td>100%</td>
<td>1.627</td>
<td>2.216</td>
</tr>
<tr>
<td>( \rho = 0.50 )</td>
<td>83.3%</td>
<td>96.2%</td>
<td>2.142</td>
<td>3.398</td>
</tr>
<tr>
<td>( \rho = 1 )</td>
<td>48.3%</td>
<td>66.3%</td>
<td>4.686</td>
<td>7.638</td>
</tr>
<tr>
<td>( \rho = 2 )</td>
<td>30%</td>
<td>40.9%</td>
<td>8.082</td>
<td>13.965</td>
</tr>
</tbody>
</table>

1,000 simulated samples of size \( n = 2,000 \).

Table 6: Observed frequency with which \( \theta_b \) was EXCLUDED from our CS when Assumptions 1 and 3 are violated

<table>
<thead>
<tr>
<th>Value of ( \rho )</th>
<th>Target coverage: 95% ( (c_{1-a} = 1.645) )</th>
<th>Target coverage: 99% ( (c_{1-a} = 2.33) )</th>
<th>95% percentile observed value of test-statistic</th>
<th>Maximum observed value of test-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = 0.25 )</td>
<td>77.1%</td>
<td>70.4%</td>
<td>14.693</td>
<td>22.551</td>
</tr>
<tr>
<td>( \rho = 0.50 )</td>
<td>71.5%</td>
<td>64.3%</td>
<td>13.292</td>
<td>24.154</td>
</tr>
<tr>
<td>( \rho = 1 )</td>
<td>79.6%</td>
<td>74.5%</td>
<td>15.951</td>
<td>21.753</td>
</tr>
<tr>
<td>( \rho = 2 )</td>
<td>85.4%</td>
<td>81.4%</td>
<td>20.797</td>
<td>29.267</td>
</tr>
</tbody>
</table>

1,000 simulated samples of size \( n = 2,000 \).

The results in Tables 5 and 6 are in line with our previous conjectures. Firstly, the ability of our approach to reject the false value \( \theta_b \) is not affected by misspecification (if anything, the propensity to reject \( \theta_b \) is increased by the misspecification). Regarding the inclusion of \( \theta_0 \) in our CS, when our model is only slightly misspecified and \( \rho \) is relatively small our results remain very much in line with the asymptotic predictions. As the model becomes increasingly misspecified and \( \rho \) is larger, our procedure rejects \( \theta_0 \) because Theorem 1 is no longer true and the model as a whole is being rejected by our econometric approach.

6 Application: Entry in the U.S drug store industry

Next we include an illustration of our results in the analysis of empirical data. One of the most important econometric applications of games has been the study of entry decisions by competing firms. Our model allows us to approach this problem by combining the usual extensive-margin enter/not enter dimension with an intensive-margin decision regarding the intensity of entry. In our application, this intensive margin is captured by the number of stores that a chain-store decides to open in a market. The key advantage of taking the intensive margin into account is that it will give us a structural interpretation of the strategic index in terms of an underlying model of supply and demand.
(see Section 2.3). This stands in contrast to the “reduced form” profit function that dominates applied work on the binary entry margin. As we show below, the intensive margin provides new insights into nature of competition in the market we study. It is important to note that our assumptions are compatible with the existence of fixed costs of entry and thus our model strictly nests the binary entry case (see Section 2.3).

Our application focuses on the U.S retail drug store industry, which we study because of three different considerations. First, it is an industry with three clearly identifiable main competitors: Walgreen’s, CVS and Rite Aid. According to IBISWorld, their market shares in 2011 were approximately 31%, 26% and 12% respectively\(^9\).

Second, there has been a recent discussion among industry watchers of a takeover of Rite-Aid by one of its competitors. This is a natural policy application for us since our approach can help us identify, for example, which is the closest competitor to Rite-Aid. Third, we believe it is a case of an industry without an obvious, compelling demand side unobservable at the market level (i.e., an unexplained taste for health) that cannot be conditioned out with observables (such as the number of doctors in the market).

Naturally, entry takes place at different points in time but these dates are largely unobserved in our data set. Our justification for modeling this as a static game is the commonly made assumption that the choices observed are the realization of a long-run equilibrium whereby firms pre-committed to their strategies before observing the strategies of others. According to this view, the fact that entry decisions take place in different points in time is incidental.

Throughout our exercise we identify these three players as:

\[
\text{player 1: CVS, player 2: Rite Aid, player 3: Walgreens.}
\]

\(p\) denotes generically any one of the three players in the model, and \(q, r\) denotes the two opponents of \(p\). Let \(Y^p\) denote the number of stores opened by \(p\) in a market. Please note that we have abstracted away the competition effect these firms may face from supermarket and big-box stores (see Ellickson and Grieco (2013)), which is likely to be significant in items such as beauty products, personal care items and over the counter medications, but less so in prescription medications and walk-in health services, although this may change in the future as Walmart and other supermarkets are added to increasingly more preferred pharmacy networks and expand their clinic services. We do this for simplicity of our empirical illustration and because we wanted to focus on the three closest competitors within the drugstore industry.

### 6.1 Data overview

#### 6.1.1 Units of observation

The decision variable \(Y^p_i\) denotes the total number of stores by \(p\) in market \(i\) in the year 2011. We define a market as a CBSA (core based statistical area) in the continental United States. Metropolitan\(^10\) CBSAs were split into the divisions determined by Office of Budget and Management and each division was considered a market. We exclude CBSAs with more than 5 million people because such large markets will likely consist of smaller sub-markets. Our final sample consists of \(N = 954\) observations.

---


*\(^10\)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 commuting to work) with the urban core. Metropolitan CBSAs are those with a population of 50,000 or more. Under certain conditions, metropolitan CBSAs with 2.5 million people or more are split into divisions.*
6.1.2 Choices and outcomes observed in the data

Table 7 summarizes some descriptive features of choices observed. It highlights the richness of the action space in this application. Table 8 shows the correlations observed across $Y^1$, $Y^2$ and $Y^3$. As we see there, a persistently positive association was observed across markets in the number of stores opened by each competitor. What is remarkable is that this pattern of positive correlation remain the same order of magnitude even after we condition on market observables such as market size, etc (we describe the market covariates in further depth in the next subsection).

Table 7: Summary statistics for $Y^p$

<table>
<thead>
<tr>
<th></th>
<th>$Y^1$</th>
<th>$Y^2$</th>
<th>$Y^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>7,004</td>
<td>4,318</td>
<td>7,283</td>
</tr>
<tr>
<td>Mean</td>
<td>7.34</td>
<td>4.52</td>
<td>7.63</td>
</tr>
<tr>
<td>Stdev</td>
<td>21.95</td>
<td>15.57</td>
<td>23.88</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>75th percentile</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>90th percentile</td>
<td>16</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>95th percentile</td>
<td>39</td>
<td>21</td>
<td>41</td>
</tr>
<tr>
<td>99th percentile</td>
<td>112</td>
<td>71</td>
<td>106</td>
</tr>
</tbody>
</table>

Table 8: Correlations observed for $Y^1$, $Y^2$ and $Y^3$

<table>
<thead>
<tr>
<th></th>
<th>$Y^1$</th>
<th>$Y^2$</th>
<th>$Y^3$</th>
<th>$Y^1 + Y^2$</th>
<th>$Y^1 + Y^3$</th>
<th>$Y^2 + Y^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y^1$</td>
<td></td>
<td>0.70</td>
<td>0.79</td>
<td>0.86</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$Y^2$</td>
<td>0.70</td>
<td></td>
<td>0.49</td>
<td>–</td>
<td>0.62</td>
<td>–</td>
</tr>
<tr>
<td>$Y^3$</td>
<td>0.79</td>
<td>0.49</td>
<td></td>
<td>–</td>
<td>–</td>
<td>0.72</td>
</tr>
</tbody>
</table>

By their nature, the drugstores of each of these competitors provide the same type of services and can be rightly deemed, in general, as demand substitutes of each other. Given this observation and recalling the underlying Cournot model discussed in Example 2.3, basic economic theory would predict that, all else equal, more aggressive entry by a competitor affects would reduce a firm’s marginal benefit to entry, leading us ex-ante to consider entry decisions as strategic substitutes. Strategic substitution is assumed numerous empirical applications of entry games (e.g., Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Berry (1992), Davis (2006)). Even though strategic substitutability is justified as the prediction of economic theory in our setting, the correlation pattern in Table 8 seems to fly in the face of it. This is especially true if we believe that there is no obvious, compelling demand side unobservable at the market level (i.e., an unexplained taste for medical drugs). Our framework can help us explore whether a model of strategic substitutes can produce this pattern of positive correlation in entry behavior.

Ignoring the intensive-margin dimension of entry and focusing only on the binary choice decision of entry immediately obscures key features of the data. For example as Table 9 shows, it wipes out much of the positive association observed in the data.

Table 9: Correlations if game is reduced to binary choice

<table>
<thead>
<tr>
<th></th>
<th>${Y^1 &gt; 0}$</th>
<th>${Y^2 &gt; 0}$</th>
<th>${Y^3 &gt; 0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${Y^1 &gt; 0}$</td>
<td>–</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>${Y^2 &gt; 0}$</td>
<td>0.23</td>
<td>–</td>
<td>0.04</td>
</tr>
<tr>
<td>${Y^3 &gt; 0}$</td>
<td>0.07</td>
<td>0.04</td>
<td>–</td>
</tr>
</tbody>
</table>
By eliminating much of the positive association observed in the intensive margin, reconciling the data with an underlying game of strategic substitutes should be easier in a binary choice representation of the game compared to one that explicitly considers the intensive margin decisions. A consequence of this would be that the inferential results for $\eta^p$ in the latter case would be more precise. Our results will confirm this.

6.1.3 Covariates included in $X$

Markets are defined as CBSAs with less than 5 million people. We included in our analysis the following market and player characteristics,

$$\text{POP} = \text{population, } \text{INC} = \text{average income per household, } \text{DENS} = \text{population density,}$$

$$\text{AGE} = \text{median age in the population, } \text{BUS} = \text{total number of business establishments,}$$

$$DIST^p = \text{distance to the nearest distribution center of } p.$$  

And we used

$$X = (\text{POP, INC, DENS, AGE, BUS, DIST}^1, DIST^2, DIST^3).$$

Population density was computed as the ratio of population/land area. $X$ was treated as jointly continuously distributed.

Most of these covariates are fairly standard in empirical work. We do note that our inclusion of the number of business establishments (which we could empirically refine to be the number of retail establishments) is designed to control for supply side unobservables in a market. If it is just costly to locate a store in a market (because of say zoning restrictions), then this should affect the entry of stores in all industries, not just pharmacies. Table 10 presents summary statistics for our covariates and makes a comparison across different markets depending on the number of total stores. The table highlights the importance in particular of market size and density, as well as distance to distribution centers, as determinants of entry. To the extent that the covariates included in $X$ may fail to fully control for unobservable market-level (demand or entry-cost) shocks and therefore leave some correlation in firms’ unobserved payoff shocks, our results in Section 5 suggest that as long as the remaining correlation is relatively small, the properties of our inferential procedure can remain approximately valid in finite samples.

<table>
<thead>
<tr>
<th>Table 10: A statistical summary of covariates and market structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Median value</strong></td>
</tr>
<tr>
<td><strong>in markets with zero stores</strong></td>
</tr>
<tr>
<td>POP</td>
</tr>
<tr>
<td>INC</td>
</tr>
<tr>
<td>DENS</td>
</tr>
<tr>
<td>AGE</td>
</tr>
<tr>
<td>BUS</td>
</tr>
<tr>
<td>DIST$^1$</td>
</tr>
<tr>
<td>DIST$^2$</td>
</tr>
<tr>
<td>DIST$^3$</td>
</tr>
</tbody>
</table>
6.2 Specifications for the strategic index $\eta^p$

We refer generically to the three players as $p, q, r$ and we consider specifications for the index of the form

$$\eta^p(y^p; X|\theta^p) = (X'\theta^{p,q}) \cdot y^q + (X'\theta^{p,r}) \cdot y^r,$$

As we discussed above, we will maintain that actions are strategic substitutes. To this end we will choose the $\Theta$ such that strategic substitutability is ensured. That is,

$$X'_i \theta^{p,q} \geq 0, \quad X'_i \theta^{p,r} \geq 0 \quad \forall i = 1, \ldots, n \quad \forall \theta \in \Theta.$$

We want to focus on simple specifications for the indices $X'_i \theta^{p,q}$ and $X'_i \theta^{p,r}$. Since $\theta$ can only be (partially) identified up to scale and location normalizations, these are also introduced in the parameter space in ways that will be described below.

**Specification 1.** Symmetry of opponents’ strategic interaction effects

First we study the special case where each $p$ weighs the actions of his two opponents equally (a maintained, key assumption in De Paula and Tang (2012)) in every market. Given our assumptions this is observationally equivalent to a strategic index of the form

$$\eta^p(y^p; X|\theta^p) = \theta^p \cdot (y^q + y^r), \quad \text{where } \theta^p = 1.$$

In this case our inferential problem simply reduces to a specification test where we evaluate whether

$$E[\mathbb{1}\{Y^p \leq y^p\} \cdot (Y^q + Y^r) | X = x] \geq E[\mathbb{1}\{Y^p \leq y^p\} | X = x] \cdot E[(Y^q + Y^r) | X = x]$$

for almost every $(x, y^p)$ in our inferential range (to be described below).

**Specification 2.** Constant, possibly asymmetric relative strategic interaction effects

Next we focus on the case where $p$ may assign different weights to each opponent, but the relative effects remain constant across all markets. Letting $\theta^p = (\theta^{p,q}, \theta^{p,r})$, the strategic index is now of the form

$$\eta^p(y^p; X|\theta^p) = \theta^{p,q} \cdot y^q + \theta^{p,r} \cdot y^r, \quad \text{where } \theta \geq 0 \forall \theta \in \Theta.$$

We normalize $\Theta$ so that $\|\theta^p\| = 1$ for each $p$ since our identified set is closed under nonnegative re-scaling of $\theta^p$ (if $\theta$ satisfies (5), then so will $c \cdot \theta$ for any $c \geq 0$). This specification is of particular interest because strategic interaction effects have been typically modeled through constant coefficients in existing work that uses “reduced form” profit functions. (e.g. Berry (1992), Tamer (2003) and many others).

**Specification 3.** A more flexible parametrization

Here we allow for asymmetry and for covariate-dependent relative interaction effects. In our specification we express $\eta^p$ solely as a function of market size ($POP$) and its distance to the nearest distribution center of each player ($DIST^1, DIST^2, DIST^3$). We wish to explore two conjectures through our parametrization:
(i) The difference in distance to the market \((DIST^p - DIST^q)\) is a determinant of the strategic interaction effect of \(q\) on \(p\). The basis for this effect is that if firm \(q\)'s distribution center is located much closer than \(p\)'s, then this will give \(q\) a cost side advantage relative to \(p\) in the market and thus make competition more intense with firm \(q\)'s entry into the market.

(ii) Strategic interaction effects change with market size. One strand of the entry literature has modeled firm profits using “per capita” variable profits (see e.g., Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a)) , which would imply that the sensitivity of a firm’s profits to another firm’s entry is increasing with market size all else equal. However one can also imagine that larger markets offer more “room” for entry not just because there exist more people but also because opportunities for market expansion relative to business stealing are larger, which would decrease the sensitivity of profit to a rival firm’s entry in larger markets.

To explore both conjectures simultaneously we use the following parametrization of \(\eta^p\). Denote \(\theta^p = (\theta^p_1, \theta^p_2, \theta^p_3, \theta^p_4)'\) and \(D^{p,q} \equiv DIST^p - DIST^q\) for every \(p \neq q\). Define

\[
\phi^{p,q}(X|\theta^p) = 
\left(\theta^p_1 + \frac{\theta^p_2}{\text{POP}} + \theta^p_3 \cdot (D^{p,q} - 200) \cdot \mathbb{I}\{D^{p,q} \geq 200\} + \theta^p_4 \cdot \frac{(D^{p,q} - 200) \cdot \mathbb{I}\{D^{p,q} \geq 200\}}{\text{POP}} \right), \quad (14)
\]

Population is measured in units of 500K inhabitants in (14). The strategic index for \(p\) is specified as

\[
\eta^p(y^{-p}; X|\theta^p) = \phi^{p,q}(X|\theta^p) \cdot y^q + \phi^{p,r}(X|\theta^p) \cdot y^r. \quad (15)
\]

Strategic substitutability is imposed by forcing the parameter space \(\Theta\) to satisfy \(\phi^{p,q}(X_i|\theta^p) \geq 0\) for each \(p, q\) and every market \(i = 1, \ldots, n\). The individual signs of each coefficient were otherwise unrestricted. For the same reason given above we normalize \(\|\theta^p\| = 1\) for \(p = 1, 2, 3\) in our parameter space.

### 6.3 Results

Our target coverage probability is 95\% throughout. Our parameter space \(\Theta\) consisted of 1 million grid points with the scale-normalization described above. An empty confidence set (CS) amounts to a rejection of the specification in question. The kernels and bandwidths used are described in detail in Appendix B.7. The kernel employed was bias-reducing of order 18, similar to the one used in Aradillas-López, Gandhi, and Quint (2013). Our bandwidths were of the form \(h_n = c \cdot \hat{\sigma}(X) \cdot n^{-\alpha_h}\) (we used individual bandwidths for each \(X\), each proportional to \(\hat{\sigma}(X)\)), \(b_n = c_b \cdot \hat{\Omega} \cdot n^{-\alpha_b}\) and \(\kappa_n = c_\kappa \cdot \hat{\Omega} \cdot \log(n)^{-1}\), where \(\hat{\Omega} = \max_{\theta \in \Theta} \hat{\sigma}(\theta)\). We chose these tuning parameters proportional to \(\hat{\Omega}\) to ensure our procedure has a scale-invariant property. The choice of the constants \(c, c_b, c_\kappa, \alpha_h\) and \(\alpha_b\) are described in Appendix B.7. For our sample size \(n = 954\) the values of these tuning parameters were \(h_n \approx 0.16 \cdot \hat{\sigma}(X)\), \(b_n \approx 10^{-3}\) and \(\kappa_n \approx 10^{-7}\). The inference range used was

\[
\mathcal{X} = \left\{ x : \hat{f}_X(x) \geq \hat{f}_X^{(0.15)}, \quad \text{POP} < 5\text{ Million} \right\},
\]

where \(\hat{f}_X^{(0.15)}\) denotes the estimated 15th percentile of the density \(\hat{f}_X\). All of our main findings were robust to moderate changes in the tuning parameters used.
6.3.1 Rejection of symmetry and of constant strategic interaction effects

Symmetry in the effects of opponents’ actions on payoffs was rejected by our results. The value of our test-statistic for testing (12) was 10.44, well above the critical value (1.645) for a 95% significance level. We conclude that if strategic substitutability is maintained across all markets, at least one player must assign different weights to the actions of his opponents in a subset of markets. Our results also rejected Specification 2 which assumed constant strategic effects. The smallest value of the test-statistic across our parameter space \( \Theta \) was 8.38, leading to an empty confidence set. Rejection of constant strategic effects is a relevant empirical finding because this is the type of specification used in the vast majority of existing parametric models. By Remark 2 rejecting any specification leads us to reject the assertion that the underlying game has a unique equilibrium w.p.1. In particular we reject the notion that there is no strategic interaction effect between the firms.

6.3.2 Results for Specification 3

Our third specification produced a nonempty CS. Our first finding was a rejection of the assertion that \( \theta^p = \theta^q \) for each \( p \neq q \) (symmetry in parameters for all players). When we imposed this restriction we obtained an empty CS, with the smallest value of the test-statistic being 2.01. Thus there is evidence of structural differences in payoff functions across these three players. We describe the main features of the CS obtained next.

6.3.3 Evidence of asymmetric weights to opponents’ strategies

Asymmetry of opponents’ interaction effects is captured by the parameters \( \theta^p_3 \) and \( \theta^p_4 \). Symmetry would hold for \( p \) in every market only if these parameters are jointly equal to zero. Figure 2 depicts the 95% joint CS for these parameters for each of the three players. As we can see, our results showed evidence of asymmetry for player 2 (Rite Aid).

![Figure 2: Asymmetry of strategic interaction effects. 95% joint CS for \( \theta^p_3 \) and \( \theta^p_4 \)](image)
We can study the asymmetry of strategic effects for specific markets. For example, figure 3 depicts our confidence region for $\phi^{2,1}(X_i|\theta^2)$ (the effect of CVS on Rite Aid) and $\phi^{2,3}(X_i|\theta^2)$ (the effect of Walgreens on Rite Aid) corresponding to CBSA 29404 (Lake County-Kenosha County, IL-WI), where $POP = 820K$, $DIST^1 = 191$, $DIST^2 = 226$ and $DIST^3 = 21$. Our results show that, from the perspective of Rite Aid, the competition effect from Walgreens is stronger than the competition effect from CVS in that market.

We wanted to learn more about what the data revealed regarding the closeness of competition between rival firms. Since symmetry could only be rejected for Rite Aid we focused only on this firm. We say that the competition effect from CVS is stronger than that of Walgreens in market $i$ if

$$\min_{\theta^2 \in CS_n(1-\alpha)} (\phi^{2,1}(X_i|\theta^2)) > \max_{\theta^2 \in CS_n(1-\alpha)} (\phi^{2,3}(X_i|\theta^2)).$$

The opposite would be true if the inequality holds with the superscripts 1 and 3 interchanged. We found that, while the competition effect from CVS was stronger than that of Walgreens only in 9 markets, the opposite was true in 160 markets. Overall, our results provide evidence that Walgreens is a closer competitor to Rite Aid than CVS is. For policy purposes this closeness in competition could suggest that a merger between Rite Aid and Walgreens could potentially have a more significant anticompetitive effect than a merger between Rite Aid and CVS.\textsuperscript{11}

### 6.3.4 Market size and strategic interaction

One of the goals of specification 3 was to study the relationship between strategic interaction and market size. Positive signs for $\theta^p_2$ and $\theta^p_4$ would be consistent with interaction effects that decrease with the size of the market. Figure 4 depicts the 95% joint CS for these parameters for each firm. As we see there most of the values included in our CS for both coefficients are positive. Some negative values (except for $\theta^2_4$) are included, but these are relatively small in absolute value.

Let us focus on cases where relative distance is not significant (i.e, less than 200 miles) and the only determinant of strategic interaction is market size. In any such market the strategic coefficients are $\phi^{p,q}(X|\theta^p) = \theta^p_1 + \theta^p_2 \cdot \frac{1}{POP}$. Figure 5 shows how these strategic coefficients change with the size of the market. As we can see there, our results suggest that the strategic effect of opponents’ strategies does not increase with market size and in fact could be less significant in larger markets.

\textsuperscript{11}Rite Aid shares jumped sharply on March 14, 2012 following speculation from a Credit Suisse analyst about a potential merger with Walgreens (source: New York Times).
6.3.5 Evidence of multiple equilibria and nondegenerate equilibrium selection

By the arguments in Remark 2, the rejection of our first two specifications along with the parameter values that were rejected in our third specification are findings that are consistent with the existence of multiple equilibria in the underlying game and with an equilibrium selection mechanism that randomizes across these equilibria.

6.3.6 Potential impact of unobserved market-level shocks

We have tried to include multiple market-level relevant covariates into $X$. If there were to remain unobserved market-level heterogeneity that is publicly known to firms this would violate our independent private shocks assumption. However, consistent with the findings in our Monte Carlo experiment section, our conjecture is that the asymptotic predictions in our model would remain approximately valid in the sample observed provided that the degree of correlation induced in firms’ private shocks is relatively minor once we control for $X$. By the arguments in Remark 2, the fact that Specification 3 was not rejected implies, in turn, that the independent private shocks assumption is not rejected by the data in this case. While this is not a definitive proof for the validity of this assumption, it does suggest that it is a reasonable approximation in this example.

6.4 Results from modeling entry as a binary decision

As Table 9 showed, much of the positive correlation in the intensive margin goes away when we look only at extensive margin decisions. This led us to conjecture that the range of models that would be consistent with strategic substitutes and with the choices observed would be larger if we limited attention to a binary choice representation of entry decisions. This intuition was confirmed by our methodology. While symmetry of weights to opponents (specification 1) and constant relative interaction effects (specification 2) were still rejected, modeling entry as a binary choice decision resulted in larger confidence sets in specification 3. Furthermore, the closeness in competition between Rite Aid and Walgreens that our results uncovered was no longer evident. Specifically, as Figure 6...
Figure 5: $\theta_1^p + \theta_2^p \cdot \frac{1}{\text{POP}}$ for a range of POP values (measured in 500K). Solid black line depicts the results for the largest value of $\theta_2^p$ in our CS. Solid red line depicts the results for the smallest value of $\theta_2^p$ in our CS. Dotted lines correspond to five randomly drawn parameter values within our CS.

Player 1 (CVS) Player 2 (Rite Aid) Player 3 (Walgreens)

shows we now failed to reject that $\theta_3^p = \theta_4^p = 0$ for Rite Aid. Thus, we failed to reject that Rite Aid gives equal weights to both opponents across all markets. Hence, we conclude that key features of strategic interaction that are captured by intensive margin strategies are obscured if we focus attention solely on binary entry/no entry decisions.

6.5 Explaining the data with strategic substitutes

Our identification results are not based on the unconditional covariance\(^{12}\) between $Y^p$ and $\eta^p(Y^{-p}; X|\theta^p)$. However, given the assumption of strategic substitutes it is interesting to see if this covariance reverses the persistent positive relationship between $Y^1$, $Y^2$ and $Y^3$ summarized in Table 8. Consider in particular the correlation lower bound

$$\min_{\theta_p \in CS_n(1-\alpha)} \left\{ \rho \left( Y^p, \eta^p(Y^{-p}, Y^3; X|\theta^p) \right) \right\}.$$ 

This lower bound was $-0.06$, $-0.14$ and $-0.03$ for CVS, Rite Aid and Walgreens, respectively. Thus our strategic index, which is a weighted average of opponents’ actions (whose weights depended on $X$), is negatively associated with the strategies of each firm. This is true despite the fact that the raw actions of the firms are strongly positively correlated. This sheds light on a key lesson that we believe is generally applicable to the empirical modeling games of entry: *functional form matters*. By allowing for a rich model of the strategic index where opponent actions interact with market observables, our approach reveals that the the standard model of strategic substitutes is consistent with the positive correlation of entry in the data even without market level unobservables.

\(^{12}\)Note that the Law of Total Covariance predicts that $\text{Cov} \left( Y^p, \eta^p(Y^{-p}; X|\theta^p) \right) = E \left[ \text{Cov} \left( Y^p, \eta^p(Y^{-p}; X|\theta^p) \right) | X \right] + \text{Cov} \left( E[Y^p|X], E[\eta^p(Y^{-p}; X|\theta^p)|X] \right)$. While our identification results are related to the sign of the first component, they do not in general predict anything about the sign of the second component.
7 Concluding remarks

We studied static games with very general strategy spaces. Making some general shape restriction assumptions on the underlying payoff functions we were able to characterize observable implications that allow us to do inference on the strategic interaction component that captures economically relevant features of strategic interaction. We showed how our assumptions can arise naturally in well-known structural economic models. Our testable implications involve inequalities of nonlinear transformations of conditional moments. We introduced an econometric approach to do inference in this setting which is computationally easy to implement even in richly parameterized models with a large collection of conditioning covariates with a rich support. We described the asymptotic properties of our approach and we applied it to a model of entry in the pharmacy store industry where entry decisions are not merely binary choices but rather strategies about the number of stores that firms will open in a market. Our results uncovered economically relevant features of the underlying structural model such as a closeness in competition between two rivals: Rite Aid and Walgreens. While our econometric theory and application were based on a parametrization of the strategic index (leaving everything else about the model nonparametrically specified), our identification results can allow us to treat the index as a nonparametric function. In that case a fully nonparametric inferential approach such as sieves estimation could be employed. In addition to our conditional moment inequality restrictions, specific conjectures about the model (e.g., substitutability, symmetry, etc.) could be incorporated into the nonparametric estimator for the index. This is the subject of ongoing work in a more general context.
A Appendix– Proofs of our identification results

A.1 Proof of Result 1

Recall from (4) that
\[ \eta^p_\sigma(X) \geq \eta^p_\sigma(X) \implies \eta^p_\sigma(v; \xi^p) - \eta^p_\sigma(u; \xi^p) \leq \eta^p_\sigma(v; \xi^p) - \eta^p_\sigma(u; \xi^p) \quad \forall \ u < v \in A^p \]

Fix any \( y^p \in A^p \) and define the following indicator function,
\[ \Pi^p_\sigma(y^p; \xi^p) = \max_{v \leq y^p} \left( \min_{u \geq y^p+1} \left( \Pi^p_\sigma(v; \xi^p) - \Pi^p_\sigma(u; \xi^p) \leq 0 \right) \right). \]

By (4), we have
\[ \Pi^p_\sigma(X) \geq \eta^p_\sigma(X) \implies \Pi^p_\sigma(y^p; \xi^p) \geq \Pi^p_\sigma(y^p; \xi^p). \]

Now suppose \( \sigma^{-p} \) and \( \sigma^{-p'} \) are any pair of beliefs that produce unique expected-payoff maximizing choices for \( p \) given the realization of \( \xi^p \), and let \( y^p_\sigma(\xi^p) \) and \( y^p_{\sigma'}(\xi^p) \) denote the corresponding optimal choices. Then for any \( y^p \in A^p \),
\[ \Pi\{ y^p_\sigma(\xi^p) \leq y^p \} = \Pi^p_\sigma(y^p; \xi^p) \text{ and } \Pi\{ y^p_{\sigma'}(\xi^p) \leq y^p \} = \Pi^p_{\sigma'}(y^p; \xi^p). \]

Therefore, for any such pair of beliefs, if \( \eta^p_\sigma(X) \geq \eta^p_{\sigma'}(X) \) then \( \Pi\{ y^p_\sigma(\xi^p) \leq y^p \} \geq \Pi\{ y^p_{\sigma'}(\xi^p) \leq y^p \} \) which proves the statement in Result 1. \( \square \)

A.2 Appendix– Proof of Theorem 1

Denote \( \xi \equiv (\xi^p)_{p=1}^P \) and \( \xi^{-p} \equiv (\xi^q)_{q \neq p} \). Given \( X \), let \( J \) denote the number of BNE \{\( \sigma_j(X) \)\}_j that the selection mechanism \( S \) can choose with positive probability, and let \( P_j^S(X) \) denote the probability that \( S \) selects the \( j \)th BNE \( (\sigma_j(X)) \), conditional on \( X \). Our assumptions maintain that \( S \) concentrates on BNE that have a unique optimal choice. Denote it as \( y^p_{\sigma_j}(\xi^p) \) for the \( j \)th BNE. First, consider
\[ E_{\xi^{-p}|X} \left[ \eta^p(y^p_{\sigma_j}(\xi^{-p}); X) \right. \left| X \right]. \]

This is the expected value of \( \eta^p \), conditional on \( X \), in the \( j \)th BNE. By definition, this is equal to \( \eta^p(y^p_{\sigma_j}(\xi^p)) \), which was defined previously as
\[ \eta^p(y^p_{\sigma_j}(\xi^p)) = \sum_{y^{-p} \in A^{-p}} \sigma_{\bar{y}}^{-p} y^{-p}|X) \cdot \eta^p(y^{-p}; X). \]

Now fix any \( y^p \in A^p \). By iterated expectations we have
\[ E \left[ \Pi\{ y^p \leq y^p \} \cdot \eta^p(Y^{-p}; X) \right| X] = \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi_1|X} \left[ \Pi\{ y^p_{\sigma_j}(\xi^p) \leq y^p \} \right. \left. \cdot \eta^p(y^p_{\sigma_j}(\xi^{-p}); X) \right| X] \]
Assumption 3 (independent private shocks, i.e. $\xi \perp \xi_p^t|X$) yields:

$$E \left[ \mathbb{I} \{ Y^p \leq y^p \} \cdot \eta^p(Y^{-p}; X)|X \right]$$

$$= \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi^p|X} \left[ \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} | X \right] \cdot E_{\xi^{-p}|X} \left[ \eta^p(y_{\sigma, j}^{-p}(\xi^{-p}); X) | X \right] ,$$

$$= \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi^p|X} \left[ \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} | X \right] \cdot \pi_{\sigma, j}^p(X).$$

Therefore, by Assumption 3 we can express

$$E \left[ \mathbb{I} \{ Y^p \leq y^p \} \cdot \eta^p(Y^{-p}; X)|X \right] = E_{\xi^p|X} \left[ \sum_{j=1}^{J} P_j^S(X) \cdot \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} \cdot \pi_{\sigma, j}^p(X) \right] \tag{A-1}$$

Next note that

$$E \left[ \mathbb{I} \{ Y^p \leq y^p \} | X \right] = E \left[ \eta^p(Y^{-p}; X)|X \right]$$

$$= \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi^p|X} \left[ \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} | X \right] \times \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi^{-p}|X} \left[ \eta^p(y_{\sigma, j}^{-p}(\xi^{-p}); X) | X \right] \tag{A-2}$$

$$= \sum_{j=1}^{J} P_j^S(X) \cdot E_{\xi^p|X} \left[ \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} | X \right] \times \sum_{j=1}^{J} P_j^S(X) \cdot \pi_{\sigma, j}^p(X)$$

Combining (A-1)-(A-2) we then have

$$E \left[ \mathbb{I} \{ Y^p \leq y^p \} \cdot \eta^p(Y^{-p}; X)|X \right] - E \left[ \mathbb{I} \{ Y^p \leq y^p \} | X \right] \cdot E \left[ \eta^p(Y^{-p}; X)|X \right] =$$

$$E_{\xi^p|X} \left[ \sum_{j=1}^{J} P_j^S(X) \cdot \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} \cdot \pi_{\sigma, j}^p(X) \right] - \left( \sum_{j=1}^{J} P_j^S(X) \cdot \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} \right) \tag{A-3}$$

$$\times \left( \sum_{j=1}^{J} P_j^S(X) \cdot \pi_{\sigma, j}^p(X) \right) \tag{A-3}$$

By Result 1, w.p.1 in $(\xi^p)$ we have

$$\sum_{j=1}^{J} P_j^S(X) \cdot \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} \cdot \pi_{\sigma, j}^p(X)$$

$$- \left( \sum_{j=1}^{J} P_j^S(X) \cdot \mathbb{I} \left\{ y_{\sigma, j}^p(\xi^p) \leq y^p \right\} \right) \times \left( \sum_{j=1}^{J} P_j^S(X) \cdot \pi_{\sigma, j}^p(X) \right) \geq 0 \quad \forall y^p \in A^p. \tag{A-4}$$
To see why, simple algebra can be used to show that

\[
\sum_{j=1}^{J} P_{j}^{S}(X) \cdot \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\} \cdot \eta_{\sigma,j}^{p}(X) \\
- \left( \sum_{j=1}^{J} P_{j}^{S}(X) \cdot \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\} \right) \times \left( \sum_{j=1}^{J} P_{j}^{S}(X) \cdot \eta_{\sigma,j}^{p}(X) \right) \\
= \sum_{J=1}^{J} \sum_{j=1}^{J} P_{j}^{S}(X) P_{j}^{S}(X) \cdot \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\} \cdot (1 - \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\}) \cdot \left( \eta_{\sigma,j}^{p}(X) - \eta_{\sigma,j}^{p}(X) \right) \geq 0,
\]

where the last inequality follows from Result 1 which implies that, w.p.1 in \(\xi^{p}\) and \(\forall y^{p}\),

\[\eta_{\sigma,j}^{p}(X) < \eta_{\sigma,j}^{p}(X) \implies \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\} \leq \mathbb{1}\{y_{\sigma,j}^{p}(\xi^{p}) \leq y^{p}\}.\]

From (A-3) and (A-4) it follows that, w.p.1 in \(\xi^{p}\) in \(X\) we must have

\[E[\mathbb{1}\{Y^{p} \leq y^{p}\} \cdot \eta^{p}(Y^{-p}; X)|X] \geq E[\mathbb{1}\{Y^{p} \leq y^{p}\}|X] \cdot E[\eta^{p}(Y^{-p}; X)|X] \quad \forall y^{p}.\]

This concludes the proof.  \(\Box\)

**B  Econometric appendix**

We focus on settings where the researcher observes an iid sample \((Y_{i}^{p})_{p=1}^{P}, X_{i})_{i=1}^{N}\), with\(^{13}\) \((Y_{i}^{p})_{p=1}^{P}, X_{i}) \sim F\). We assume that \(X\) can be split as \(X = (X^{c}, X^{d})\), where \(X^{c}\) have absolutely continuous distribution with respect to Lebesgue measure and \(X^{d}\) have a discrete distribution. We denote the dimension of \(X^{c}\) by \(q\). We begin by describing the preliminary conditions needed for our construction.

**B.1 Specifying an “inference range”**

Let \(X \subset \text{Supp}(X)\) denote a prespecified set such that

\[X \cap \text{Supp}(X^{c}) \subset \text{int}(\text{Supp}(X^{c})).\]

We maintain the assumption that \(f_{X}(x) \geq f > 0\) for all \(x \in X\). Let\(^{14}\) \(I_{X}(x) = \mathbb{1}\{x \in X\}\). Let

\[T_{X}^{p}(\theta^{p}) = E_{Y_{p}, X} \left[ \max \{r^{p}(Y^{p}|X; \theta^{p}), 0\} \cdot I_{X}(X) \right]. \tag{B-1}\]

By construction, \(T_{X}^{p}(\theta^{p}) \geq 0\), and \(T_{X}^{p}(\theta^{p}) = 0\) if and only if \(Pr \left(r^{p}(Y^{p}|X; \theta^{p}) \leq 0 | X \in X\right) = 1\). We aggregate these one-sided expectations as

\[T_{X}(\theta) = \sum_{p=1}^{P} T_{X}^{p}(\theta^{p}).\]

\(^{13}\)We generalize our assumptions to a setting where \((Y_{i}^{p})_{p=1}^{P}, X_{i})_{i=1}^{N}\) is a triangular array in Section B.6.

\(^{14}\)The indicator function \(I_{X}\) could be replaced with a smooth “trimming” function.
Note that $T_X(\theta) \geq 0$, and $T_X(\theta) = 0$ if and only if $Pr \left( \tau^p(Y^p|X; \theta^p) \leq 0 \middle| X \in \mathcal{X} \right) = 1$ for $p = 1, \ldots, P$. The inference range $\mathcal{X}$ will be assumed to be such that the nonparametric estimators involved in our construction have appropriate asymptotic properties uniformly over it. Given our choice of $\mathcal{X}$, we focus attention of the following superset of the identified set $\Theta^I$,

$$\Theta^I_{\mathcal{X}} = \{ \theta \in \Theta : T^p_X(\theta) = 0 \text{ for } p = 1, \ldots, P \} .$$

Note that $\Theta^I \subseteq \Theta^I_{\mathcal{X}}$, where $\Theta^I = \{ \theta \in \Theta : Pr \left( \tau^p(Y^p|X; \theta^p) \leq 0 \right) = 1 \text{ for } p = 1, \ldots, P \}$.

### B.2 Estimators involved in our construction

We employ kernel-based nonparametric estimators. $K : \mathbb{R}^q \to \mathbb{R}$ will denote our kernel function. For a given $x \equiv (x^c, x^d)$ and $h > 0$ define

$$\mathcal{H}(X_i - x; h) = K \left( \frac{X_i^c - x^c}{h} \right) \cdot \mathbb{I} \{ X_i^d - x^d = 0 \} .$$

Let $h_n \to 0$ be a nonnegative bandwidth sequence. For a given $x \equiv (x^c, x^d)$, $y^p$ and $\theta^p$ our estimators are of the form

$$\hat{f}_X(x) = (nh_n^2)^{-1} \sum_{i=1}^{n} \mathcal{H}(X_i - x; h_n) ,$$

$$\hat{F}_{Y^p}(y^p|x) = \left( nh_n^2 \cdot \hat{f}_X(x) \right)^{-1} \sum_{i=1}^{n} \mathbb{I} \{ Y_i^p \leq y^p \} \cdot \mathcal{H}(X_i - x; h_n) ,$$

$$\hat{\lambda}^p(x; \theta^p) = \left( nh_n^2 \cdot \hat{f}_X(x) \right)^{-1} \sum_{i=1}^{n} \eta^p(Y_i^{-p}; x|\theta^p) \cdot \mathcal{H}(X_i - x; h_n) ,$$

$$\hat{\mu}^p(y^p|x; \theta^p) = \hat{F}_{Y^p}(y^p|x) \cdot \hat{\lambda}^p(x; \theta^p) - \hat{\mu}^p(x; \theta^p) .$$

Our estimators for $T^p_X(\theta^p)$ and $T_X(\theta)$ are

$$\hat{T}^p_X(\theta^p) = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}^p(\theta^p) \cdot \mathbb{I} \{ \hat{\tau}^p(X_i; \theta^p) \geq -b_n \} \cdot \mathcal{H}(X_i) ,$$

$$\hat{T}_X(\theta) = \sum_{p=1}^{P} \hat{T}^p_X(\theta^p) .$$

(B-2)

where $b_n \to 0$ is a nonnegative sequence whose properties will be described below.

### B.3 Basic Assumptions

**Assumption B1. (Smoothness)**

As before, express any $x \in \text{Supp}(X)$ generically as $x \equiv (x^c, x^d)$ with $x^c$ corresponding to the continuously distributed elements in $X$. Denote

$$\mathcal{W} = \{(x,y) \in \text{Supp}(X,Y); x \in \mathcal{X} \} .$$

39
Recall that we defined before

\[ F_{Y^n}(y^n|x) = E_{Y^n|X} \left[ \mathbb{1}\{Y^n \leq y^n\} \mid X = x \right], \]

\[ \lambda^n(x; \theta^n) = E_{Y^n|X} \left[ \eta^n(Y^n; x\theta^n) \mid X = x \right], \]

\[ \mu^n(y^n; x; \theta^n) = E_{Y^n|X} \left[ \mathbb{1}\{Y^n \leq y^n\} \cdot \eta^n(Y^n; x\theta^n) \mid X = x \right], \]

\[ \tau^n(y^n; x; \theta^n) = F_{Y^n}(y^n|x) \cdot \lambda^n(x; \theta^n) - \mu^n(y^n; x; \theta^n). \]

For almost every \((x, y^n) \in \mathcal{W}, x' \in X\) and every \(\theta^n \in \Theta\), the following objects are \(M\) times differentiable with respect to \(x^n\) with bounded derivatives,

\[ F_{Y^n}(y^n|x), \quad f^n(x), \quad E_{Y^n|X} \left[ \eta^n(Y^n; x'|\theta^n) \mid X = x \right], \]

\[ E_{Y^n|X} \left[ \mathbb{1}\{Y^n \leq y^n\} \cdot \eta^n(Y^n; x'|\theta^n) \mid X = x \right]. \]

Now let

\[ \gamma^n_I(y^n, x; \theta^n) = E_{Y^n|X} \left[ \mathbb{1}\{y^n \leq Y^n\} \cdot \mathbb{1}\{\tau^n(Y^n|x; \theta^n) \geq 0\} \mid X = x \right], \]

\[ \gamma^n_{II}(y^n, x; \theta^n) = E_{Y^n|X} \left[ F_{Y^n}(Y^n|x) \cdot \mathbb{1}\{\tau^n(Y^n|x; \theta^n) \geq 0\} \mid X = x \right], \]

\[ \gamma^n_{III}(x; \theta^n) = E_{Y^n|X} \left[ \eta^n(Y^n|x; \theta^n) \cdot \mathbb{1}\{\tau^n(Y^n|x; \theta^n) \geq 0\} \mid X = x \right]. \]

For almost every \((x, y^n) \in \mathcal{W}\) and every \(\theta^n \in \Theta\), the three objects defined above are \(M\) times differentiable with respect to \(x^n\) with bounded derivatives, and this is also satisfied by the trimming function \(\mathbb{1}(x)\). For given \(y^n, x\) and \(\theta^n\) define

\[ Q^{\mu}_{F_{Y^n}}(y^n|x) = F_{Y^n}(y^n|x) \cdot f^n(x), \quad Q^{\lambda^n}(x; \theta^n) = \lambda^n(x; \theta^n) \cdot f^n(x), \]

\[ Q^{\mu^n}(y^n|x; \theta^n) = \mu^n(y^n|x; \theta^n) \cdot f^n(x). \]

Then for some \(\overline{Q} < \infty\),

\[ \sup_{(x, y^n) \in \mathcal{W}} \left| Q^{\mu}_{F_{Y^n}}(y^n|x) \right| \leq \overline{Q}, \quad \sup_{x \in X, \theta^n \in \Theta} \left| Q^{\lambda^n}(x; \theta^n) \right| \leq \overline{Q}, \]

\[ \sup_{(x, y^n) \in \mathcal{W}, \theta^n \in \Theta} \left| Q^{\mu^n}(y^n|x; \theta^n) \right| \leq \overline{Q}. \]

Note that the restrictions in Assumption B1 would likely rule out equilibrium selection rules that generate nonsmoothness. Following the analysis in Bajari, Hong, Kreiner, and Nekipelov (2009), the smoothness conditions with respect to continuous state variables in the equilibrium selection mechanism require that the equilibrium paths to not bifurcate for almost all values of the continuous state variable, or that a smooth path is chosen at the points of bifurcation.

**Assumption B2. (Kernels and Bandwidths)** Let \(M\) be as described in Assumption B1. We use a bias-reducing kernel \(K\) of order \(M\) with bounded support. The kernel is a function of bounded variation, symmetric around zero and satisfies \(\sup_{v \in \mathbb{R}^v} |K(v)| \leq \overline{K} < \infty\). The bandwidth sequences \(b_n\) and \(h_n\) are such that, for a small enough \(\epsilon > 0\),

\[ n^{1/2-\epsilon_1} \cdot h_n^{\rho} \cdot b_n \rightarrow \infty, \quad n^{1/2+\epsilon_2} \cdot b_n^2 \rightarrow 0, \quad n^{1/2+\epsilon_2} \cdot h_n^M \rightarrow 0. \]
Focus on bandwidths of the type \( h_n \propto n^{-\alpha_h} \) and \( b_n \propto n^{-\alpha_b} \). Let \( \tau > 0 \) be an arbitrarily small, but strictly positive constant and let \( \alpha_h = \frac{1}{2M} + \tau \) and \( \alpha_b = \frac{1}{4} + \tau \). The conditions in Assumption B2 will be satisfied if

\[
M \geq \left[ \frac{2 \cdot q}{1 - 4 \cdot \tau (2 + q)} \right].
\]

For example, suppose \( q = 8 \) (as in our empirical application). Then we need \( M \geq 17 \). Recall that \( M \) is the number of derivatives assumed to exist in Assumption B1 and it also corresponds to the order of the kernel employed.

Our framework must allow for the existence of parameter values \( \theta^p \in \Theta \) such that \( \tau^p(Y^p|X; \theta^p) \) has a point mass at zero. While we allow for that, the following assumption restricts the way in which the distribution of \( \tau^p(Y^p|X; \theta^p) \) approaches zero from the left. In essence the condition assumes that the density of \( \tau^p(Y^p|X; \theta^p) \) is bounded in a neighborhood of the type \([-\overline{b}, 0)\) where \( \overline{b} > 0 \).

**Assumption B3. (A regularity condition)**

There exist constants \( \overline{b} > 0 \) and \( A > 0 \) such that, for each \( p \) and each \( \theta^p \in \Theta \),

\[
Pr \left( -b \leq \tau^p(Y^p|X; \theta^p) < 0 \middle| X \in X \right) \leq b \cdot A \quad \forall \, 0 < b \leq \overline{b}.
\]

Note that Assumption B3 allows for \( \tau^p(Y^p|X; \theta^p) \) to have a point mass at zero. It merely assumes the existence of a neighborhood \([-\overline{b}, 0)\) such that the density of \( \tau^p(Y^p|X; \theta^p) \) is bounded, uniformly over \( \theta^p \in \Theta \) in that neighborhood.

**Assumption B4. (Empirical process and manageability conditions)**

For each \( p \) the following conditions are satisfied. Let

\[
\tau^p(y^p) = \sup_{x \in X, \theta^p \in \Theta} |\eta^p(y^p; x|\theta^p)|.
\]

Then \( E \left[ \exp \left\{ \left( \tau^p(Y^p) \right)^2 \cdot \epsilon \right\} \right] \leq C < \infty \) for some \( \epsilon > 0 \). That is, \( (\tau^p(Y^p))^2 \) possesses a moment generating function.

(i) The classes of functions

\[
\mathcal{F} = \left\{ f \colon f(y^p) = \eta^p(y^p; x|\theta^p) \text{ for some } (x, \theta^p) \in X \times \Theta \right\},
\]

\[
\mathcal{F}' = \left\{ f \colon f(x) = \lambda^p(x; \theta^p) \text{ for some } \theta^p \in \Theta \right\},
\]

\[
\mathcal{F}'' = \left\{ f \colon f(y^p, x) = \mu^p(y^p|x; \theta^p) \text{ for some } \theta^p \in \Theta \right\},
\]

are Euclidean (see Definition 2.7 in Pakes and Pollard (1989)) with respect to envelopes \( \tau^p(\cdot) \), \( \overline{F}^p(\cdot) \) and \( \overline{F}''(\cdot) \) respectively, where \( \overline{F}^p(Y^p) \) satisfies the existence-of-moments conditions described above, and \( \overline{F}'(\cdot) \) and \( \overline{F}''(\cdot) \) satisfy \( E \left[ \overline{F}'(X)^2 \right] < \infty \) and \( E \left[ \overline{F}''(Y^p, X)^2 \right] < \infty \).

(ii) Let \( \overline{b} > 0 \) be as described in Assumption B3. The class of functions

\[
\mathcal{G} = \left\{ g \colon g(x, y) = 1 \cdot \{ -b \leq \tau^p(x, y; \theta^p) < 0 \} \cdot 1_{X}(x) \text{ for some } \theta^p \in \Theta, 0 < b \leq \overline{b} \right\},
\]

41
is Euclidean with respect to envelope 1.

Sufficient conditions for a class of functions to be Euclidean can be found, e.g. in Nolan and Pollard (1987) and Pakes and Pollard (1989). Once a parametric family is chosen for \( \eta^p \), those conditions can be used to verify part (i) of Assumption B4. In particular, \( \eta^p \) does not have to be smooth (or even continuous) to satisfy the Euclidean property. For part (ii) fix \( b \in \mathbb{R} \) and let \( \mathcal{N}(x, y; b) \) denote the number of points in \( \Theta \) where \( \tau(x, y; \theta^p) - b \) changes sign. Suppose \( \sup_{(x, y) \in X \times A} \mathcal{N}(x, y; b) \leq \bar{N} < \infty \) for all \( 0 < b \leq \bar{b} \). By Lemma 1 in Asparouhova, Golanski, Kasprzyk, Sherman, and Asparouhov (2002) this ensures that the class of sets indexed by the indicator functions in part (ii) of our assumption is a VC class of sets (see Definition 2.2 in Pakes and Pollard (1989)). The Euclidean property for said class of functions follows from here by the results in Pakes and Pollard (1989).

### B.4 Asymptotic properties of \( \hat{T}_X^p(\theta^p) \)

The following theorem summarizes the key asymptotic properties of \( \hat{T}_X^p(\theta^p) \) under our assumptions.

**Theorem 2.** Let

\[
\psi_U^p(Y, X; \theta^p) = \left[ (\gamma_p^I(Y^p, X; \theta^p) - \gamma_p^{II}(X; \theta^p)) \cdot \lambda^p(X; \theta^p) + (\eta^p(Y^{-p}, X|\theta^p) - \lambda^p(X; \theta^p)) \cdot \gamma_p^{II}(X; \theta^p) \right] \cdot I(X),
\]

and

\[
\psi^p(Y_i, X_i; \theta^p) = (\max \{ \tau^p(Y_i^p|X_i; \theta^p), 0 \} \cdot I(X_i) - T_X^p(\theta^p)) + \psi_U^p(Y_i, X_i; \theta^p).
\]

If Assumptions B1-B4 hold, then

\[
\hat{T}_X^p(\theta^p) = T_X^p(\theta^p) + \frac{1}{n} \sum_{i=1}^n \psi^p(Y_i, X_i; \theta^p) + \varepsilon_{p,n}(\theta^p),
\]

where

\[
\psi^p(Y_i, X_i; \theta^p) = (\max \{ \tau^p(Y_i^p|X_i; \theta^p), 0 \} \cdot I(X_i) - T_X^p(\theta^p)) + \psi_U^p(Y_i, X_i; \theta^p),
\]

and

\[
\sup_{\theta^p \in \Theta} |\varepsilon_{p,n}(\theta^p)| = O_p\left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0.
\]

The “influence function” \( \psi^p \) has two key properties:

(i) \( E[\psi^p(Y_i, X_i; \theta^p)] = 0 \) \( \forall \theta^p \in \Theta \).

(ii) \( \psi^p(Y_i, X_i; \theta^p) = 0 \) \( \forall \theta^p : \tau^p(Y^p|X; \theta^p) < 0 \) w.p.1.

Property (ii) can be verified immediately by inspection. Property (i) can be verified using iterated expectations and we prove it in Appendix B.4.2, below. Let \( \psi(Y_i, X_i; \theta) = \sum_{p=1}^P \psi^p(Y_i, X_i; \theta^p) \). By Theorem 2,

\[
\hat{T}_X(\theta) = T_X(\theta) + \frac{1}{n} \sum_{i=1}^n \psi(Y_i, X_i; \theta) + \varepsilon_n(\theta),
\]

where

\[
\sup_{\theta \in \Theta} |\varepsilon_n(\theta)| = O_p\left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0.
\]

(B-3)
And \(\psi(Y_i, X_i; \theta)\) is identified and has two key properties:

(i) \(E [\psi(Y_i, X_i; \theta)] = 0 \) \(\forall \theta \in \Theta\).

(ii) Let
\[
\Theta^I_X = \{ \theta \in \Theta : \tau^p(Y^p|X; \theta^p) < 0 \ \text{w.p.1.} \ \forall p = 1, \ldots, P. \}
\]

Then \(\psi(Y_i, X_i; \theta) = 0 \) w.p.1 \(\forall \theta \in \Theta^I_X\).

We now proceed to prove Theorem 2.

**B.4.1 Proof of Theorem 2**

In Assumption B1 we described \(\mathcal{W}\) as
\[
\mathcal{W} = \{(x, y) \in \text{Supp}(X, Y) : x \in \mathcal{X}\},
\]
where \(\mathcal{X} \subseteq \text{Supp}(X)\) is a prespecified set such that \(\mathcal{X} \cap \text{Supp}(X^c) \subseteq \text{int}(\text{Supp}(X^c))\). We maintain the assumption that \(f_X(x) \geq \underline{f} > 0\) for all \(x \in \mathcal{X}\). We will split the proof in three steps.

**Step 1**

Our first step is to show that under our assumptions, there exist \(D_1 > 0, D_2 > 0\) and \(D_3 > 0\) such that
\[
\Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \tau^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq b_n \right) \\
\leq D_1 \exp \left\{ -\sqrt{n} h_n^4 \left( D_2 \cdot b_n - D_3 \cdot h_n^M \right) \right\}.
\]

Fix \(y^p, x\) and \(\theta^p\) and let \(Q^p_{F_{Y^p}}(y^p|x), Q^p_{\lambda^p}(x; \theta^p)\) and \(Q^p_{\mu^p}(y^p|x; \theta^p)\) be as defined in Assumption B1. We estimate these functionals with
\[
\tilde{Q}^p_{F_{Y^p}}(y^p|x) = (n h_n^p)^{-1} \sum_{i=1}^n \mathbb{1} \{ Y_{i}^{p} \leq y^p \} \cdot \mathcal{H}(X_i - x; h_n),
\]
\[
\tilde{Q}^p_{\lambda^p}(x; \theta^p) = (n h_n^p)^{-1} \sum_{i=1}^n \eta^p \left( Y_{i}^{p-}; x|\theta^p \right) \cdot \mathcal{H}(X_i - x; h_n),
\]
\[
\tilde{Q}^p_{\mu^p}(y^p|x; \theta^p) = (n h_n^p)^{-1} \sum_{i=1}^n \mathbb{1} \{ Y_{i}^{p} \leq y^p \} \cdot \eta^p \left( Y_{i}^{p-}; x|\theta^p \right) \cdot \mathcal{H}(X_i - x; h_n).
\]

Using an \(M^{th}\) order approximation, our smoothness restrictions in Assumption B1 imply the existence of a finite constant \(\overline{M}\) such that,
\[
\sup_{x \in \mathcal{X}} \left| E \left[ \tilde{f}_X(x) \right] - f_X(x) \right| \leq \overline{M} \cdot h_n^M,
\]
\[
\sup_{(x, y^p) \in \mathcal{W}} \left| E \left[ \tilde{Q}^p_{F_{Y^p}}(y^p|x) \right] - Q^p_{F_{Y^p}}(y^p|x) \right| \leq \overline{M} \cdot h_n^M,
\]
\[
\sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| E \left[ \tilde{Q}^p_{\lambda^p}(x; \theta^p) \right] - Q^p_{\lambda^p}(x; \theta^p) \right| \leq \overline{M} \cdot h_n^M,
\]
\[
\sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| E \left[ \tilde{Q}^p_{\mu^p}(y^p|x; \theta^p) \right] - Q^p_{\mu^p}(y^p|x; \theta^p) \right| \leq \overline{M} \cdot h_n^M.
\]
Invoking Lemma 22 in Nolan and Pollard (1987) and Lemmas 2.4 and 2.14 in Pakes and Pollard (1989), having a kernel of bounded variation implies that the class of functions

$$G = \left\{ g : g(x) = \mathcal{H}(x - v; h) \text{ for some } v \in \mathbb{R}^{dim(X)} \text{ and some } h > 0 \right\}$$

is Euclidean with respect to the constant envelope $\mathcal{K}$. Lemma 2.4 in Pakes and Pollard (1989) also implies that the class of functions

$$G = \left\{ g : g(y^p) = 1 \{ y^p \leq v \} \text{ for some } v \in \mathbb{R} \right\}$$

is Euclidean with respect to the envelope 1. Combined with Assumption B4(i) and Lemma 2.14 in Pakes and Pollard (1989) we have that the classes of functions

$$\mathcal{F}_1 = \left\{ f : f(y^p, x) = \eta^p(y^p; u|\theta^p) \cdot \mathcal{H}(x - u; h) \text{ for some } u \in \mathcal{X} \text{ and } \theta^p \in \Theta \right\},$$

$$\mathcal{F}_2 = \left\{ f : f(y, x) = 1 \{ y^p \leq v \} \cdot \eta^p(y^p; u|\theta^p) \cdot \mathcal{H}(x - u; h) \text{ for some } v \in \mathbb{R}, u \in \mathcal{X} \text{ and } \theta^p \in \Theta \right\}$$

are Euclidean with respect to the envelope $\mathcal{K} \cdot \eta^p(\cdot)$. Since this envelope has a moment generating function by Assumption B4(i), the maximal inequality results in Chapter 7 of Pollard (1990) combined with the bias conditions in B-4 imply that there exist positive constants $A_1$, $A_2$ and $A_3$ such that for any $\delta > 0$,

$$\Pr \left( \sup_{x \in \mathcal{X}} \left| \begin{array}{c} \tilde{f}_X(x) - f_X(x) \end{array} \right| \geq \delta \right) \leq A_1 \cdot \exp \left\{ - \frac{\gamma n}{h_n^q} \left( A_2 \cdot \delta - A_3 \cdot h_n^M \right) \right\},$$

$$\Pr \left( \sup_{(x,y^p) \in \mathcal{W}} \left| \begin{array}{c} \tilde{Q}_{Y^p,X}(y^p|x) - Q_{Y^p,X}(y^p|x) \end{array} \right| \geq \delta \right) \leq A_1 \cdot \exp \left\{ - \frac{\gamma n}{h_n^q} \left( A_2 \cdot \delta - A_3 \cdot h_n^M \right) \right\},$$

$$\Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \begin{array}{c} \tilde{Q}_{\mu^p}(y^p|x; \theta^p) - Q_{\mu^p}(y^p|x; \theta^p) \end{array} \right| \geq \delta \right) \leq A_1 \cdot \exp \left\{ - \frac{\gamma n}{h_n^q} \left( A_2 \cdot \delta - A_3 \cdot h_n^M \right) \right\}. \quad (B-5)$$

For any $x$ such that $f_X(x) > 0$ define

$$\psi_{Y^p,X_i,y^p,x;h} = \frac{1 \{ y^p \leq y^p \} - F_{Y^p}(y^p|x)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h),$$

$$\psi_{\lambda^p,Y_i^{-p},X_i,x;\theta^p,h} = \frac{\eta^p(Y_i^{-p}; x|\theta^p) - \lambda^p(x|\theta^p)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h),$$

$$\psi_{\mu^p,Y_i,X_i,y^p,x;\theta^p,h} = \frac{1 \{ y^p \leq y^p \} \cdot \eta^p(Y_i^{-p}; x|\theta^p) - \mu^p(y^p|x; \theta^p)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h). \quad (B-6)$$

And let

$$\tilde{\zeta}_{Y^p,X}(y^p,x) = \left[ \begin{array}{c} \tilde{Q}_{Y^p,X}(y^p|x) - Q_{Y^p,X}(y^p|x) \\ \tilde{f}_X(x) - f_X(x) \end{array} \right] \cdot \mathcal{H}(X_i - x; h),$$

$$\tilde{\zeta}_{\lambda^p,X}(x; \theta^p) = \left[ \begin{array}{c} \tilde{Q}_{\lambda^p,X}(x; \theta^p) - Q_{\lambda^p,X}(x; \theta^p) \\ \tilde{f}_X(x) - f_X(x) \end{array} \right] \cdot \mathcal{H}(X_i - x; h),$$

$$\tilde{\zeta}_{\mu^p}(y^p,x; \theta^p) = \left[ \begin{array}{c} \tilde{Q}_{\mu^p}(y^p|x; \theta^p) - Q_{\mu^p}(y^p|x; \theta^p) \\ \tilde{f}_X(x) - f_X(x) \end{array} \right] \cdot \mathcal{H}(X_i - x; h).$$

---

15 See Definition 2.7 in Pakes and Pollard (1989).
Note that (B-5) implies that for any $\delta > 0$,

$$
Pr \left( \sup_{(x,y^p) \in W} \left| \tilde{\xi}_{F_{V^p}}(y^p, x) \right| \geq \delta \right) \leq Pr \left( \sup_{(x,y^p) \in W} \left| \tilde{Q}_{F_{V^p}}(y^p|x) - Q_{F_{V^p}}(y^p|x) \right| \geq \frac{\delta}{\sqrt{2}} \right)
$$

$$
+ Pr \left( \sup_{x \in X} \left| \tilde{f}_X(x) - f_X(x) \right| \geq \frac{\delta}{\sqrt{2}} \right)
$$

$$
\leq A_1 \cdot \exp \left\{ - \left( \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right)^2 \right\} + A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\}
$$

Similarly (B-5) yields

$$
Pr \left( \sup_{x \in X, \theta \in \Theta} \left| \tilde{\xi}_{\lambda \rho}(x, \theta^p) \right| \geq \delta \right) \leq 2 \cdot A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\},
$$

$$
Pr \left( \sup_{(x,y^p) \in W, \theta \in \Theta} \left| \tilde{\xi}_{\lambda \rho}(y^p, x, \theta^p) \right| \geq \delta \right) \leq 2 \cdot A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\}.
$$

Whenever $\tilde{f}_X(x) > 0$ and $f_X(x) > 0$, a second order approximation yields the following results:

$$
\tilde{F}_{V^p}(y^p|x) - F_{V^p}(y^p|x) = \frac{1}{n h_n^p} \sum_{i=1}^n \psi_{F_{V^p}} (Y_i^p, X_i, y^p, h_n^p) + \xi_{F_{V^p}}(y^p, x),
$$

where $\xi_{F_{V^p}}(y^p, x) = \frac{1}{2} \zeta_{F_{V^p}}(y^p, x)' \begin{pmatrix} 0 & -\frac{1}{f_{F_{V^p}}(x)} \\ -\frac{1}{f_{F_{V^p}}(x)} & \frac{2 \tilde{Q}_{F_{V^p}}(y^p|x)}{f_{F_{V^p}}(x)} \end{pmatrix} \tilde{\xi}_{F_{V^p}}(y^p, x)$

where $\left( \tilde{f}_X(x), \tilde{Q}_{F_{V^p}}(y^p|x) \right)$ belongs in the line segment connecting $\left( \tilde{f}_X(x), \tilde{Q}_{F_{V^p}}(y^p|x) \right)$ and $\left( f_X(x), Q_{F_{V^p}}(y^p|x) \right)$.

$$
\tilde{\lambda}(x; \theta^p) - \lambda(x; \theta^p) = \frac{1}{n h_n^p} \sum_{i=1}^n \psi_{\lambda \rho} (Y_i^p, X_i, \theta^p, h_n) + \xi_{\lambda \rho}(x, \theta^p),
$$

where $\xi_{\lambda \rho}(x, \theta^p) = \frac{1}{2} \zeta_{\lambda \rho}(x, \theta^p)' \begin{pmatrix} 0 & -\frac{1}{f_{\lambda \rho}(x)} \\ -\frac{1}{f_{\lambda \rho}(x)} & \frac{2 \tilde{Q}_{\lambda \rho}(y^p|x, \theta^p)}{f_{\lambda \rho}(x)} \end{pmatrix} \tilde{\xi}_{\lambda \rho}(x, \theta^p)$

where $\left( \tilde{f}_X(x), \tilde{Q}_{\lambda \rho}(x; \theta^p) \right)$ belongs in the line segment connecting $\left( \tilde{f}_X(x), \tilde{Q}_{\lambda \rho}(x; \theta^p) \right)$ and $\left( f_X(x), Q_{\lambda \rho}(x; \theta^p) \right)$.

$$
\tilde{\mu}(y^p|x; \theta^p) - \mu(y^p|x; \theta^p) = \frac{1}{n h_n^p} \sum_{i=1}^n \psi_{\mu \rho} (Y_i, X_i, y^p, x, \theta^p, h_n) + \xi_{\mu \rho}(y^p, x, \theta^p),
$$

where $\xi_{\mu \rho}(y^p, x, \theta^p) = \frac{1}{2} \zeta_{\mu \rho}(y^p, x, \theta^p)' \begin{pmatrix} 0 & -\frac{1}{f_{\mu \rho}(y^p,x,\theta^p)} \\ -\frac{1}{f_{\mu \rho}(y^p,x,\theta^p)} & \frac{2 \tilde{Q}_{\mu \rho}(y^p|x,\theta^p)}{f_{\mu \rho}(y^p,x,\theta^p)} \end{pmatrix} \tilde{\xi}_{\mu \rho}(y^p, x, \theta^p)$

where $\left( \tilde{f}_X(x), \tilde{Q}_{\mu \rho}(y^p|x; \theta^p) \right)$ belongs in the line segment connecting $\left( \tilde{f}_X(x), \tilde{Q}_{\mu \rho}(y^p|x; \theta^p) \right)$ and $\left( f_X(x), Q_{\mu \rho}(y^p|x; \theta^p) \right)$.
Let $Q$ be as described in Assumption B1. For any $0 < f^* < f$, define
\[
D(f^*) = \begin{vmatrix}
0 & -\frac{1}{f^*} \\
\frac{1}{Q} & 0
\end{vmatrix}.
\] (B-7)

Let $0 < f^* < f$ and $D(f^*)$ be as described in (B-7). Combining our previous results, for any $\delta > 0$,
\[
Pr\left(\sup_{(x,y^p)\in W} |\xi_n^{F_Y} (y^p, x)| \geq \delta\right) \leq Pr\left(\sup_{(x,y^p)\in W} |\hat{Q}_{F_Y} (y^p | x) - Q_{F_Y} (y^p | x)| \geq \overline{Q}\right)
+ Pr\left(\sup_{x\in X} |\hat{f}_X (x) - f_X (x)| \geq f - f^*\right) + Pr\left(\sup_{(x,y^p)\in W} |\xi_n^{F_Y} (y^p, x)| \geq \sqrt{\frac{2\delta}{D(f^*)}}\right)
\leq 4A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^\eta \left( A_2 \cdot \min \left\{ \sqrt{\frac{\delta}{D(f^*)}}, \overline{Q}, \frac{f - f^*}{\overline{Q}} \right\} - A_3 \cdot h_n^M \right) \right\}.
\]

And the same bound holds for
\[
Pr\left(\sup_{x\in X, \theta^p\in \Theta} |\xi_n^{\psi_{F_Y}} (x, \theta^p)| \geq \delta\right) \quad \text{and} \quad Pr\left(\sup_{(x,y^p)\in W, \theta^p\in \Theta} |\xi_n^{\psi_{F_Y}} (y^p, x, \theta^p)| \geq \delta\right).
\]

Assumption B4 and Lemma 2.14 in Pakes and Pollard (1989) we have that the classes of functions
\[
\mathcal{G}_1 = \{ g: g(y^p, x) = \psi_{F_Y} (y^p, x, v^p, u; h) : (v^p, u) \in W, h > 0 \},
\]
\[
\mathcal{G}_2 = \{ g: g(y^{-p}, x) = \psi_{\lambda_Y} (y^{-p}, x, u; \theta^p) : u \in X, \theta^p \in \Theta, h > 0 \},
\]
\[
\mathcal{G}_3 = \{ g: g(y, x) = \psi_{\mu_Y} (y, x, v^p, u; \theta^p; h) : (v^p, u) \in W, \theta^p \in \Theta, h > 0 \}
\]
are Euclidean with respect to envelopes $\frac{2\kappa^\psi (\cdot)}{L}$ and $\frac{2\kappa^\psi (\cdot)}{L}$, respectively. The existence of moments feature of $\overline{\psi} (\cdot)$ in Assumption B4 and the results in Chapter 7 of Pollard (1990) combined with the bias conditions in B-4 imply that there exist positive constants $A'_1, A'_2$ and $A'_3$ such that for any $\delta > 0$, the probabilities
\[
Pr\left(\sup_{(x,y^p)\in W} \left| \frac{1}{nh_n^\psi} \sum_{i=1}^n \psi_{F_Y} (Y_i^p, X_i, y^p, x; h_n) \right| \geq \delta\right),
\]
\[
Pr\left(\sup_{x\in X, \theta^p\in \Theta} \left| \frac{1}{nh_n^\psi} \sum_{i=1}^n \psi_{\lambda_Y} (Y_i^{-p}, X_i, x, \theta^p; h_n) \right| \geq \delta\right),
\]
\[
Pr\left(\sup_{(x,y^p)\in W, \theta^p\in \Theta} \left| \frac{1}{nh_n^\psi} \sum_{i=1}^n \psi_{\mu_Y} (Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \delta\right),
\]
are bounded above by
\[
A'_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^\eta \left( A'_2 \cdot \delta - A'_3 \cdot h_n^M \right) \right\}.
\]

46
Let \( 0 < f^* \leq f \) and \( D(f^*) \) be as described in (B-7). Combining our results, for any \( \delta > 0 \) we have

\[
Pr \left( \sup_{(x,y_p) \in W} \left| \hat{F}_{Y_p} (y_p|x) - F_{Y_p} (y_p|x) \right| \geq \delta \right) \leq Pr \left( \sup_{x \in X} \left| \hat{f}_X (x) - f_X (x) \right| \geq f - f^* \right) \\
+ Pr \left( \sup_{(x,y_p) \in W} \left| \frac{1}{n h_n^2} \sum_{i=1}^n \psi_{F_{Y_p}} (Y_i, X_i, y_p, x; h_n) \right| \geq \frac{\delta}{2} \right) \\
+ Pr \left( \sup_{(x,y_p) \in W} \left| \xi_{F_{Y_p}} (y_p, x) \right| \geq \frac{\delta}{2} \right) \\
\leq A_1 \cdot \exp \left\{ - \left( \sqrt{n} \cdot h_n^2 \left( A_2 \cdot (f - f^*) - A_3 \cdot h_n^2 \right) \right)^2 \right\} \\
+ A'_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^2 \left( A'_2 - A'_3 \cdot h_n^2 \right) \right\} \\
+ 4A_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^2 \left( A_2 \cdot \min \left\{ \sqrt{\frac{\delta}{2D}}, \frac{\sqrt{\delta}}{2D}, \frac{f - f^*}{x} \right\} - A_3 \cdot \hat{h}_n^2 \right) \right\} \\
\leq B_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^2 \left( \min \left\{ \frac{\delta}{2}, \sqrt{\frac{\delta}{2D}}, \frac{f - f^*}{x} \right\} - B_3 \cdot \hat{h}_n^2 \right) \right\}
\]

where \( B_1 = 6 \cdot \max \{ A_1, A'_1 \}, B_2 = \min \{ A_2, A'_2 \} \) and \( B_3 = \max \{ A_3, A'_3 \} \). The same type of bound is valid for

\[
Pr \left( \sup_{x \in X, \theta \in \Theta} \left| \hat{\lambda}^p (x; \theta^p) - \lambda^p (x; \theta^p) \right| \geq \delta \right), \\
Pr \left( \sup_{(x,y_p) \in W, \theta \in \Theta} \left| \hat{\mu}^p (y_p|x; \theta^p) - \mu^p (y_p|x; \theta^p) \right| \geq \delta \right).
\]

The previous results allow us now to turn our attention to \( \hat{\tau}^p (y_p|x; \theta^p) \). For \( h > 0 \) let

\[
\psi_{\tau^p} (Y_i, X_i, y_p, x, \theta^p; h) \\
= \lambda^p (x; \theta^p) \cdot \psi_{F_{Y_p}} (Y_i, X_i, y_p, x; h) + F_{Y_p} (y_p|x) \cdot \psi_{\lambda^p} (Y_i, X_i, y_p, x, \theta^p; h) - \psi_{\mu^p} (Y_i, X_i, y_p, x, \theta^p; h) \\
= \left[ \lambda^p (x; \theta^p) \cdot (1 \{ Y_i^p \leq y_p \} - F_{Y_p} (y_p|x)) + F_{Y_p} (y_p|x) \cdot (\mu^p (Y_i^p, x|\theta^p) - \lambda^p (x; \theta^p)) \right] - (1 \{ Y_i^p \leq y_p \} \cdot \mu^p (Y_i^p; x|\theta^p) - \mu^p (y_p|x; \theta^p)) \\
\times \frac{\mathcal{H}(X_i - x; h)}{f_X (x)}
\]

From our previous results we have

\[
\hat{\tau}^p (y_p|x; \theta^p) - \tau^p (y_p|x; \theta^p) = \frac{1}{n h_n^2} \sum_{i=1}^n \psi_{\tau^p} (Y_i, X_i, y_p, x, \theta^p; h_n) + \xi_{\tau^p} (y_p, x, \theta^p),
\]

where

\[
\xi_{\tau^p} (y_p, x, \theta^p) = \lambda^p (x; \theta^p) \cdot \xi_{F_{Y_p}} (y_p, x) + F_{Y_p} (y_p|x) \cdot \xi_{\lambda^p} (x, \theta^p) - \xi_{\mu^p} (y_p, x, \theta^p) \\
+ \left( F_{Y_p} (y_p|x) - F_{Y_p} (y_p|x) \right) \cdot \left( \hat{\lambda}^p (x; \theta^p) - \lambda^p (x; \theta^p) \right).
\]

Let

\[
\sup_{x \in X, \theta \in \Theta} |\lambda^p (x; \theta^p)| = \bar{\lambda}^p.
\]
For any $\delta > 0$,

$$
\begin{align*}
&\Pr \left( \sup_{(x,y)\in W, \theta \in \Theta} \left| \xi_n^p(y^p, x, \theta^p) \right| \geq \delta \right) \\
&\quad \leq \Pr \left( \sup_{(x,y)\in W, \theta \in \Theta} \left| \xi_n^V(y^p, x) \right| \geq \frac{\delta}{4A^p} \right) \\
&\quad + \Pr \left( \sup_{(x,y)\in W, \theta \in \Theta} \left| \xi_n^{\lambda}(x, \theta^p) \right| \geq \frac{\delta}{4} \right) \\
&\quad + \Pr \left( \sup_{(x,y)\in W} \left| \tilde{F}_V(y^p|x) - F_V(y^p|x) \right| \geq \frac{\sqrt{\delta}}{2} \right) \\
&\quad + \Pr \left( \sup_{x \in X, \theta^p \in \Theta} \left| \tilde{\lambda}^p(x; \theta^p) - \lambda^p(x; \theta^p) \right| \geq \frac{\sqrt{\delta}}{2} \right)
\end{align*}
$$

Let $0 < f^* < \tilde{f}$ and $D(f^*)$ be as described in (B-7), the previous expression is bounded above by

$$
\begin{align*}
&4A_1 \exp \left\{ -\sqrt{n}h_n^5 \left( A_2 \min \left\{ \frac{1}{2} \sqrt{\frac{\delta}{D(f^*)}}, \frac{\sqrt{\delta}}{D(f^*)} \right\} - A_3 \cdot h_n^M \right) \right\} \\
&+ 8A_1 \exp \left\{ -\sqrt{n}h_n^5 \left( A_2 \min \left\{ \frac{1}{2} \sqrt{\frac{\delta}{D(f^*)}}, \frac{\sqrt{\delta}}{D(f^*)} \right\} - A_3 \cdot h_n^M \right) \right\} \\
&+ 2B_1 \exp \left\{ -\sqrt{n}h_n^5 \left( B_2 \min \left\{ \frac{1}{2} \sqrt{\delta}, \frac{1}{2} \sqrt{\frac{\delta^{1/4}}{D(f^*)}}, \frac{\sqrt{\delta}}{D(f^*)} \right\} - B_3h_n^M \right) \right\}
\end{align*}
$$

Let $B = \frac{1}{2} \cdot \min \left\{ \frac{1}{\sqrt{D_{\lambda^p}}}, \frac{1}{\sqrt{\delta}}, 1, 2q, 2(\tilde{f} - f^*) \right\}$ and define $C_1 \equiv 4 \cdot B_1$, $C_2 \equiv B_2 \cdot B$, $C_3 \equiv B_3$. We have

$$
\Pr \left( \sup_{(x,y)\in W, \theta \in \Theta} \left| \xi_n^p(y^p, x, \theta^p) \right| \geq \delta \right) \\
\quad \leq C_1 \exp \left\{ -\sqrt{n}h_n^5 \left( C_2 \cdot \min \left\{ \delta^{1/2}, \delta^{1/4}, 1 \right\} - C_3 \cdot h_n^M \right) \right\}
$$

By Assumption B4 and Lemma 2.14 in Pakes and Pollard (1989), the class of functions

$$
\mathcal{G}_4 = \{ g: g(y, x) = \psi_{\tau^p}(y, x, v^p, u, \theta^p; h): (v^p, u) \in W, \theta^p \in \Theta, h > 0 \}
$$

is Euclidean with respect to the envelope $\frac{2\sqrt{\mathcal{X}}}{\epsilon} + \frac{2K \tau^p(\cdot)}{\epsilon}$. The existence of moments feature of $\tilde{\tau}^p(\cdot)$ in Assumption B4 and the results in Chapter 7 of Pollard (1990) combined with the bias conditions in B-4 imply that there exist positive constants $C'_1$, $C'_2$ and $C'_3$ such that for any $\delta > 0$,

$$
\Pr \left( \sup_{(x,y)\in W, \theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \delta \right) \\
\quad \leq C'_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^5 \left( C'_2 \cdot \delta - C'_3 \cdot h_n^M \right) \right\}.
$$
As before, if we let $0 < \int^* < \int$ be as described in (B-7)

$$
Pr \left( \sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \hat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq \delta \right) \\
\leq Pr \left( \sup_{x \in X} \left| \hat{f}_X(x) - f_X(x) \right| \geq \int - \int^* \right) \\
+ Pr \left( \sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \frac{1}{n h_n^m} \sum_{i=1}^{n} \psi_{\theta^p}(Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \delta \frac{1}{2} \right) \\
+ Pr \left( \sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \xi_n^p(y^p, x, \theta^p) \right| \geq \delta \frac{1}{2} \right).
$$

From here, putting our results together we have that for any $\delta > 0$,

$$
Pr \left( \sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \hat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq \delta \right) \\
\leq D_1 \exp \left\{ -\sqrt{n h_n^m} \left( D_2 \cdot \min \left\{ \delta, \delta^{1/2}, \delta^{1/4}, 1 \right\} - D_3 \cdot h_n^M \right) \right\},
$$

where $D_1 = 3 \cdot \max \{ A_1, C_1^2, C_1 \}$, $D_2 = \frac{1}{2} \cdot \min \{ C_2^2, C_2, 2A_2(\int - \int^*) \}$, $D_3 = \max \{ A_3, C_3, C_3^3 \}$. Our results also imply

$$
\hat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) = \frac{1}{n h_n^m} \sum_{i=1}^{n} \psi_{\theta^p}(Y_i, X_i, y^p, x, \theta^p; h_n) + \xi_n^p(y^p, x, \theta^p),
$$

where

$$
\sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \xi_n^p(y^p, x, \theta^p) \right| = O_p \left( \frac{\log(n)}{n h_n^m} \right),
$$

and

$$
\sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \hat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| = O_p \left( \frac{\log(n)}{\sqrt{n h_n^m}} \right).
$$

Let $b_n$ be the sequence used in our construction. For $n$ large enough we have $\min \{ b_n, b_n^{1/2}, b_n^{1/4}, 1 \} = b_n$ and therefore (B-10) yields

$$
Pr \left( \sup_{(x,y^p) \in W, \theta^p \in \Theta} \left| \hat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq b_n \right) \\
\leq D_1 \exp \left\{ -\sqrt{n h_n^m} \left( D_2 \cdot b_n - D_3 \cdot h_n^M \right) \right\},
$$

This concludes Step 1 of our proof.

**Step 2**

Here we use the results from Step 1 to show that

$$
\hat{\tau}_X^p(\theta^p) = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}^p(Y_i^p|X_i; \theta^p) \cdot \mathbb{I} \left\{ \tau^p(Y_i^p|X_i; \theta^p) \geq 0 \right\} \cdot f_X(X_i) + \psi_n^p(\theta^p),
$$

where

$$
\sup_{\theta^p \in \Theta} \left| \psi_n^p(\theta^p) \right| = O_p \left( n^{-1/2} \right) \quad \text{for some } \epsilon > 0.
$$
We begin by noting that we can express
\[
\hat{\varphi}_n^p(\theta^p) = \frac{1}{n} \sum_{i=1}^{n} \tilde{\varphi}(Y_i^p|X_i; \theta^p) \cdot \mathbb{I}\{\tau(Y_i^p|X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i) + \varphi_n^p(\theta^p),
\]
where
\[
|\varphi_n^p(\theta^p)| \leq \frac{1}{n} \sum_{i=1}^{n} \tilde{\varphi}(Y_i^p|X_i; \theta^p) \cdot \mathbb{I}\{-2b_n \leq \tau(Y_i^p|X_i; \theta^p) < 0\} \cdot \mathbb{I}_X(X_i)
\]
\[
+ \frac{2}{n} \sum_{i=1}^{n} \tilde{\varphi}(Y_i^p|X_i; \theta^p) \cdot \mathbb{I}\{\tau(Y_i^p|X_i; \theta^p) \leq b_n\} \cdot \mathbb{I}_X(X_i).
\]

We begin by examining \(\varphi_n^{p,2}\). Using (B-11),
\[
\sup_{(x,y)\in W, \theta^p \in \Theta} \tilde{\varphi}(y^p|x; \theta^p) = \mathcal{O}_p(1).
\]
Therefore,
\[
\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| \leq \mathcal{O}_p(1) \cdot \sup_{\theta^p \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\tilde{\varphi}(Y_i^p|X_i; \theta^p) - \tau(Y_i^p|X_i; \theta^p) \geq b_n\} \cdot \mathbb{I}_X(X_i) \right|
\]

Take any \(\alpha > 0\) and any \(\varepsilon > 0\). Then,
\[
Pr \left( n^\alpha \cdot \sup_{\theta^p \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\tilde{\varphi}(Y_i^p|X_i; \theta^p) - \tau(Y_i^p|X_i; \theta^p) \geq b_n\} \cdot \mathbb{I}_X(X_i) \right| > \varepsilon \right)
\]
\[
\leq Pr \left( \prod_{\theta^p \in \Theta} \left\{ \sup_{(x,y)\in W, \theta^p \in \Theta} \mathbb{I}\{\tilde{\varphi}(Y_i^p|X_i; \theta^p) - \tau(Y_i^p|X_i; \theta^p) \geq b_n\} \cdot \mathbb{I}_X(X_i) \neq 0 \right\} \right)
\]
\[
\leq \sum_{i=1}^{n} Pr \left( \prod_{\theta^p \in \Theta} \left\{ \sup_{(x,y)\in W, \theta^p \in \Theta} \mathbb{I}\{\tilde{\varphi}(Y_i^p|X_i; \theta^p) - \tau(Y_i^p|X_i; \theta^p) \geq b_n\} \cdot \mathbb{I}_X(X_i) \neq 0 \right\} \right)
\]
\[
\leq n \cdot Pr \left( \sup_{(x,y)\in W, \theta^p \in \Theta} \mathbb{I}\{\tilde{\varphi}(y^p|x; \theta^p) - \tau(y^p|x; \theta^p) \geq b_n\} \right)
\]
\[
\leq n \cdot D_1 \exp \left\{ -\frac{1}{2} \sqrt{\log n} \left( D_2 \cdot b_n - D_3 \cdot h_n^M \right) \right\} = D_1 \exp \left\{ -\frac{1}{2} \sqrt{\log n} \left( D_2 \cdot b_n - D_3 \cdot h_n^M + \log(n) \right) \right\} \rightarrow 0
\]

Therefore, \(\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| = o_p(n^{-\alpha})\). In particular, the following much weaker (but useful for our purposes) result holds,
\[
\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| = O_p \left( n^{-1/2-\epsilon} \right) \quad \text{for some } \epsilon > 0.
\]

We move on to \(\varphi_n^{p,1}(\theta^p)\). Note that
\[
\tilde{\varphi}(Y_i^p|X_i; \theta^p) = \sum_{j=0}^{1} \tau(Y_i^p|X_i; \theta^p)^{1-j} \cdot \left( \tilde{\varphi}(Y_i^p|X_i; \theta^p) - \tau(Y_i^p|X_i; \theta^p) \right)^j.
\]
Therefore,

$$|\varphi_{n}^{\log}(\theta^n)|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=0}^{1} \left| \tau^p(Y_i^n|X_i;\theta^n) - \tau^p(Y_i^n|X_i;\theta^n) \right| \right] \cdot \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \mathbb{I}_{\mathcal{X}}(X_i)$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=0}^{1} 2b_n \left| \tau^p(Y_i^n|X_i;\theta^n) - \tau^p(Y_i^n|X_i;\theta^n) \right| \right] \cdot \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \mathbb{I}_{\mathcal{X}}(X_i).$$

Using (B-11) we have

$$\sup_{(x,y^n)\in\mathcal{W},\theta^n\in\Theta} \left| \frac{1}{n} \sum_{j=0}^{1} 2b_n \left| \tau^p(y^n|x;\theta^n) - \tau^p(y^n|x;\theta^n) \right| \right| = \sum_{j=0}^{1} O\left( b_n^{1-j} \right) \cdot O_p \left( \left( \frac{\log(n)}{\sqrt{nh_n}} \right) \right)$$

$$= O_p\left( b_n \right),$$

where the last equality follows from the bandwidth convergence restrictions in Assumption B2 since they imply that $\frac{\log(n)}{\sqrt{nh_n}b_n} \rightarrow 0$. Therefore,

$$\sup_{\theta^n\in\Theta} |\varphi_{n}^{\log}(\theta^n)| \leq O_p\left( b_n \right) - \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \mathbb{I}_{\mathcal{X}}(X_i)$$

For a given $b > 0$ denote

$$g_{n}^{\log}(\theta^n, b) = \mathbb{I}\left\{ -b \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i).$$

And let

$$\nu_{n}^{\log}(\theta^n) = \frac{1}{n} \sum_{i=1}^{n} \left( g_{n}^{\log}(\theta^n, 2b_n) - E\left[ g_{n}^{\log}(\theta^n, 2b_n) \right] \right).$$

Let $A$ and $b$ be the constants described in Assumption B3. For large enough $n$ we have $2b_n \leq b$ and therefore we can express

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) = \nu_{n}^{\log}(\theta^n) + \xi_{n}^{\log}(\theta^n),$$

where

$$\sup_{\theta^n\in\Theta} |\xi_{n}^{\log}(\theta^n)| = 2Ab_n = O\left( b_n \right) \quad \text{and} \quad \sup_{\theta^n\in\Theta} \text{Var}\left( \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right) = O\left( b_n \right).$$

by Assumption B3. Using part (ii) of Assumption B4(ii),

$$\sup_{\theta^n\in\Theta} |\nu_{n}^{\log}(\theta^n)| = O_p\left( \sqrt{\frac{b_n}{n}} \right) = O_p\left( b_n \right).$$

Combining these results, we have

$$\sup_{\theta^n\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ -2b_n \leq \tau^p(Y_i^n|X_i;\theta^n) < 0 \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right| = O_p\left( b_n \right).$$
And therefore
\[
\sup_{\theta^p \in \Theta} |\phi_n^{p-1}(\theta^p)| \leq O(b_n) \times O_p(b_n) = O_p(b_n^2) = O_p\left(n^{-1/2-\varepsilon}\right) \text{ for some } \varepsilon > 0.
\]

Where the last line follows from the bandwidth convergence restrictions in Assumption B2. Combining the results for \(\phi_n^{p-1}\) and \(\phi_n^{p-2}\):
\[
\tilde{T}_X^p(\theta^p) = \frac{1}{n} \sum_{i=1}^{n} \tilde{T}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{I}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i) + \phi_n^{p}(\theta^p),
\]
where
\[
\sup_{\theta^p \in \Theta} |\phi_n^{p}(\theta^p)| = O_p\left(n^{-1/2-\varepsilon}\right) \text{ for some } \varepsilon > 0.
\]

**Step 3**

This is the last step in the proof. We take the results from Step 2 to show that
\[
\frac{1}{n} \sum_{i=1}^{n} \left(\tilde{T}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)\right) \cdot \mathbb{I}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i)
\]
\[
= \frac{1}{n^2} \sum_{j \neq i}^{n} g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) + \phi_n^{p-1}(\theta^p),
\]
where
\[
\sup_{\theta^p \in \Theta} |\phi_n^{p-1}(\theta^p)| = O_p\left(n^{-1/2-\varepsilon}\right) \text{ for some } \varepsilon > 0.
\]

We then examine the Hoeffding decomposition of the U-statistic described above and, using our assumptions, we obtain the result in Theorem 2. We have
\[
\frac{1}{n} \sum_{i=1}^{n} \tilde{T}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{I}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i) = \frac{1}{n} \sum_{i=1}^{n} \max_{\theta^p} \{\tau^p(Y_i^p | X_i; \theta^p), 0\} \cdot \mathbb{I}_X(X_i)
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \left(\tilde{T}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)\right) \cdot \mathbb{I}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i).
\]

Let \(\psi_{\tau^p}\) be as defined in (B-8). For any pair of observations \(i, j\) in \(1, \ldots, n\) and \(h > 0\) let
\[
g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h) = \frac{1}{h^d} \cdot \psi_{\tau^p}(Y_j, X_j, Y_i^p, X_i, \theta^p; h) \cdot \mathbb{I}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_X(X_i).
\]

Note that
\[
\sup_{\theta^p \in \Theta} \left|\frac{1}{n^2} \sum_{i=1}^{n} g_{\tau^p}(X_i, Y_i, X_i, Y_i; \theta^p, h_n)\right| = O_p\left(\frac{1}{n h_n}\right) = O_p\left(n^{-1/2-\varepsilon}\right) \text{ for some } \varepsilon > 0.
\]
Combined with (B-11), this yields

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \tau_p(Y^p_i | X_i; \theta^p) - \tau(Y^p_i | X_i; \theta^p) \right) \cdot \mathbb{1} \{ \tau_p(Y^p_i | X_i; \theta^p) \geq 0 \} \cdot \mathbb{1}_{X}(X_i)
\]

\[
= \frac{1}{n^2} \sum_{j \neq i}^{n} g_{\tau p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) + g_{\tau p}^{1}(\theta^p),
\]

where

\[
\sup_{\theta^p \in \Theta} \left| g_{\tau p}^{1}(\theta^p) \right| = O_p \left( \frac{(\log(n))^2}{nh_n} \right) + O_p \left( \frac{1}{nh_n} \right) = O_p \left( n^{-1/2-\epsilon} \right) \quad \text{for some } \epsilon > 0.
\]

We will examine the U-statistic in (B-16). Using (B-8) we can express

\[
g_{\tau p}(X_i, Y_i, X_j, Y_j; \theta^p, h) = g_{\tau p}^{a}(X_i, Y_i, X_j, Y_j; \theta^p, h) + g_{\tau p}^{b}(X_i, Y_i, X_j, Y_j; \theta^p, h) + g_{\tau p}^{c}(X_i, Y_i, X_j, Y_j; \theta^p, h),
\]

where

\[
g_{\tau p}^{a}(X_i, Y_i, X_j, Y_j; \theta^p, h) =
\]

\[
\frac{1}{h^2} \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{1} \{ Y_j^p \leq Y_i^p \} - F_{Y p}(Y_i^p | X_i)) \cdot \mathbb{1} \{ \tau_p(Y_i^p | X_i; \theta^p) \geq 0 \} \cdot \mathbb{1}_{X}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},
\]

\[
g_{\tau p}^{b}(X_i, Y_i, X_j, Y_j; \theta^p, h) =
\]

\[
\frac{1}{h^2} \cdot F_{Y p}(Y_i^p | X_i) \cdot (\eta^p(Y_j^p; X_i | \theta^p) - \lambda^p(X_i; \theta^p)) \cdot \mathbb{1} \{ \tau_p(Y_i^p | X_i; \theta^p) \geq 0 \} \cdot \mathbb{1}_{X}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},
\]

\[
g_{\tau p}^{c}(X_i, Y_i, X_j, Y_j; \theta^p, h) =
\]

\[
\frac{1}{h^2} \cdot (\mathbb{1} \{ Y_j^p \leq Y_i^p \} \cdot \eta^p(Y_j^p; X_i | \theta^p) - \mu^p(Y_i^p | X_i; \theta^p)) \cdot \mathbb{1} \{ \tau_p(Y_i^p | X_i; \theta^p) \geq 0 \} \cdot \mathbb{1}_{X}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},
\]

Let \( \gamma_p^I \), \( \gamma_p^{II} \) and \( \gamma_p^{III} \) be as defined in Assumption B1. By the smoothness conditions in Assumption B1, there exists a \( C < \infty \) such that

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| E \left[ g_{\tau p}^{a}(x, y, X, Y; \theta^p, h) \right] \right| \leq C \cdot h^M,
\]

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| E \left[ g_{\tau p}^{b}(x, y, X, Y; \theta^p, h) \right] \right| \leq C \cdot h^M,
\]

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| E \left[ g_{\tau p}^{c}(x, y, X, Y; \theta^p, h) \right] \right| \leq C \cdot h^M.
\]

And

\[
E \left[ g_{\tau p}^{a}(X, Y, x, y; \theta^p, h) \right] = \left( \gamma_p^I(y^p, x; \theta^p) - \gamma_p^{II}(x; \theta^p) \right) \cdot \mathbb{1}_{X}(x) + \zeta_p^a(y, x; \theta^p, h),
\]

\[
E \left[ g_{\tau p}^{b}(X, Y, x, y; \theta^p, h) \right] = \left( \eta^p(y^p; x | \theta^p) - \lambda^p(x; \theta^p) \right) \cdot \gamma_p^{II}(x; \theta^p) \cdot \mathbb{1}_{X}(x) + \zeta_p^b(y, x; \theta^p, h),
\]

\[
E \left[ g_{\tau p}^{c}(X, Y, x, y; \theta^p, h) \right] = \left( \gamma_p^I(y^p, x; \theta^p) \cdot \eta^p(y^p; x | \theta^p) - \gamma_p^{III}(x; \theta^p) \right) \cdot \mathbb{1}_{X}(x) + \zeta_p^c(y, x; \theta^p, h),
\]

where

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| \zeta_p^a(y, x; \theta^p, h) \right| \leq C \cdot h^M,
\]

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| \zeta_p^b(y, x; \theta^p, h) \right| \leq C \cdot h^M,
\]

\[
\sup_{(x,y) \in W, \theta^p \in \Theta} \left| \zeta_p^c(y, x; \theta^p, h) \right| \leq C \cdot h^M.
\]
In particular, this implies that

\[ \sup_{\theta^p \in \Theta} \left| E \left[ g_{r^p} (X_i, Y_i, X_j, Y_j; \theta^p, h_n) | X_i, Y_i \right] \right| \leq C \cdot h_n^M, \]

and if we define

\[
\psi_{i,\ell}^p (Y, X; \theta^p) = \left[ (\gamma_p^I (Y^p, X; \theta^p) - \gamma_p^{II} (X; \theta^p)) \cdot \lambda_p (X; \theta^p) + \left( \eta_p (Y^{-p}; X|\theta^p) - \lambda_p (X; \theta^p) \right) \cdot \gamma_p^{II} (X; \theta^p) \right] \cdot \mathbb{I}_X (X), \tag{B-17}
\]

then

\[ E \left[ g_{r^p} (X_i, Y_i, X_j, Y_j; \theta^p, h_n) | X_j, Y_j \right] = \psi_{i,\ell}^p (Y_j, X_j; \theta^p) + s_{p,n}(\theta^p), \quad \text{where } \sup_{\theta^p \in \Theta} |s_{p,n}(\theta^p)| = O_p \left( h_n^M \right) \]

Combining Assumptions B1, B2 and B4 we can show that the class of functions

\[ \mathcal{F} = \left\{ f : \mathcal{W} \times \mathcal{W} \rightarrow \mathbb{R} : f(x_1, y_1, x_2, y_2) = g_{r^p} (x_1, y_1, x_2, y_2; \theta^p, h) \text{ for some } \theta^p \in \Theta \text{ and some } h > 0 \right\} \]

is Euclidean with respect to an envelope with finite second moment. Combining this with our previous results, a Hoeffding decomposition (Serfling (1980)) and Corollary 4 in Sherman (1994) imply that (B-16) can be expressed as

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \tilde{T}^p (Y_i^p|X_i; \theta^p) - \tau^p (Y_i^p|X_i; \theta^p) \right) \cdot \mathbb{I} \left\{ \tau^p (Y_i^p|X_i; \theta^p) \geq 0 \right\} \cdot \mathbb{I}_X (X_i) = \frac{1}{n} \sum_{i=1}^{n} \psi_{i,\ell}^p (Y_i, X_i; \theta^p) + \theta_{p,n}(\theta^p),
\]

where

\[
\sup_{\theta^p \in \Theta} |\theta_{p,n}(\theta^p)| = O_p \left( \frac{\log(n)^2}{nh_n^2} \right) + O_p \left( \frac{1}{nh_n^2} \right) + O_p \left( h_n^M \right) = O_p \left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0,
\]

where the last line follows from our bandwidth convergence conditions. Going back to (B-13) and (B-14) we obtain

\[
\tilde{T}^p_{X^p}(\theta^p) = T^p_{X^p}(\theta^p) + \frac{1}{n} \sum_{i=1}^{n} \psi^p (Y_i, X_i; \theta^p) + \varepsilon_{p,n}(\theta^p),
\]

where

\[
\psi^p (Y_i, X_i; \theta^p) = \max \left\{ \tau^p (Y_i^p|X_i; \theta^p), 0 \right\} \cdot \mathbb{I}_X (X_i) - T^p_{X^p}(\theta^p) + \psi_{i,\ell}^p (Y_i, X_i; \theta^p), \tag{B-18}
\]

and

\[
\sup_{\theta^p \in \Theta} |\varepsilon_{p,n}(\theta^p)| = O_p \left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0.
\]

This concludes Step 3 and finishes the proof of Theorem 2. \qed

B.4.2 Two key properties of \( \psi^p \)

The “influence function” \( \psi^p \) has two key properties:

(i) \( E \left[ \psi^p (Y_i, X_i; \theta^p) \right] = 0 \ \forall \ \theta^p \in \Theta. \)

(ii) \( \psi^p (Y_i, X_i; \theta^p) = 0 \ \forall \ \theta^p : \tau^p (Y^p|X; \theta^p) < 0 \ \text{w.p.1.} \)
Part (ii) is obvious by inspection. To see why (i) is true we can show how it holds for each one of the summands that comprise $\psi^p$. Note first that by definition,

$$E \left[ \max \left\{ \tau^p(Y_i^p|X_i; \theta^p), 0 \right\} \cdot \mathbb{I}_X(X) - T^p_X(\theta^p) \right] = 0.$$  

We will show how each of the three summands that comprise $\psi^p$ has mean zero. We begin with the first term. Exchanging the order of integration, we have

$$E \left[ \left( \gamma^p_{\theta^p}(Y_i^p, X_i; \theta^p) - \gamma_{\theta^p}^{II}(X_i; \theta^p) \right) \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_X(X_i) \right]$$

$$= E_X \left[ E_{Y_i|X_i} \left[ E_{Y_j|X_j} \left[ \left( \mathbb{I} \{ Y_i^p \leq Y_j^p \} - F_{Y_j^p}(Y_j^p|X_j) \right) |X_i, Y_j, X_j \right] \cdot \mathbb{I} \{ \tau^p(Y_j^p|X_j; \theta^p) \geq 0 \} \right] | X_j = X_i, X_i \right]$$

$$\times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_X(X_i)$$

$$= E_X \left[ E_{Y_i|X_i} \left[ E_{Y_j|X_j} \left[ \left( F_{Y_j^p}(Y_j^p|X_j) - F_{Y_j^p}(Y_j^p|X_j) \right) |X_i, Y_j, X_j \right] \cdot \mathbb{I} \{ \tau^p(Y_j^p|X_j; \theta^p) \geq 0 \} \right] | X_j = X_i, X_i \right]$$

$$\times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_X(X_i) = 0$$

For the second term we have

$$E \left[ \left( \eta^p(Y_i^{-p}; X_i|\theta^p) - \lambda^p(X_i; \theta^p) \right) \cdot \gamma_{\theta^p}^{II}(X_i; \theta^p) \cdot \mathbb{I}_X(X_i) \right]$$

$$= E_X \left[ \left( \lambda^p(X_i; \theta^p) - \lambda^p(X_i; \theta^p) \right) \cdot \gamma_{\theta^p}^{II}(X_i; \theta^p) \cdot \mathbb{I}_X(X_i) \right] = 0,$$

where we simply used the fact that $\lambda^p(X_i; \theta^p) = E_{Y_i|X_i}[\eta^p(Y_i^{-p}; X_i|\theta^p)|X_i]$. For the third term, exchanging the order of integration we have

$$E \left[ \left( \gamma^p_{\theta^p}(Y_i^p, X_i; \theta^p) - \gamma_{\theta^p}^{II}(X_i; \theta^p) \right) \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_X(X_i) \right]$$

$$= E_X \left[ E_{Y_i|X_i} \left[ E_{Y_j|X_j} \left[ \left( \mathbb{I} \{ Y_i^p \leq Y_j^p \} \cdot \eta^p(Y_i^{-p}; X_i|\theta^p) - \mu^p(Y_j|X_j; \theta^p) \right) |X_i, Y_j, X_j \right] \cdot \mathbb{I} \{ \tau^p(Y_j^p|X_j; \theta^p) \geq 0 \} \right] | X_j = X_i, X_i \right]$$

$$\times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_X(X_i)$$

$$= E_X \left[ E_{Y_i|X_i} \left[ \left( \mu^p(Y_j|X_j; \theta^p) - \mu^p(Y_j|X_j; \theta^p) \right) \times \mathbb{I} \{ \tau^p(Y_j^p|X_j; \theta^p) \geq 0 \} \right] \right] = 0.$$

Combining these results we have $E \left[ \psi^p(Y_i, X_i; \theta^p) \right] = 0 \ \forall \ \theta^p \in \Theta$, as claimed.

### B.5 Constructing a confidence set

Let $\kappa_n$ denote any sequence of positive numbers such that $\kappa_n \to 0$ and $n' \kappa_n \to \infty$ for any $\epsilon > 0$. For each $\theta \in \Theta$ define $t_n(\theta) = \frac{\sqrt{n} \cdot T_X(\theta)}{\max \{ \kappa_n, \sigma(\theta) \}}$. By Theorem 2 and (B-3),

$$t_n(\theta) = \frac{\sqrt{n} \cdot T_X(\theta)}{\max \{ \kappa_n, \sigma(\theta) \}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Y_i, X_i; \theta) + \zeta_n(\theta).$$

By Theorem 2 and (B-3), $\sup_{\theta \in \Theta} |\kappa_n(\theta)| = o_p(1)$ since

$$\sup_{\theta \in \Theta} |\kappa_n(\theta)| = \sup_{\theta \in \Theta} \left| \frac{\sqrt{n} \cdot \varepsilon_n(\theta)}{\max \{ \kappa_n, \sigma(\theta) \}} \right| = O_p \left( \frac{1}{n^{\epsilon} \kappa_n} \right) \quad \text{for some} \ \epsilon > 0,$$
and $n^r \kappa_n \to \infty$ for any $\epsilon > 0$. Let
\[ \Theta_X^I = \{ \theta \in \Theta : \tau^p(Y^p|X; \theta^p) < 0 \text{ w.p.1 for all } p = 1, \ldots, P_0 \} \]
$\Theta_X^I$ is the collection of parameter values that satisfy our inequalities as strict inequalities w.p.1 over our inference range. Inspecting the terms that comprise $\psi(Y_i, X_i; \theta)$, we can see that $\psi(Y_i, X_i; \theta) = 0$ w.p.1 $\forall \theta \in \Theta_X^I$. On the other hand, inspecting the terms that comprise $\psi(Y_i, X_i; \theta^p)$ we can verify that $P(\psi(Y_i, X_i; \theta^p) \neq 0) > 0$ for any \( \theta \in \Theta_X^I \setminus \Theta_X \) and therefore $\sigma^2(\theta) > 0$ for any such $\theta$. Therefore,

(i) If $\theta \in \Theta \setminus \Theta_X^I$, then $T_{X_\theta}(\theta) > 0$ and therefore $t_n(\theta) \to +\infty$ w.p.1.

(ii) If $\theta \in \Theta_X^I$, then $t_n(\theta) = o_p(1)$.

(iii) If $\theta \in \Theta_X \setminus \Theta_X^I$, then $t_n(\theta) \to \mathcal{N}(0, 1)$.

t_n(\theta) is unfeasible because $\sigma^2(\theta)$ is unknown. However it can be estimated, we use $\hat{t}_n(\theta) = \frac{\sqrt{n}T_{X_\theta}(\theta)}{\max[\kappa_n, \sigma(\theta)]}$, where
\[
\hat{\psi}^p(Y_i, X_i; \theta^p) = \frac{1}{(n-1)} \sum_{j \neq i} \hat{g}_{ij}^p(X_j, Y_j, X_i, Y_i; \theta^p, h_u),
\]

\[
\hat{\psi}(Y_i, X_i; \theta^p) = \hat{\psi}^p(Y_i, X_i; \theta^p) \cdot \mathbb{I}(\hat{\psi}^p(Y_i, X_i; \theta^p) \geq -b_n) \cdot \mathbb{I}_{X_\theta}(X_i) - \hat{T}_{X_\theta}(\theta^p) + \hat{\psi}^p(Y_i, X_i; \theta^p), 
\]
\[
\hat{\sigma}^2(\theta) = \frac{1}{n} \sum_{i=1}^n \hat{\psi}(Y_i, X_i; \theta)^2.
\]
$g_{ij}$ is as described in (B-15). Under our assumptions we have $\hat{\sigma}^2(\theta) \overset{P}{\to} \sigma^2(\theta)$ for each $\theta \in \Theta$.

**Confidence set and pointwise asymptotic properties**

For a desired coverage probability $1 - \alpha$, our confidence set (CS) for $\theta_0$ is of the form
\[
CS_n(1 - \alpha) = \{ \theta \in \Theta : \hat{t}_n(\theta) \leq c_{1-\alpha} \}, \quad (B-20)
\]
where $c_{1-\alpha}$ is the Standard Normal critical value for $1 - \alpha$. By the features outlined above our CS will have correct pointwise coverage properties. Namely,

\[ \inf_{\theta \in \Theta : \theta = \theta_0} \liminf_{n \to \infty} P(\theta \in CS_n(1 - \alpha)) \geq 1 - \alpha \]

And if $\Theta_X^I \setminus \Theta_X \neq \emptyset$, then
\[ \inf_{\theta \in \Theta : \theta = \theta_0} \liminf_{n \to \infty} P(\theta \in CS_n(1 - \alpha)) = 1 - \alpha \]

Our CS will also satisfy
\[ \lim_{n \to \infty} P(\theta \in CS_n(1 - \alpha)) = 0 \quad \forall \theta \in \Theta \setminus \Theta_X^I. \]

By the design of our CS, its pointwise properties have the potential to hold uniformly (i.e., over sequences of parameter values and distributions) under appropriate assumptions about the underlying space of distributions. We describe those assumptions next and we characterize the asymptotic properties that would follow from them.
B.6 Analysis of uniform properties of our CS

Let us generalize our basic setup and assume that \( \{(Y_i^p)_{p=1}^P, X_i\} : 1 \leq i \leq n, n \geq 1 \) is a triangular array, row-wise iid with distribution \( F_n \in \mathcal{F} \). For a given \( F \in \mathcal{F} \) we will now index all the objects that depend on the distribution of the data by \( F \). Thus, we denote \( \psi(Y, X; \theta, F) \), \( \sigma^2(\theta, F) \), \( \Theta^I(F) \), and so on. We assume the following conditions about \( \mathcal{F} \).

**Assumption B5.** The space of distributions \( \mathcal{F} \) has common support and satisfies \( P_F(X \in X) \geq p > 0 \) for all \( F \in \mathcal{F} \). In addition:

(i) The conditions in Assumptions B1, B3 and B4 are satisfied by every \( F \in \mathcal{F} \).

(ii) For some \( \delta > 0 \) and \( b < \infty \),

\[
\sup_{\theta \in \Theta^I(F)} \mathbb{E}_F \left[ \frac{\left| \psi(Y, X; \theta, F) \right|^{2+\delta}}{\sigma^{2+\delta}(\theta, F)} \right] \leq b
\]

**B.6.1 Coverage properties**

Part (i) of Assumption B5 is meant to ensure that the linear representation in (B-3) holds uniformly over \( \mathcal{F} \). Part (ii) is sufficient to ensure the Lindeberg condition,

\[
\lim_{\lambda \to \infty} \sup_{\theta \in \Theta^I(F)} \mathbb{E}_F \left[ \frac{\left| \psi(Y, X; \theta, F) \right|^2}{\sigma^2(\theta, F)} \right] \cdot \mathbb{P} \left\{ \left| \psi(Y, X; \theta, F) \right| > \lambda \right\} = 0.
\]

To see why, note that for any \( \tilde{\lambda} > 0 \) and \( \delta > 0 \), \( \tilde{\lambda}^2 \cdot \psi(Y, X; \theta, F)^2 \cdot \mathbb{P} \left\{ \left| \psi(Y, X; \theta, F) \right| > \tilde{\lambda} \right\} \leq \left| \psi(Y, X; \theta, F) \right|^{2+\delta} \).

Therefore \( E \left[ \psi(Y, X; \theta, F)^2 \cdot \mathbb{P} \left\{ \left| \psi(Y, X; \theta, F) \right| > \tilde{\lambda} \right\} \right] \leq E \left[ \sup_{\theta \in \Theta^I(F)} \left| \psi(Y, X; \theta, F) \right|^{2+\delta} \right] \). The Lindeberg condition follows by using the \( \delta \) described in Assumption B5, letting \( \lambda = \sigma(\theta, F) \) and dividing both sides of the inequality by \( \sigma^2(\theta, F) \). Combined with the kernel and bandwidth conditions in Assumption B2, part (i) and the Lindeberg condition implied by part (ii) of Assumption B5 imply that for any sequence \( (F_n, \theta_n) \) such that \( F_n \in \mathcal{F} \) and \( \theta_n \in \Theta^I(F_n) \setminus \Theta^I_X(F_n) \),

\[
\frac{\sqrt{n} \cdot \hat{T}_X(\theta_n)}{\sigma(\theta_n, F_n)} \xrightarrow{d} \mathcal{N}(0, 1).
\]

And for any sequence \( (F_n, \theta_n) \) such that \( F_n \in \mathcal{F} \) and \( \theta_n \in \Theta^I_X(F_n) \),

\[
\frac{\sqrt{n} \cdot \hat{T}_X(\theta_n)}{\max \{ \kappa_n, \sigma(\theta_n, F_n) \}} \xrightarrow{p} 0.
\]

Let \( t_n(\theta) = \frac{\sqrt{n} \hat{T}_F(\theta)}{\max \{ \kappa_n, \sigma(\theta, F_n) \}} \) denote the unfeasible test-statistic that uses \( \sigma(\theta, F_n) \) instead of \( \hat{\sigma}(\theta) \). Combined, parts (i) and (ii) of Assumption B5 would yield

\[
\lim_{n \to \infty} \inf_{\theta \in \Theta^I(\theta) = \theta_0} \inf_{F \in \mathcal{F}} P_F(t_n(\theta) \leq c_{1-\alpha}) \geq 1 - \alpha, \quad (B-21)
\]

57
with
\[
\liminf_{n \to \infty} \inf_{\theta \in \Theta}\inf_{\theta = \theta_0} P_F (t_n(\theta) \leq c_{1-\alpha}) = 1 - \alpha \quad \text{if} \quad \Theta_X(F) \cap \overline{\Theta}_X^i(F) \neq \emptyset \quad \text{for some} \quad F \in \mathcal{F}.
\]

Of course, our CS is based on \(\hat{\theta}(\theta) = \sqrt{n} T^2_X(\theta)\), where \(\hat{\theta}^2(\theta)\) is estimated as described in (B-19). We need to endow \(\mathcal{F}\) with conditions that ensure that the necessary Laws of Large Numbers for triangular arrays hold in a way that ensures that \(\hat{\theta}^2(\theta_n) - \sigma^2(\theta_n, F_n) \xrightarrow{p} 0\) over sequences \((F_n, \theta_n) \in \mathcal{F} \times \Theta\). For this we can look at the type of sufficient conditions found in Romano (2004, Lemma 2). To this end we impose the following conditions.

**Assumption B6.** Let \(\psi_{\psi_{Fp}}, \psi_{\lambda p}, \psi_{\mu p}\) and \(\psi_{\tau p}\) be as described in (B-6), (B-8) and (B-15). Then, for some \(\delta > 0\) and \(b < \infty\) the following holds for each \(p = 1, \ldots, P\),

\[
\sup_{F \in \mathcal{F}} E_F \left[ \frac{1}{h^q} \psi_{\psi_{Fp}} (Y_i^p, X_i, y^p, x; h, F) - E \left[ \frac{1}{h^q} \psi_{\psi_{Fp}} (Y_i^p, X_i, y^p, x; F) \right] \right]^{1+\delta} \leq b,
\]

\[
\sup_{F \in \mathcal{F}} E_F \left[ \frac{1}{h^q} \psi_{\lambda p} (Y_i^p, X_i, x, \theta^p; h) - E \left[ \frac{1}{h^q} \psi_{\lambda p} (Y_i^p, X_i, x, \theta^p; h) \right] \right]^{1+\delta} \leq b,
\]

\[
\sup_{F \in \mathcal{F}} E_F \left[ \frac{1}{h^q} \psi_{\mu p} (Y_i, X_i, y^p, x, \theta^p; h) - E \left[ \frac{1}{h^q} \psi_{\mu p} (Y_i, X_i, y^p, x, \theta^p; h) \right] \right]^{1+\delta} \leq b,
\]

\[
\sup_{F \in \mathcal{F}} E_F \left[ \frac{1}{h^q} \psi_{\tau p} (Y, X, y^p, x, \theta^p; h, F) - E \left[ \frac{1}{h^q} \psi_{\tau p} (Y, X, y^p, x, \theta^p; h) \right] \right]^{1+\delta} \leq b,
\]

Assumption B6 is sufficient to satisfy the conditions for the Law of Large Numbers for triangular arrays in Romano (2004, Lemma 2). Combined with Assumption B5, the smoothness conditions in Assumption B1 and the linear representation in (B-9), Assumption B6 and Romano (2004, Lemma 2) can be used to show that for any sequence \((F_n, \theta_n) \in \mathcal{F} \times \Theta\),

\[
\hat{\theta}^2(\theta_n) - \sigma^2(\theta_n, F_n) \xrightarrow{p} 0.
\]

Combining Assumptions B5 and B6, our confidence sets would inherit the coverage properties in (B-21). Namely,

\[
\liminf_{n \to \infty} \inf_{\theta \in \Theta}\inf_{\theta = \theta_0} P_F (\theta \in CS_n(1 - \alpha)) = 1 - \alpha,
\]

with

\[
\liminf_{n \to \infty} \inf_{\theta \in \Theta}\inf_{\theta = \theta_0} P_F (\theta \in CS_n(1 - \alpha)) = 1 - \alpha \quad \text{if} \quad \Theta_X^i(F) \cap \overline{\Theta}_X^i(F) \neq \emptyset \quad \text{for some} \quad F \in \mathcal{F}.
\]
B.6.2 Power properties

The linear representation in (B-3) facilitates the study of the power features of our procedure. Take a sequence $(F_n, \theta_n)$ such that $F_n \in \mathcal{F}$ and $\theta_n \in \Theta \setminus \Theta_X^I(F_n)$. By Assumption B5(ii), for any $c$ we have

$$\lim_{n \to \infty} P_{F_n}\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(Y_i, X_i; \theta_n, F_n)}{\sigma(\theta_n, F_n)} > c \right) = 1 - \Phi(c).$$

The key to the power properties of our test over such a sequence is the behavior of $\sigma^2(\theta_n, F_n) = \text{Var}_{F_n}(\psi(Y, X; \theta_n, F_n))$. Recall that $T_X(\theta_n, F_n) = \sum_{p=1}^{P} E[\max\{\tau^p(Y^p|X; \theta_n, F_n), 0\} \cdot I_X(X)]$. By Assumption B5, $\lim_{n \to \infty} P_{F_n}(X \in \mathcal{X}) \geq p > 0$ for any sequence $F_n \in \mathcal{F}$. Therefore we have $T_X(\theta_n, F_n) \to 0$ if and only if $P_{F_n}(\tau^p(Y^p|X; \theta_n, F_n) > 0|X \in \mathcal{X}) \to 0$ for each $p = 1, \ldots, P$. If we inspect the structure of $\psi(Y, X; \theta_n, F_n)$ we will see that the key will be the behavior of the sequence

$$P_{F_n}(\tau^p(Y^p|X; \theta_n, F_n) = 0 \text{ for some } p = 1, \ldots, P|X \in \mathcal{X}) \equiv \Delta_X(\theta_n, F_n).$$

$\Delta_X(\theta_n, F_n)$ is the probability that the inequalities are binding for some $p$ over our inference range. We have the following:

(i) If $T_X(\theta_n, F_n) \to 0$ and $\Delta_X(\theta_n, F_n) \to 0$, then $\sigma(\theta_n, F_n) \to 0$.

(ii) If $T_X(\theta_n, F_n) \to 0$ but $\Delta_X(\theta_n, F_n) \not\to 0$, then $\sigma(\theta_n, F_n) \not\to 0$.

(iii) If $T_X(\theta_n, F_n) \not\to 0$, then $\sigma(\theta_n, F_n) \not\to 0$.

The asymptotic power of our approach will be determined by the behavior of the following two sequences,

$$s_{1,n}(\theta_n, F_n) = \max\{\kappa_n, \sigma(\theta_n, F_n)\} / \sigma(\theta_n, F_n), \quad \text{and} \quad s_{2,n}(\theta_n, F_n) = \sqrt{n} \cdot T_X(\theta_n, F_n) / \max\{\kappa_n, \sigma(\theta_n, F_n)\}.$$

Suppose $s_{1,n}(\theta_n, F_n) \to s_1$ and $s_{2,n}(\theta_n, F_n) \to s_2$. Note that $s_1 \geq 1$ by construction. If Assumptions B5 and B6 hold, the conditions in Romano (2004, Theorem 5) are satisfied and we can use this to show that

$$\lim_{n \to \infty} P_{F_n}(\hat{\theta}(\theta_n) > c_{1-\alpha}) = 1 - \Phi(s_1 \cdot (c_{1-\alpha} - s_2)).$$

From here we conclude that our procedure will have asymptotic power of 1 if either:

(a) $s_2 = \infty$: This includes as a special case any sequence such that $T_X(\theta_n, F_n) = O(n^{-\alpha})$ for some $\alpha < 1/2$. In this case we would have $s_{2,n}(\theta_n, F_n) = O\left(\frac{n^{1/2-\alpha}}{\kappa_n}\right) \to \infty$ by the convergence restrictions of $\kappa_n$.

(b) $s_1 = \infty$ and $s_2 > c_{1-\alpha}$: Firstly, our discussion above implies that $s_1 = \infty$ can occur only if $\Delta_X(\theta_n, F_n) \to 0$ and $T_X(\theta_n, F_n) \to 0$. The additional condition $s_2 > c_{1-\alpha}$ forbids $T_X(\theta_n, F_n)$ from converging to zero “too fast”.

Part (a) shows that our procedure will have asymptotic power of 1 whenever $T_X(\theta_n, F_n) = O(n^{-\alpha})$ for some $\alpha < 1/2$. Suppose $T_X(\theta_n, F_n) = O\left(n^{-\alpha}\right)$ for some $\alpha > 1/2$. Then we have $s_2 = 0$ by the bandwidth convergence restrictions of $\kappa_n$. In this case our approach will have asymptotic power of zero if $s_1 = \infty$ (i.e., if $\sigma(\theta_n, F_n)/\kappa_n \to 0$). On the other hand if $\sigma(\theta_n, F_n)/\kappa_n \to \infty$ then the asymptotic power will be $\alpha$. This will be the case, for
example, for any sequence such that \( T_X(\theta_n, F_n) = O(n^{-\alpha}) \) for some \( \alpha > 1/2 \) but \( \lim_{n \to \infty} \Delta_X(\theta_n, F_n) > 0 \). On the other hand, our asymptotic power would be zero if \( \lim_{n \to \infty} \Delta_X(\theta_n, F_n) = 0 \). If \( \sigma(\theta_n, F_n) \propto \kappa_n \), the power will be bounded between zero and \( \alpha \). Finally, suppose \( T_X(\theta_n, F_n) = O\left(n^{-1/2}\right) \). Our procedure will have asymptotic power of 1 for any such sequence as long as \( \lim_{n \to \infty} \Delta_X(\theta_n, F_n) = 0 \), as this would yield \( s_2 = \infty \). If \( \lim_{n \to \infty} \Delta_X(\theta_n, F_n) \neq 0 \), then \( s_2 < \infty \). In this case our asymptotic power will be 1 if \( s_2 > c_{1-\alpha} \) but it will be zero if \( s_2 < c_{1-\alpha} \). Thus, our asymptotic power for any sequence \( T_X(\theta_n, F_n) = O\left(n^{-1/2}\right) \) will be determined by the limit of the sequence \( \Delta_X(\theta_n, F_n) \). Note that –as one should expect– choosing the maximum rate of convergence for \( \kappa_n \) that is consistent with our assumptions is beneficial for power. Given our bandwidth convergence restrictions, this rate is \( \kappa_n \propto \log(n) \). Our analysis shows the power advantages of our approach vis-a-vis using a test-statistic based on a least-favorable configuration, as this would be based on normalizing our test statistic by a standard deviation that does not converge to zero when \( T_X(\theta_n, F_n) \to 0 \).

### B.7 Kernels and bandwidths used in our empirical application

Our covariate vector \( X \) includes \( q = 8 \) continuous random variables. The smallest kernel order \( M \) compatible with Assumption B2 is \( M = 2 \cdot q + 1 = 17 \). We employed a multiplicative kernel \( K(\psi_1, \ldots, \psi_n) = k(\psi_1) \cdot k(\psi_2) \cdots k(\psi_n) \), where each \( k(\cdot) \) is a bias-reducing Biweight-type kernel of order \( M = 18 \) of the form,

\[
k(u) = \sum_{j=1}^{9} c_j \cdot (1 - u^2)^{2j} \cdot I\{|u| \leq s\},
\]

where \( c_1, \ldots, c_5 \) were chosen to satisfy the restriction of a bias-reducing kernel of order 18. As in Aradillas-López, Gandhi, and Quint (2013) we set \( s = 30 \). Following the guidelines in Assumption B2 we employed a bandwidth of the form \( h_n = c \cdot \hat{\sigma}(X) \cdot n^{-\alpha_b} \) (note that each \( X \) has its own bandwidth), where \( \alpha_b = \frac{1}{M} + \tau \) and \( \tau = 10^{-5} \).

As a guidance to select the constant ‘\( c \)’ we used the “rule of thumb” formula (Silverman (1986)), using the Normal distribution as the reference distribution. We set

\[
c = 2 \cdot \left( \frac{\pi^{1/2} (M!)^3 \cdot R_k}{(2M) \cdot (2M)! \cdot (k^2_M)} \right)^{\frac{3}{2M+1}}, \text{ where } R_k = \int_{-1}^{1} k^2(u)du, \quad k_M = \int_{-1}^{1} u^M k(u)du.
\]

This yielded \( c \approx 0.2 \) and therefore \( h_n \approx 0.16 \cdot \hat{\sigma}(X) \) (for our sample size \( n = 954 \)). Let \( \Omega = \max_{\theta \in \Theta} \left| \hat{\sigma}(\theta) \right| \). We used \( b_n = c_b \cdot \Omega \cdot n^{-\alpha_b} \), where \( \alpha_b = \frac{1}{3} + \tau \) and \( \kappa_n = c_K \cdot \Omega \cdot \log(n)^{-1} \) with \( c_b = 10^{-6} \) and \( c_K = 10^{-8} \). We chose these tuning parameters proportional to \( \Omega \) to ensure our procedure is scale-invariant. These bandwidth choices satisfy Assumption B2. For our sample size \( n = 954 \) this resulted in \( b_n \approx 10^{-5} \) and \( \kappa_n \approx 10^{-7} \). The inference range used was

\[
\mathcal{X} = \left\{ x : \hat{f}_X(x) \geq f_X^{(0.15)}, \quad POP < 5 \text{ Million} \right\},
\]

where \( f_X^{(0.15)} \) denotes the estimated 15th percentile of the density \( \hat{f}_X \). Our main findings were qualitatively robust to moderate changes in these tuning parameters. Our results were qualitatively robust to moderate changes in the constants \( c, c_b, c_K, \alpha_b \) and \( \alpha_b \) used to construct our bandwidths.
References


