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

March 26, 2014

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/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 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 non-parametrically 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/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 the parameters underlying each player's strategic index. We are the first to exploit multiplicity of equilibria for the purposes of inference of parameters that govern strategic interactions.

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 a 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 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.[1] 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 anti-competitive 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 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 6 concludes. All proofs are included in the appendix.

[1] Aradillas-Lopez (2011) also focuses on rich strategy spaces but the goal there is to answer a different question than the one posed here.

2 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 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 $Y^p \in \mathcal{A}^p$. The strategy space $\mathcal{A}^p$ 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 $\mathcal{A}^p$. The only restriction is that it must possess a natural order. We let $\mathcal{A}^{-p} = \prod^{q \neq p} \mathcal{A}^q$ denote the action space of p's opponents. We use lower case y^p to denote a potential action (in $\mathcal{A}^p$) for p and $y^{-p} \equiv (y^q)^{q \neq p}$ to denote a potential action profile (in $\mathcal{A}^{-p}$) for p's opponents. We use upper case letters (Y^p 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

$$\nu^p(y^p, y^{-p}; \xi^p), \tag{1}$$

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

$$\xi^p = (X, \varepsilon^p),$$

where X is observed by the econometrician and ε^p is not. The dimension of ε^p is left unspecified and we allow ε^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 will assume that payoff functions can be expressed in the following way.

Assumption 1. ***(Generic expression of payoff functions)***
ν^p can be expressed as follows,

$$\nu^p(y^p, y^{-p}; \xi^p) = \nu^{p,a}(y^p; \xi^p) - \nu^{p,b}(y^p; \xi^p) \cdot \eta^p(y^{-p}; X), \tag{2}$$

where $\nu^{p,b}$ and η^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 η^p is that it depends on ξ^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 will 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 $\mathcal{A}^{-p}$. For any set of beliefs $\sigma^{-p} : \mathcal{A}^{-p} \longrightarrow [0,1]$, the associated expected utility for p of choosing $Y^p = y^p$ is

$$\begin{aligned}
\overline{\nu}^p_\sigma(y^p;\xi^p) &= \sum_{y^{-p}\in\mathcal{A}^{-p}} \sigma^{-p}(y^{-p})\cdot\nu^p(y^p,y^{-p};\xi^p) \\
&= \nu^{p,a}(y^p;\xi^p) - \nu^{p,b}(y^p;\xi^p)\cdot\overline{\eta}^p_\sigma(X), \quad \text{where} \\
\overline{\eta}^p_\sigma(X) &= \sum_{y^{-p}\in\mathcal{A}^{-p}} \sigma^{-p}(y^{-p})\cdot\eta^p(y^{-p};X).
\end{aligned} \tag{3}$$

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 η^p.

Assumption 2. ***(Marginal benefit of*** Y^p ***is nonincreasing in*** $\boldsymbol{\eta^p}$***)*** *With probability one in* ξ^p*, the function* $\nu^{p,b}(\cdot;\xi^p)$ *is nondecreasing over* $\mathcal{A}^p$*. That is, for any* $v > u$ *in* $\mathcal{A}^p$ *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 $\mathcal{A}^p$. Take any pair of beliefs σ^{-p} and $\sigma^{-p'}$. Then,

$$[\overline{\nu}^p_\sigma(v;\xi^p) - \overline{\nu}^p_\sigma(u;\xi^p)] - [\overline{\nu}^p_{\sigma'}(v;\xi^p) - \overline{\nu}^p_{\sigma'}(u;\xi^p)] = [\overline{\eta}^p_{\sigma'}(X) - \overline{\eta}^p_\sigma(X)]\cdot\left[\nu^{p,b}(v;\xi^p) - \nu^{p,b}(u;\xi^p)\right].$$

Therefore by Assumption 2,

$$\overline{\eta}^p_\sigma(X) \geq \overline{\eta'}^p_\sigma(X) \quad\Longrightarrow\quad \overline{\nu}^p_\sigma(v;\xi^p) - \overline{\nu}^p_\sigma(u;\xi^p) \leq \overline{\nu}^p_{\sigma'}(v;\xi^p) - \overline{\nu}^p_{\sigma'}(u;\xi^p) \quad \forall\, u < v \in \mathcal{A}^p \tag{4}$$

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)). In this paper we will not make any assumptions[2] regarding how payoffs shift with specific elements in ξ^p.

[2] If economic theory provides ex-ante information about how payoffs should shift with some specific elements in ξ^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 η^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 η^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

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

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

$$\mathcal{Q}^p = \phi^p(\varepsilon^p)\cdot\sum_{q=1}^{P} a^{p,q}(X)\cdot\Big(\mathcal{P}^q + \zeta^p(\xi^p)\Big),\ \text{for } p = 1,\ldots,P.$$

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

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

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 it need not have to display increasing marginal costs). Profit functions are of the form,

$$\pi^p\left(\mathcal{Q}^p, \mathcal{Q}^{-p}; \xi\right) = \left(\frac{1}{\phi^p(\varepsilon^p)} \sum_{q=1}^{P} b^{p,q}(X) \cdot \mathcal{Q}^q - \zeta^p(\xi^p)\right) \cdot \mathcal{Q}^p - C^p(\mathcal{Q}^p; \xi^p).$$

In a Cournot model firms compete in quantities, so $Y^p = \mathcal{Q}^p$. 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

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

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 $\left[A(X)^{-1}\right]_{p,q} = b^{p,q}(X)$, the strategic indices η^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,

$$\log\left(\mathcal{Q}^p\right) = \sum_{q=1}^{P} d^{p,q}(\xi^p) \cdot \log\left(\mathcal{P}^q\right) + f^p(\xi^p), \text{ for } p = 1, \ldots, P.$$

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 ε^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

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

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 $\left[D(X)^{-1}\right]_{p,q} \equiv r^{p,q}(X)$. Inverting the demand system we obtain the following inverse demands,

$$\mathcal{P}^p = e^{-\lambda^p(\xi^p)} \cdot \prod_{q=1}^{P} \left(\mathcal{Q}^q\right)^{r^{p,q}(X)}, \text{ for } p = 1, \ldots, P.$$

Profit functions are now of the form

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

Define $\nu^{p,a}(y^p;\xi^p) = -C^p(y^p;\xi^p)$. For $\nu^{p,b}$ and η^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

$$\begin{aligned}
\nu^{p,b}(y^p;\xi^p) &= e^{-\lambda^p(\xi^p)} \cdot \Big(\mathbb{1}\left\{r^{p,p}(X) \geq -1\right\} - \mathbb{1}\left\{r^{p,p}(X) < -1\right\}\Big) \cdot \left(y^p\right)^{r^{p,p}(X)}, \\
\eta^p(y^{-p};X) &= \Big(\mathbb{1}\left\{r^{p,p}(X) \geq -1\right\} - \mathbb{1}\left\{r^{p,p}(X) < -1\right\}\Big) \cdot \prod_{q \neq p} \left(y^q\right)^{r^{p,q}(X)}
\end{aligned}$$

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 η^p

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

2.5 Information and behavior

We now introduce a key assumption on the information set of the players. Before introducing our assumption it is useful to take a step back and ask why we are adopting an incomplete information perspective. We adopt an incomplete information perspective because it generically will allow 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.[3] 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.[4] 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.

Assumption 3. ***(Independent private shocks)***

X is perfectly observed by all players, but ε^p is only privately observed by p. We assume that each ε^p

[3]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).

[4]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.

is independent of ε^{-p} conditional on X. The true distribution of $(X, (\varepsilon^p)_{p=1}^P)$ is common knowledge among the players, as are the functional forms of payoff functions $(\nu^p)_{p=1}^P$. Thus, the only source of incomplete information for p is the realization of ε^{-p}.

The dimension of ε^p is left unspecified and we allow ε^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 $\xi^p \subseteq X$. Thus, a game of complete information would be a special case of our setting as long as $\xi^p \subseteq X$ for each p. In this case the only source of unobserved heterogeneity would be the equilibrium selection mechanism.

2.5.1 Bayesian Nash equilibrium (BNE) behavior

We will 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 $\{\sigma_*^p(\cdot|X) : \mathcal{A}^p \longrightarrow [0,1]\}_{p=1}^P \equiv \sigma_*(X)$ with corresponding expected utility functions

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

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

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

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

$$Y^p \in \underset{y \in \mathcal{A}^p}{\mathit{argmax}}\, \overline{\nu}^p_{\sigma_*}(y;\xi^p) \ \textit{ for some BNE } \sigma_*(X).$$

*For a given realization of payoff shifters, multiple BNE may exist and we leave the underlying selection mechanism $\mathcal{S}$ unspecified except for the assumption that it always picks a BNE $\sigma_*(X)$ such that the resulting expected payoff function $\overline{\nu}^p_{\sigma_*}(\,\cdot\,;\xi^p)$ 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.[5]

3 Implications of Assumptions 1-4

3.1 Properties of players' best-responses

As we stated above, we will 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 σ^{-p} and $\sigma^{-p'}$ denote any pair of beliefs that produce unique expected-payoff maximizing choices for p given the realization of ξ^p, and let $y^p_\sigma(\xi^p)$ and $y^p_{\sigma'}(\xi^p)$ denote the corresponding optimal choices. If Assumptions 1-2 hold, then w.p.1 we have,

$$If\ \overline{\eta}^p_\sigma(X) \geq \overline{\eta}^p_{\sigma'}(X), \text{ then } \mathbb{1}\big\{y^p_\sigma(\xi^p) \leq y^p\big\} \geq \mathbb{1}\big\{y^p_{\sigma'}(\xi^p) \leq y^p\big\} \quad \forall\, y^p \in \mathcal{A}^p.$$

Proof: In Appendix A. $\square$

3.2 Main result

Let σ_{*j} and σ_{*k} denote any pair of existing BNE that the selection mechanism $\mathcal{S}$ could choose with positive probability. By Result 1, w.p.1 we must have,

$$If\ \overline{\eta}^p_{\sigma_{*j}}(X) \geq \overline{\eta}^p_{\sigma_{*k}}(X), \text{ then } \mathbb{1}\big\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\big\} \geq \mathbb{1}\big\{y^p_{\sigma_{*k}}(\xi^p) \leq y^p\big\} \quad \forall\, y^p \in \mathcal{A}^p.$$

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*

$$E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p};X)\big|X\big] \geq E\big[\mathbb{1}\{Y^p \leq y^p\}\big|X\big] \cdot E\big[\eta^p(Y^{-p};X)\big|X\big] \quad \forall\, y^p. \tag{5}$$

[5]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).

Proof: In Appendix A. $\square$

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 $Y^p \perp Y^{-p}|X$ and therefore any measurable function $g(Y^{-p}; X)$ should satisfy Theorem 1 as an equality. Therefore Theorem 1 provides identification power for η^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. The identification power of multiple equilibria has been used before in econometric work. A notable recent example is De Paula and Tang (2012), which is also the existing paper more closely related to ours. De Paula and Tang (2012) consider binary choice games and focus on identifying the direction (the sign) of strategic interaction under the symmetry assumption that each player cares *equally* about the decisions of each opponent. We will focus on a more general object –the strategic interaction index η^p– which not only captures the direction of strategic interaction, but also the relative effects of opponents' actions on p's payoff. Instead of maintaining symmetry, we will ultimately test for it by estimating the relevant primitive of the game.

Remark 2. Two important results follow from Theorem 1.

(i) **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 some 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.

(ii) **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 η^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

Suppose the strategic interaction index η^p belongs 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 $\theta = \bigcup_{p=1}^{P} \theta^p$ and let Θ denote the parameter space. The true value of θ will be denoted by θ_0. For given y^p, x and θ^p define

$$\begin{aligned}
F_{Y^p}(y^p|x) &= E_{Y^p|X}\left[\mathbb{1}\left\{Y^p \leq y^p\right\}\big|X = x\right], \\
\lambda^p(x;\theta^p) &= E_{Y^{-p}|X}\left[\eta^p(Y^{-p}; x|\theta^p)\big|X = x\right], \\
\mu^p(y^p|x;\theta^p) &= E_{Y|X}\left[\mathbb{1}\left\{Y^p \leq y^p\right\} \cdot \eta^p(Y^{-p}; x|\theta^p)\big|X = x\right], \\
\tau^p(y^p|x;\theta^p) &= F_{Y^p}(y^p|x) \cdot \lambda^p(x;\theta^p) - \mu^p(y^p|x;\theta^p).
\end{aligned}$$

Theorem 1 predicts that for each p,

$$\tau^p(y^p|X;\theta_0^p) \leq 0 \ \text{ w.p.1 in } X \ \ \forall\, y^p \in \mathcal{A}^p.$$

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

$$Pr\left(\tau^p(Y^p|X;\theta_0^p) \leq 0\right) = 1. \tag{6}$$

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

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

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 . Suppose $m(W;\theta)$ is a vector of *known* functions such that $E\left[m(W;\theta_0)\big|X\right] \leq 0$ w.p.1. Let $\mathcal{G}$ be a space of measurable, nonnegative functions of X. Then the previous inequality implies that we must have $E\left[m(W;\theta_0)\cdot g(X)\right] \leq 0$ 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 $E\left[m(W;\theta_0)\big|X\right] \leq 0$ 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 θ^p satisfies our restrictions for player p over a range $(y^p, x) \in \mathcal{W}$ by using at a test-statistic of the form $\widehat{\nu}^p_\alpha(\theta^p) = \underset{(y^p,x)\in\mathcal{W}}{inf}\left[(-\widehat{\tau}^p(y^p|x;\theta^p)) + \widehat{k}(\alpha)\cdot\widehat{\sigma}^p(y^p|x;\theta^p)\right]$, where $\widehat{\sigma}^p$ is an estimator of the standard error of $\widehat{\tau}^p$ and $\widehat{k}(\alpha)$ is a critical value based on the α^{th} quantile of a particular process. We would reject the inequalities for θ^p if $\widehat{\nu}^p_\alpha(\theta^p) < 0$ 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 η^p is such that $\tau^p(y|x;\theta^p)$ is nonseparable in θ^p. In such cases the critical value $\widehat{k}(\alpha)$ would also depend on θ^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 θ^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 θ^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 and we will focus 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 prespecified set such that

$$\mathcal{X} \cap \text{Supp}(X^c) \subset \text{int}\left(\text{Supp}(X^c)\right).$$

We will maintain the assumption that $f_X(x) \geq \underline{f} > 0$ for all $x \in \mathcal{X}$. Let $\mathbb{I}_{\mathcal{X}}(x)$ denote a "trimming" function such that $\mathbb{I}_{\mathcal{X}}(x) = 0$ if $x \notin \mathcal{X}$ and $\mathbb{I}_{\mathcal{X}}(x) > 0$ otherwise. Let

$$T^p_{\mathcal{X}}(\theta^p) = E_{Y^p,X}\Big[\max\{\tau^p(Y^p|X;\theta^p),0\} \cdot \mathbb{I}_{\mathcal{X}}(X)\Big], \quad T_{\mathcal{X}}(\theta) = \sum_{p=1}^{P} T^p_{\mathcal{X}}(\theta^p). \tag{8}$$

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 Θ^I,

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

Note that $\Theta^I \subseteq \Theta^I_{\mathcal{X}}$. 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.

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 $\left((Y_i^p)_{p=1}^P, X_i\right)_{i=1}^n$ produced by a model satisfying our assumptions. We replace the objects in (B-1) with estimators of the form

$$\widehat{T}^p_{\mathcal{X}}(\theta^p) = \frac{1}{n}\sum_{i=1}^n \widehat{\tau}^p(Y_i^p|X_i;\theta^p)\cdot \mathbb{1}\left\{\widehat{\tau}^p(Y_i^p|X_i;\theta^p) \geq -b_n\right\}\cdot \mathbb{I}_{\mathcal{X}}(X_i), \quad \widehat{T}_{\mathcal{X}}(\theta) = \sum_{p=1}^P \widehat{T}^p_{\mathcal{X}}(\theta^p)$$

where $b_n \longrightarrow 0$ 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 $\widehat{\tau}^p$ we use kernel-based estimators.

In the econometric appendix we describe conditions under which

$$\widehat{T}_{\mathcal{X}}(\theta) = T_{\mathcal{X}}(\theta) + \frac{1}{n}\sum_{i=1}^n \psi(Y_i, X_i;\theta) + \varepsilon_n(\theta),$$

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

The "influence function" ψ can be expressed as

$$\psi(Y_i, X_i;\theta) = \sum_{p=1}^P\Big(\max\left\{\tau^p(Y_i^p|X_i;\theta_p), 0\right\}\cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}(\theta^p)\Big) + \sum_{p=1}^P \psi_U^p(Y_i, X_i;\theta^p).$$

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

(ii) Let

$$\overline{\Theta}^I_{\mathcal{X}} = \{\theta\in\Theta : \tau^p(Y^p|X;\theta^p) < 0 \ \text{ w.p.1.} \, \forall\, p = 1,\ldots,P.\}$$

Then $\psi(Y_i, X_i;\theta) = 0$ w.p.1 $\forall\, \theta \in \overline{\Theta}^I_{\mathcal{X}}$.

$\overline{\Theta}^{I}_{\mathcal{X}}$ is the collection of parameter values that satisfy our inequalities as *strict inequalities* w.p.1 over our inference range. Let $\sigma^2(\theta) = Var\left(\psi(Y_i, X_i; \theta)\right)$. Based on the properties outlined above, we will have

$$\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta) = \sqrt{n}T_{\mathcal{X}}(\theta) + V_n(\theta) + \xi_n(\theta),$$

where $V_n(\theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\theta)\right)$ and $\underset{\theta\in\Theta}{sup}\left|\xi_n(\theta)\right| = o_p\left(n^{-\epsilon}\right)$ for some $\epsilon > 0$. Given these features, our statistic will be of the form

$$\widehat{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\max\left\{\kappa_n, \widehat{\sigma}(\theta)\right\}},$$

where $\widehat{\sigma}^2(\theta)$ is an estimator of $\sigma^2(\theta)$ and κ_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 $\underset{\theta\in\overline{\Theta}^{I}_{\mathcal{X}}}{sup}\left|\widehat{T}_{\mathcal{X}}(\theta)\right| = O_p\left(n^{-1/2-\epsilon}\right)$ for some $\epsilon > 0$. The use of κ_n allows our statistic to satisfy $\underset{\theta\in\overline{\Theta}^{I}_{\mathcal{X}}}{sup}\left|\sqrt{n}\cdot\widehat{t}_n(\theta)\right| = o_p(1)$.

For a desired coverage probability $1-\alpha$, our CS for θ_0 is of the form

$$CS_n\left(1-\alpha\right) = \left\{\theta \in \Theta\colon \widehat{t}_n(\theta) \leq c_{1-\alpha}\right\},$$

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,

$$\underset{\theta\in\Theta:\theta=\theta_0}{inf} \ \underset{n\to\infty}{\lim inf}\ P\left(\theta \in CS_n(1-\alpha)\right) \geq 1-\alpha$$

Suppose we generalize our basic setting and assume that $\left\{\left((Y_i^p)_{p=1}^P, X_i\right) : 1 \leq i \leq n\right\}$ 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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi(Y_i, X_i; \theta_n, F_n)}{\sigma(\theta_n, F_n)}$$

for any sequence $F_n \in \mathcal{F}$ and $\theta_n \in \Theta \setminus \overline{\Theta}^{I}_{\mathcal{X}}(F_n)$.

(ii) The necessary Laws of Large Numbers for triangular arrays hold to ensure that $\big|\widehat{\sigma}^2(\theta_n) -$

$\sigma^2(\theta_n, F_n)\big| = o_p(1)$ over any sequence $F_n \in \mathcal{F}$ and $\theta_n \in \Theta$.

We describe such conditions in the appendix. If they hold, then

$$\lim_{n\to\infty} inf \inf_{\substack{\theta\in\Theta:\theta=\theta_0 \\ F\in\mathcal{F}}} P_F\left(\theta \in CS_n(1-\alpha)\right) \geq 1-\alpha$$

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

$$\widetilde{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\widehat{\Omega}(\theta)},$$

where $\widehat{\Omega}(\theta)$ is the estimator of $\widehat{\sigma}(\theta)$ that would result if the inequalities were *binding* a.s. To construct it we would replace each indicator function $\mathbb{1}\left\{\tau^p(Y^p|X;\theta^p) \geq 0\right\}$ 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 κ_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.

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

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[6]. Second, there has been a recent discussion among industry watchers of a takeover of Rite-Aid by one of its competitors. This issue provides a natural policy application of our results that we will explore. 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. Our justification for modelling 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:

player 1: CVS,   player 2: Rite Aid,   player 3: Walgreens.

We will use p to refer generically to any one of the three players in the model and we will use q, r to denote his opponents. Let Y^p denote the number of stores opened by p in a market.

5.1 Data overview

5.1.1 Units of observation

The decision variable Y_i^p 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[7] CBSAs were split into the divisions determined by Office of Budget and Management and each

[6]Source: http://clients.ibisworld.com/industryus/ataglance.aspx?indid=1054

[7]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.

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.

5.1.2 Choices and outcomes observed in the data

Table 1 summarizes some descriptive features of choices observed. It highlights the richness of the action space in this application. Table 2 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 1: Summary statistics for Y^p

	Y^1	Y^2	Y^3
Total	$7,004$	$4,318$	$7,283$
Mean	7.34	4.52	7.63
Stdev	21.95	15.57	23.88
25^{th} percentile	0	0	1
Median	1	0	1
75^{th} percentile	4	3	4
90^{th} percentile	16	10	17
95^{th} percentile	39	21	41
99^{th} percentile	112	71	106

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

	Y^1	Y^2	Y^3	Y^2+Y^3	Y^1+Y^3	Y^1+Y^2
Y^1	–	0.70	0.79	0.86	–	–
Y^2	0.70	–	0.49	–	0.62	–
Y^3	0.79	0.49	–	–	–	0.72

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), Tamer (2003), Davis (2006)). Even though strategic substitutability is justified as the prediction of economic theory in our setting, the correlation pattern in Table 2 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). One empirical question we will use our model explore is whether a model of strategic substitutes can explain 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 3 shows, it wipes out much of the positive association observed in the data.

Table 3: Correlations if game is reduced to binary choice

	$\mathbb{1}\{Y^1 > 0\}$	$\mathbb{1}\{Y^2 > 0\}$	$\mathbb{1}\{Y^3 > 0\}$
$\mathbb{1}\{Y^1 > 0\}$	–	0.23	0.07
$\mathbb{1}\{Y^2 > 0\}$	0.23	–	0.04
$\mathbb{1}\{Y^3 > 0\}$	0.07	0.04	–

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 η^p in the latter case would be more precise. We will see below that our results confirm this.

5.1.3 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,

POP=population, INC=average income per household, DENS=population density,

AGE=median age in the population, BUS=total number of business establishments,

$DIST^p$=distance to the nearest distribution center of p.

And we used

$$X = \left(POP, INC, DENS, AGE, BUS, DIST^1, DIST^2, DIST^3\right).$$

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.

5.2 Specifications for the strategic index η^p

We will refer generically to the three players as p, q, r and we will 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 Θ such that strategic substitutability is ensured. That is,

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

We want to focus on simple specifications for the indices $X_i'\theta^{p,q}$ and $X_i'\theta^{p,r}$. Since θ 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

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

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

$$E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot (Y^q + Y^r)\,\big|X = x\big] \geq E\big[\mathbb{1}\{Y^p \leq y^p\}\big|X = x\big] \cdot E\big[(Y^q + Y^r)\,\big|X = x\big] \quad (9)$$

for almost every (x, y^p) in our inferential range (which we will describe 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

$$\eta^p(y^{-p}; X|\theta^p) = \theta^{p,q} \cdot y^q + \theta^{p,r} \cdot y^r, \quad where \quad \theta \geq 0 \, \forall \, \theta \in \Theta. \tag{10}$$

We normalize Θ so that $\|\theta^p\| = 1$ for each p since our identified set is closed under nonnegative re-scaling of θ^p (if θ 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 η^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 η^p. Denote $\theta^p = (\theta_1^p, \theta_2^p, \theta_3^p, \theta_4^p)'$ and $D^{p,q} \equiv DIST^p - DIST^q$ for every $p \neq q$. Define

$$\phi^{p,q}(X|\theta^p) =$$
$$\left(\theta_1^p + \theta_2^p \cdot \frac{1}{POP} + \theta_3^p \cdot (D^{p,q} - 200) \cdot \mathbb{1}\{D^{p,q} \geq 200\} + \theta_4^p \cdot \frac{(D^{p,q} - 200) \cdot \mathbb{1}\{D^{p,q} \geq 200\}}{POP}\right), \tag{11}$$

Population is measured in units of 500K inhabitants in (11). The strategic index for p is specified as

$$\eta^p(y^{-p}; X|\theta^p) = \phi^{p,q}(X|\theta^p) \cdot y^q + \phi^{p,r}(X|\theta^p) \cdot y^r. \tag{12}$$

Strategic substitutability is imposed by forcing the parameter space Θ 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.

5.3 Results

Our target coverage probability is 95% throughout. Our parameter space Θ 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 $h_n = c \cdot \widehat{\sigma}(X) \cdot n^{-\alpha_h}$ (we used individual bandwidths for each X, each proportional to $\widehat{\sigma}(X)$), $b_n = c_b \cdot \overline{\Omega} \cdot n^{-\alpha_b}$ and $\kappa_n = c_\kappa \cdot \overline{\Omega} \cdot \log(n)^{-1}$, where $\overline{\Omega} = \underset{\theta \in \Theta}{max} \left|\widehat{\sigma}(\theta)\right|$. We chose these tuning parameters proportional to $\overline{\Omega}$ to ensure our procedure has a scale-invariant property. The choice of the constants c, c_b, c_κ, α_h and α_b are described in Appendix B.7. For our sample size $n = 954$ the values of these tuning parameters were $h_n \approx 0.16 \cdot \widehat{\sigma}(X)$, $b_n \approx 10^{-5}$ and $\kappa_n \approx 10^{-7}$. The inference range used was

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

where $\widehat{f}_X^{(0.15)}$ denotes the estimated 15th percentile of the density $\widehat{f}_X$. All of our main findings were robust to moderate changes in the tuning parameters used.

5.3.1 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 (9) 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 Θ 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.

5.3.2 Results for Specification 3

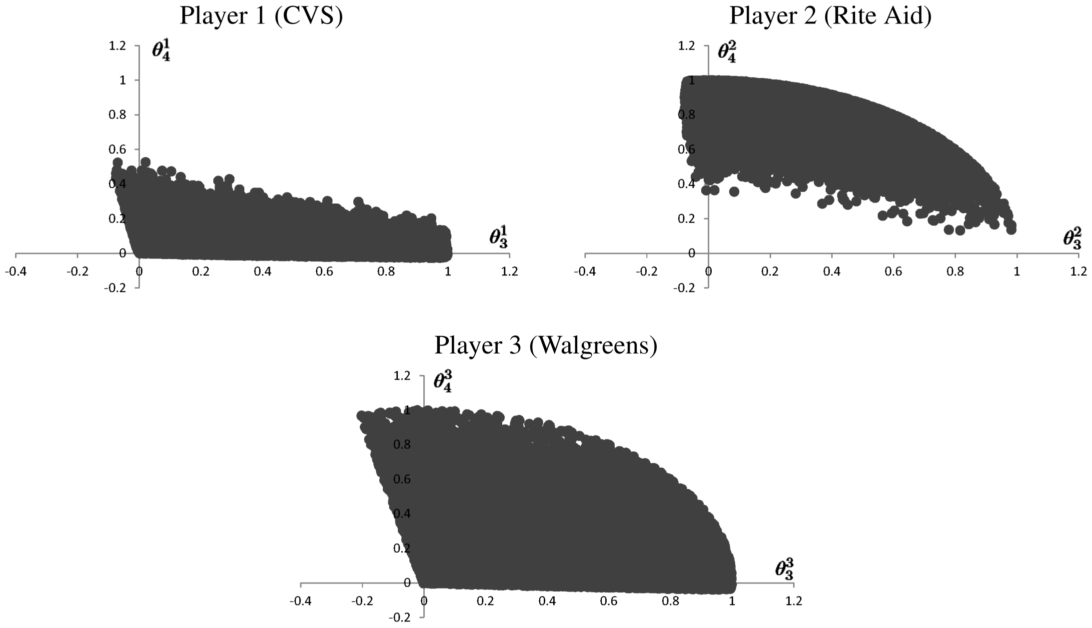
Our third specification produced a nonempty CS. Our first finding was a rejection of the assertion that $\theta^p = \theta^q$ for each $p \neq q$ (symmetry in parameters for all players). When we imposed this restriction we obtained an empty CS, with the smallest value of the test-statistic being 2.01.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

5.3.3 Evidence of asymmetric weights to opponents' strategies

Asymmetry of opponents' interaction effects is captured by the parameters θ_3^p and θ_4^p. Symmetry would hold for p in every market only if these parameters are jointly equal to zero. Figure 1 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).

We can study the asymmetry of strategic effects for specific markets. For example, figure 2 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.

Figure 1: Asymmetry of strategic interaction effects. 95% joint CS for θ_3^p and θ_4^p



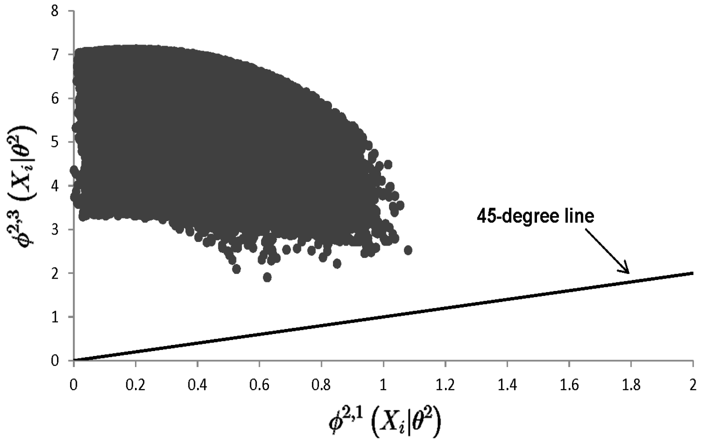
We wanted to learn more about what the data revealed regarding the closeness of competition between rival firms. Since symmetry could only be rejected for Rite Aid we focused only on this firm. We say that the competition effect from CVS is stronger than that of Walgreens in market i if $\underset{\theta^2 \in CS_n(1-\alpha)}{min}\left(\phi^{2,1}(X_i|\theta^2)\right) > \underset{\theta^2 \in CS_n(1-\alpha)}{max}\left(\phi^{2,3}(X_i|\theta^2)\right)$. 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.[8]

5.3.4 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 θ_2^p and θ_4^p would be consistent with interaction effects that decrease

[8]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).

Figure 2: CS for $\phi^{2,1}(X_i|\theta^2)$ and $\phi^{2,3}(X_i|\theta^2)$, for market i =CBSA 29404 (Lake County-Kenosha County, IL-WI)



with the size of the market. Figure 3 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 θ_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 $\phi^{p,q}(X|\theta^p) = \theta_1^p + \theta_2^p \cdot \frac{1}{POP}$. Figure 4 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 is less significant in larger markets.

5.4 Results from modeling entry as a binary decision

As Table 3 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, modelling 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 3: Market size and strategic interaction. 95% joint CS for θ_2^p and θ_4^p

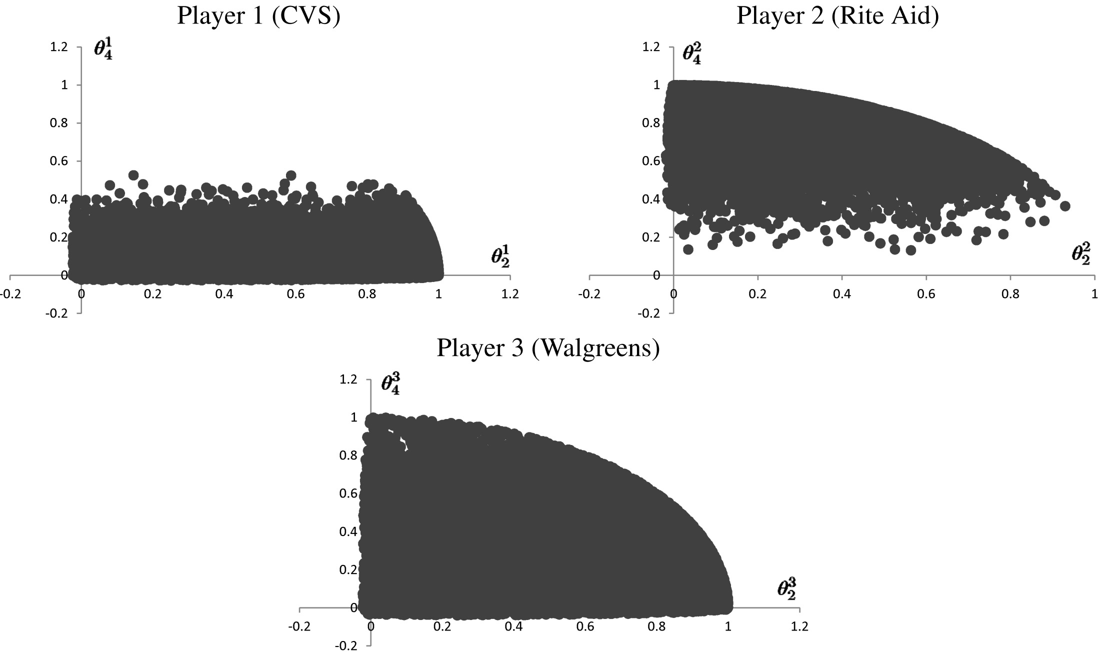


Figure 5 shows we now failed to reject that $\theta_3^p = \theta_4^p = 0$ for Rite Aid. Thus, we failed to reject that Rite Aid gives equal weights to both opponents across all markets. Hence, we conclude that key features of strategic interaction that are captured by intensive margin strategies are obscured if we focus attention solely on binary entry/no entry decisions.

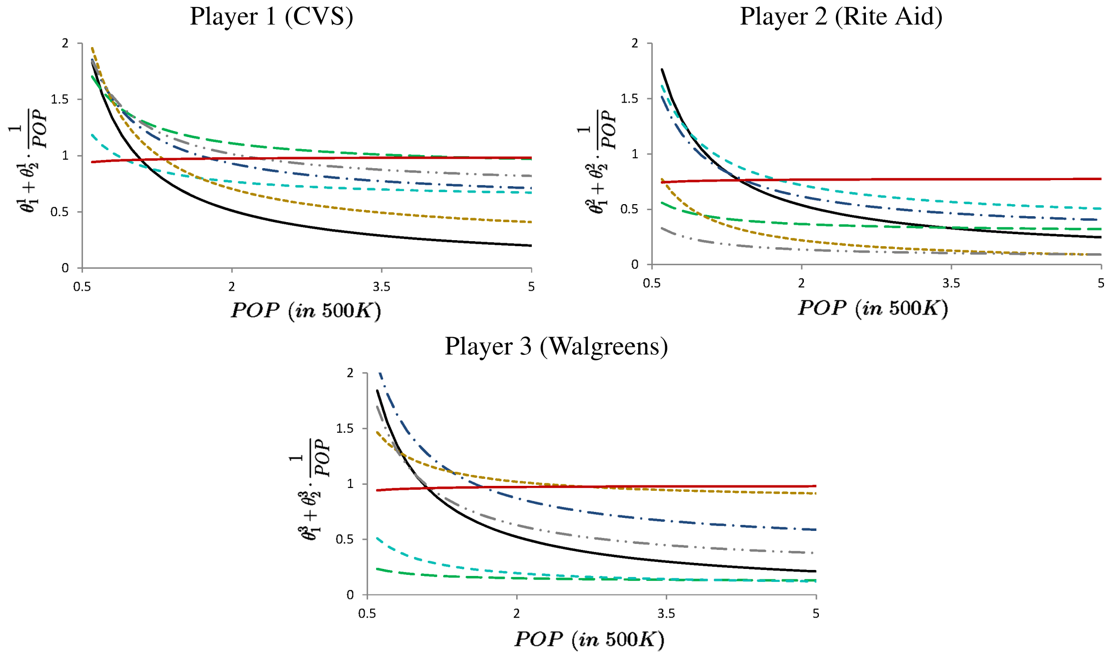
5.5 Explaining the data with strategic substitutes

Our identification results are not based on the unconditional covariance[9] 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 2. Consider in particular the correlation lower bound

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

[9]Note that the Law of Total Covariance predicts that $Cov\left(Y^p, \eta^p(Y^{-p}; X|\theta^p)\right) = E\left[Cov\left(Y^p, \eta^p(Y^{-p}; X|\theta^p)\right) \big| X\right] + Cov\left(E[Y^p|X], E[\eta^p(Y^{-p}; X|\theta^p)|X]\right)$. 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.

Figure 4: $\theta_1^p + \theta_2^p \cdot \frac{1}{POP}$ for a range of POP values (measured in 500K). Solid **black** line depicts the results for the largest value of θ_2^p in our CS. Solid **red** line depicts the results for the smallest value of θ_2^p in our CS. Dotted lines correspond to five randomly drawn parameter values within our CS.

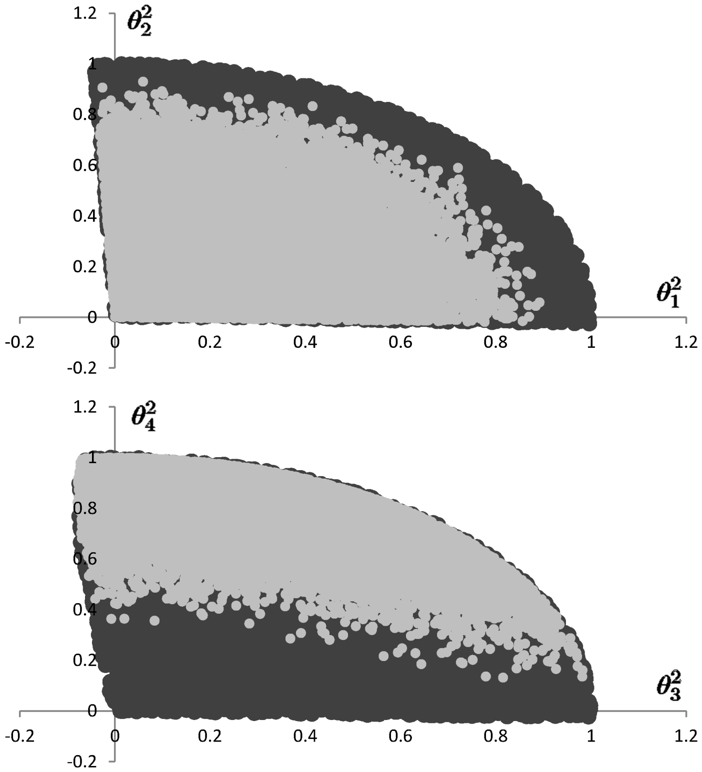


This lower bound was -0.06, -0.14 and -0.03 for CVS, Rite Aid and Walgreens, respectively. Thus our strategic index, which is a weighted average of opponents' actions (whose weights depended on X), is negatively associated with the strategies of each firm. This is true despite the fact that the raw actions 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.

6 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

Figure 5: Comparing our previous results (shown in clear gray) with the CS from a binary choice entry model (shown in dark gray). Player 2 (Rite Aid).



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.

A Appendix– Proofs of our identification results

A.1 Proof of Result 1

Recall from (4) that

$$\overline{\eta}^p_\sigma(X) \geq \overline{\eta'}^p_\sigma(X) \quad \Longrightarrow \quad \overline{\nu}^p_\sigma(v;\xi^p) - \overline{\nu}^p_\sigma(u;\xi^p) \leq \overline{\nu}^p_{\sigma'}(v;\xi^p) - \overline{\nu}^p_{\sigma'}(u;\xi^p) \quad \forall\, u < v \in \mathcal{A}^p$$

Fix any $y^p \in \mathcal{A}^p$ and define the following indicator function,

$$\mathbb{I}^p_\sigma\left(y^p;\xi^p\right) = \max_{u \leq y^p}\left(\min_{v \geq y^p+1}\Big(\mathbb{1}\left\{\overline{\nu}^p_\sigma\left(v;\xi^p\right) - \overline{\nu}^p_\sigma\left(u;\xi^p\right) \leq 0\right\}\Big)\right).$$

By (4), we have

$$\overline{\eta}^p_\sigma(X) \geq \overline{\eta}^p_{\sigma'}(X) \implies \mathbb{I}^p_\sigma\left(y^p;\xi^p\right) \geq \mathbb{I}^p_{\sigma'}\left(y^p;\xi^p\right).$$

Now suppose σ^{-p} and $\sigma^{-p'}$ are any pair of beliefs that produce unique expected-payoff maximizing choices for p given the realization of ξ^p, and let $y^p_\sigma(\xi^p)$ and $y^p_{\sigma'}(\xi^p)$ denote the corresponding optimal choices. Then for any $y^p \in \mathcal{A}^p$,

$$\mathbb{1}\left\{y^p_\sigma(\xi^p) \leq y^p\right\} = \mathbb{I}^p_\sigma\left(y^p;\xi^p\right) \;\text{ and }\; \mathbb{1}\left\{y^p_{\sigma'}(\xi^p) \leq y^p\right\} = \mathbb{I}^p_{\sigma'}\left(y^p;\xi^p\right).$$

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

A.2 Appendix– Proof of Theorem 1

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

$$E_{\xi^{-p}|X}\left[\eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p});X) \;\middle|\; X\right].$$

This is the expected value of η^p, conditional on X, in the j^{th} BNE. By definition, this is equal to $\overline{\eta}^p_{\sigma_{*j}}(X)$, which was defined previously as

$$\overline{\eta}^p_{\sigma_{*j}}(X) = \sum_{y^{-p} \in \mathcal{A}^{-p}} \sigma^{-p}_{*j}(y^{-p}|X) \cdot \eta^p(y^{-p}; X).$$

Now fix any $y^p \in \mathcal{A}^p$. By iterated expectations we have

$$E\left[\mathbb{1}\left\{Y^p \leq y^p\right\} \cdot \eta^p(Y^{-p}; X)\middle|X\right] = \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi|X}\left[\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\} \cdot \eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X) \mid X\right]$$

Assumption 3 (independent private shocks, i.e $\xi^p \perp \xi^{-p}|X$) yields

$$\begin{aligned}
&E\left[\mathbb{1}\left\{Y^p \leq y^p\right\} \cdot \eta^p(Y^{-p}; X)\middle|X\right] \\
&= \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi^p|X}\left[\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\} \mid X\right] \cdot E_{\xi^{-p}|X}\left[\eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X) \mid X\right], \\
&= \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi^p|X}\left[\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\} \mid X\right] \cdot \overline{\eta}^p_{\sigma_{*j}}(X).
\end{aligned}$$

Therefore, by Assumption 3 we can express

$$E\left[\mathbb{1}\left\{Y^p \leq y^p\right\} \cdot \eta^p(Y^{-p}; X)\middle|X\right] = E_{\xi^p|X}\left[\sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot \mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\} \cdot \overline{\eta}^p_{\sigma_{*j}}(X) \middle| X\right] \tag{A-1}$$

Next note that

$$\begin{aligned}
&E\left[\mathbb{1}\left\{Y^p \leq y^p\right\}|X\right] \cdot E\left[\eta^p(Y^{-p}; X)|X\right] \\
&= \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi^p|X}\left[\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\}|X\right] \times \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi^{-p}|X}\left[\eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X) \middle| X\right] \\
&= \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot E_{\xi^p|X}\left[\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\right\}|X\right] \times \sum_{j=1}^{J} P^{\mathcal{S}}_j(X) \cdot \overline{\eta}^p_{\sigma_{*j}}(X)
\end{aligned} \tag{A-2}$$

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

$$E\left[\mathbb{1}\left\{Y^p\leq y^p\right\}\cdot\eta^p(Y^{-p};X)\middle|X\right]-E\left[\mathbb{1}\left\{Y^p\leq y^p\right\}|X\right]\cdot E\left[\eta^p(Y^{-p};X)|X\right]=$$
$$E_{\xi^p|X}\left[\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\cdot\overline{\eta}^p_{\sigma_{*j}}(X)-\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\right)\right.$$
$$\left.\times\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\overline{\eta}^p_{\sigma_{*j}}(X)\right)\middle|X\right] \tag{A-3}$$

By Result 1, w.p.1 in (ξ^p) we have

$$\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\cdot\overline{\eta}^p_{\sigma_{*j}}(X)$$
$$-\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\right)\times\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\overline{\eta}^p_{\sigma_{*j}}(X)\right)\geq 0\quad\forall\, y^p\in\mathcal{A}^p. \tag{A-4}$$

To see why, simple algebra can be used to show that

$$\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\cdot\overline{\eta}^p_{\sigma_{*j}}(X)$$
$$-\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\right)\times\left(\sum_{j=1}^J P_j^{\mathcal{S}}(X)\cdot\overline{\eta}^p_{\sigma_{*j}}(X)\right)$$
$$=\sum_{\ell=1}^J\sum_{j=1}^J P_\ell^{\mathcal{S}}(X)P_j^{\mathcal{S}}(X)\cdot\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\cdot\left(1-\mathbb{1}\left\{y^p_{\sigma_{*\ell}}(\xi^p)\leq y^p\right\}\right)\cdot\left(\overline{\eta}^p_{\sigma_{*j}}(X)-\overline{\eta}^p_{*\ell}(X)\right)\geq 0,$$

where the last inequality follows from Result 1 which implies that, w.p.1 in ξ^p and $\forall\;y^p$,

$$\overline{\eta}^p_{\sigma_{*j}}(X)<\overline{\eta}^p_{*\ell}(X)\quad\Longrightarrow\quad\mathbb{1}\left\{y^p_{\sigma_{*j}}(\xi^p)\leq y^p\right\}\leq\mathbb{1}\left\{y^p_{\sigma_{*\ell}}(\xi^p)\leq y^p\right\}.$$

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

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

This concludes the proof. $\square$

B Econometric appendix

We focus on settings where the researcher observes an iid sample $\left((Y_i^p)_{p=1}^P, X_i\right)_{i=1}^n$, with[10] $\left((Y_i^p)_{p=1}^P, X_i\right) \sim F$. We assume that X can be split as $X = \left(X^c, X^d\right)$, where X^c have absolutely continuous distribution with respect to Lebesgue measure and X^d have a discrete distribution. We will 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 $\mathcal{X} \subset \text{Supp}(X)$ denote a prespecified set such that

$$\mathcal{X} \cap \text{Supp}(X^c) \subset \text{int}\left(\text{Supp}(X^c)\right).$$

We will maintain the assumption that $f_X(x) \geq \underline{f} > 0$ for all $x \in \mathcal{X}$. Let[11] $\mathbb{I}_{\mathcal{X}}(x) = \mathbb{1}\{x \in \mathcal{X}\}$. Let

$$T_{\mathcal{X}}^p(\theta^p) = E_{Y^p,X}\left[\max\left\{\tau^p(Y^p|X;\theta^p), 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X)\right]. \tag{B-1}$$

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

$$T_{\mathcal{X}}(\theta) = \sum_{p=1}^P T_{\mathcal{X}}^p(\theta^p).$$

Note that $T_{\mathcal{X}}(\theta) \geq 0$, and $T_{\mathcal{X}}(\theta) = 0$ if and only if $Pr\left(\tau^p(Y^p|X;\theta^p) \leq 0 \big| X \in \mathcal{X}\right) = 1$ 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 Θ^I,

$$\Theta_{\mathcal{X}}^I = \left\{\theta \in \Theta : T_{\mathcal{X}}^p(\theta^p) = 0 \text{ for } p = 1, \ldots, P\right\}.$$

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

[10] We will generalize our assumptions to a setting where $\left((Y_i^p)_{p=1}^P, X_i\right)_{i=1}^n$ is a triangular array in Section B.6, below.

[11] The indicator function $\mathbb{I}_{\mathcal{X}}$ could be replaced with a smooth "trimming" function.

B.2 Estimators involved in our construction

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

$$\mathcal{H}\left(X_i - x; h\right) = K\left(\frac{X_i^c - x^c}{h}\right) \cdot \mathbb{1}\left\{X_i^d - x^d = 0\right\}.$$

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

$$\begin{aligned}
\widehat{f}_X(x) &= (nh_n^q)^{-1}\sum_{i=1}^n \mathcal{H}\left(X_i - x; h_n\right),\\
\widehat{F}_{Y^p}(y^p|x) &= \left(nh_n^q \cdot \widehat{f}_X(x)\right)^{-1}\sum_{i=1}^n \mathbb{1}\left\{Y_i^p \leq y^p\right\} \cdot \mathcal{H}\left(X_i - x; h_n\right),\\
\widehat{\lambda}^p(x;\theta^p) &= \left(nh_n^q \cdot \widehat{f}_X(x)\right)^{-1}\sum_{i=1}^n \eta^p\left(Y_i^{-p}; x|\theta^p\right) \cdot \mathcal{H}\left(X_i - x; h_n\right),\\
\widehat{\mu}^p(y^p|x;\theta^p) &= \left(nh_n^q \cdot \widehat{f}_X(x)\right)^{-1}\sum_{i=1}^n \mathbb{1}\left\{Y_i^p \leq y^p\right\} \cdot \eta^p\left(Y_i^{-p}; x|\theta^p\right) \cdot \mathcal{H}\left(X_i - x; h_n\right),\\
\widehat{\tau}^p(y^p|x;\theta^p) =& \widehat{F}_{Y^p}(y^p|x) \cdot \widehat{\lambda}^p(x;\theta^p) - \widehat{\mu}^p(x;\theta^p).
\end{aligned}$$

Our estimators for $T_{\mathcal{X}}^p(\theta^p)$ and $T_{\mathcal{X}}(\theta)$ are

$$\begin{aligned}
\widehat{T}_{\mathcal{X}}^p(\theta^p) &= \frac{1}{n}\sum_{i=1}^n \widehat{\tau}^p(Y_i^p|X_i;\theta^p) \cdot \mathbb{1}\left\{\widehat{\tau}^p(Y_i^p|X_i;\theta^p) \geq -b_n\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i),\\
\widehat{T}_{\mathcal{X}}(\theta) &= \sum_{p=1}^P \widehat{T}_{\mathcal{X}}^p(\theta^p).
\end{aligned} \tag{B-2}$$

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

B.3 Basic Assumptions

Assumption B1. ***(Smoothness)***
As before, express any $x \in$ Supp(X) generically as $x \equiv \left(x^c, x^d\right)$ with x^c corresponding to the continuously distributed elements in X. Denote

$$\mathcal{W} = \left\{(x,y) \in \textit{Supp}(X,Y)\colon x \in \mathcal{X}\right\}.$$

Recall that we defined before

$$\begin{aligned}
F_{Y^p}(y^p|x) &= E_{Y^p|X}\left[\mathbb{1}\left\{Y^p \leq y^p\right\}\big|X=x\right],\\
\lambda^p(x;\theta^p) &= E_{Y^{-p}|X}\left[\eta^p(Y^{-p};x|\theta^p)\big|X=x\right],\\
\mu^p(y^p|x;\theta^p) &= E_{Y|X}\left[\mathbb{1}\left\{Y^p \leq y^p\right\}\cdot\eta^p(Y^{-p};x|\theta^p)\big|X=x\right],\\
\tau^p(y^p|x;\theta^p) &= F_{Y^p}(y^p|x)\cdot\lambda^p(x;\theta^p) - \mu^p(y^p|x;\theta^p).
\end{aligned}$$

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

$$F_{Y^p}\left(y^p\big|x\right), \quad f_X(x), \quad E_{Y^{-p}|X}\left[\eta^p\left(Y^{-p};x'\big|\theta^p\right)\big|X=x\right],$$

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

Now let

$$\begin{aligned}
\gamma_p^I(y^p,x;\theta^p) &= E_{Y^p|X}\left[\mathbb{1}\left\{y^p \leq Y^p\right\}\cdot\mathbb{1}\left\{\tau^p(Y^p|x;\theta^p)\geq 0\right\}\big|X=x\right],\\
\gamma_p^{II}(x;\theta^p) &= E_{Y^p|X}\left[F_{Y^p}(Y^p|x)\cdot\mathbb{1}\left\{\tau^p(Y^p|x;\theta^p)\geq 0\right\}\big|X=x\right],\\
\gamma_p^{III}(x;\theta^p) &= E_{Y^p|X}\left[\mu^p(Y^p|x;\theta^p)\cdot\mathbb{1}\left\{\tau^p(Y^p|x;\theta^p)\geq 0\right\}\big|X=x\right],
\end{aligned}$$

For almost every $(x,y^p) \in \mathcal{W}$ and every $\theta^p \in \Theta$, the three objects defined above are$M$ times differentiable with respect to x^c with bounded derivatives, and this is also satisfied by the trimming function $\mathbb{I}(x)$. Finally, for some $\overline{Q} < \infty$,

$$\underset{(x,y^p)\in\mathcal{W}}{sup}\left|Q_{F_{Y^p}}(y^p|x)\right| \leq \overline{Q}, \qquad \underset{x\in\mathcal{X},\theta^p\in\Theta}{sup}\left|Q_{\lambda^p}(x;\theta^p)\right| \leq \overline{Q},$$

$$\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}\left|Q_{\mu^p}(y^p|x;\theta^p)\right| \leq \overline{Q}.$$

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 $\underset{v\in\mathbb{R}^q}{sup}\left|K(v)\right| \leq \overline{K} < \infty$. The bandwidth sequences b_n and h_n are such that, for a small enough $\epsilon_1 > 0$,*

$$n^{1/2-\epsilon_1}\cdot h_n^q\cdot b_n \longrightarrow \infty, \quad n^{1/2+\epsilon_1}\cdot b_n^2 \longrightarrow 0, \quad n^{1/2+\epsilon_1}\cdot h_n^M \longrightarrow 0.$$

Focus on bandwidths of the type $h_n \propto n^{-\alpha_h}$ and $b_n \propto n^{-\alpha_b}$. Let $\overline{\epsilon} > 0$ be an arbitrarily small, but strictly positive constant and let $\alpha_h = \frac{1}{2M} + \overline{\epsilon}$ and $\alpha_b = \frac{1}{4} + \overline{\epsilon}$. The conditions in Assumption B2 will be satisfied if

$$M \geq \left\lceil \frac{2 \cdot q}{1 - 4 \cdot \overline{\epsilon}(2+q)} \right\rceil.$$

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 $\theta^p \in \Theta$,

$$Pr\Big(-b \leq \tau^p(Y^p|X;\theta^p) < 0 \Big| X \in \mathcal{X}\Big) \leq b \cdot A \quad \forall\, 0 < b \leq \overline{b}.$$

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

$$\overline{\eta}^p(y^{-p}) = \underset{x\in\mathcal{X},\theta^p\in\Theta}{sup} \left|\eta^p(y^{-p};x|\theta^p)\right|.$$

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

(i) The classes of functions

$$\mathscr{F} = \Big\{ f\colon\ f(y^{-p}) = \eta^p \left(y^{-p}; x | \theta^p \right) \ \textit{for some } (x, \theta^p) \in \mathcal{X} \times \Theta \Big\},$$
$$\mathscr{F}' = \Big\{ f\colon\ f(x) = \lambda^p(x; \theta^p) \ \textit{for some } \theta^p \in \Theta \Big\},$$
$$\mathscr{F}'' = \Big\{ f\colon\ f(y^p, x) = \mu^p(y^p | x; \theta^p) \ \textit{for some } \theta^p \in \Theta \Big\},$$

are Euclidean (see Definition 2.7 in Pakes and Pollard (1989)) with respect to envelopes $\overline{\eta}^p(\cdot)$, $\overline{F}'(\cdot)$ and $\overline{F}''(\cdot)$ respectively, where $\overline{\eta}^p(Y^{-p})$ 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

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

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 η^p, those conditions can be used to verify part (i) of Assumption B4. In particular, η^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 Θ where $\tau(x, y; \theta^p) - b$ changes sign. Suppose $\underset{(x,y)\in\mathcal{X}\times\mathcal{A}}{sup} \mathcal{N}(x, y; b) \leq \overline{\mathcal{N}} < \infty$ for all $0 < b \leq \overline{b}$. 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 $\widehat{T}^p_{\mathcal{X}}(\theta^p)$

The following theorem summarizes the key asymptotic properties of $\widehat{T}^p_{\mathcal{X}}(\theta^p)$ under our assumptions.

Theorem 2. *Let*

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

and

$$\psi^p(Y_i,X_i;\theta^p) = \left(\max\left\{\tau^p(Y_i^p|X_i;\theta^p),0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)-T_{\mathcal{X}}^p(\theta^p)\right)+\psi_U^p(Y_i,X_i;\theta^p).$$

If Assumptions B1-B4 hold, then

$$\widehat{T}_{\mathcal{X}}^p(\theta^p) = T_{\mathcal{X}}^p(\theta^p)+\frac{1}{n}\sum_{i=1}^n\psi^p(Y_i,X_i;\theta^p)+\varepsilon_{p,n}(\theta^p),$$

$$\textit{where}\quad \psi^p(Y_i,X_i;\theta^p) = \left(\max\left\{\tau^p(Y_i^p|X_i;\theta^p),0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)-T_{\mathcal{X}}^p(\theta^p)\right)+\psi_U^p(Y_i,X_i;\theta^p),$$

$$\textit{and}\quad \underset{\theta^p\in\Theta}{sup}\left|\varepsilon_{p,n}(\theta^p)\right| = O_p\left(n^{-1/2-\epsilon}\right)\ \textit{for some } \epsilon>0.$$

The "influence function" ψ^p has two key properties:

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

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

$$\widehat{T}_{\mathcal{X}}(\theta) = T_{\mathcal{X}}(\theta)+\frac{1}{n}\sum_{i=1}^n\psi(Y_i,X_i;\theta)+\varepsilon_n(\theta), \\ \text{where}\quad \underset{\theta\in\Theta}{sup}\left|\varepsilon_n(\theta)\right| = O_p\left(n^{-1/2-\epsilon}\right)\ \text{ for some }\epsilon>0. \tag{B-3}$$

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

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

(ii) Let

$$\overline{\Theta}^I_{\mathcal{X}} = \{\theta \in \Theta : \tau^p(Y^p|X;\theta^p) < 0 \text{ w.p.1.} \; \forall\, p = 1,\ldots,P.\}$$

Then $\psi(Y_i, X_i;\theta) = 0$ w.p.1 $\forall\, \theta \in \overline{\Theta}^I_{\mathcal{X}}$.

We now proceed to prove Theorem 2.

B.4.1 Proof of Theorem 2

In Assumption B1 we described $\mathcal{W}$ as

$$\mathcal{W} = \Big\{(x,y) \in Supp(X,Y)\colon x \in \mathcal{X}\Big\},$$

where $\mathcal{X} \subset \text{Supp}(X)$ is a prespecified set such that $\mathcal{X} \cap \text{Supp}(X^c) \subset \text{int}\,(\text{Supp}(X^c))$. 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

$$\begin{aligned} &Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup} \Big|\widehat{\tau}^p(y^p|x;\theta^p) - \tau^p(y^p|x;\theta^p)\Big| \geq b_n \right) \\ &\leq D_1 \exp\left\{-\sqrt{n}h_n^q \left(D_2 \cdot b_n - D_3 \cdot h_n^M\right)\right\}. \end{aligned}$$

For given y^p, x and θ^p define

$$Q^p_{F_{Y^p}}(y^p|x) = F_{Y^p}(y^p|x)\cdot f_X(x), \quad Q_{\lambda^p}(x;\theta^p) = \lambda^p(x;\theta^p)\cdot f_X(x),$$

$$Q_{\mu^p}(y^p|x;\theta^p) = \mu^p(y^p|x;\theta^p)\cdot f_X(x).$$

and let

$$\begin{aligned} \widehat{Q}_{F_{Y^p}}(y^p|x) &= (nh_n^q)^{-1}\sum_{i=1}^n \mathbb{1}\left\{Y_i^p \leq y^p\right\}\cdot \mathcal{H}\left(X_i - x; h_n\right), \\ \widehat{Q}_{\lambda^p}(x;\theta^p) &= (nh_n^q)^{-1}\sum_{i=1}^n \eta^p\left(Y_i^{-p}; x|\theta^p\right)\cdot \mathcal{H}\left(X_i - x; h_n\right), \\ \widehat{Q}_{\mu^p}(y^p|x;\theta^p) &= (nh_n^q)^{-1}\sum_{i=1}^n \mathbb{1}\left\{Y_i^p \leq y^p\right\}\cdot \eta^p\left(Y_i^{-p}; x|\theta^p\right)\cdot \mathcal{H}\left(X_i - x; h_n\right). \end{aligned}$$

Using an M^{th} order approximation, our smoothness restrictions in Assumption B1 imply the existence of a finite constant $\overline{M}$ such that,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|E\left[\widehat{f}_X(x)\right]-f_X(x)\right|\leq\overline{M}\cdot h_n^M,\\
&\sup_{(x,y^p)\in\mathcal{W}}\left|E\left[\widehat{Q}_F^p(y^p|x)\right]-Q_{F_{Y^p}}(y^p|x)\right|\leq\overline{M}\cdot h_n^M,\\
&\sup_{x\in\mathcal{X},\theta^p\in\Theta}\left|E\left[\widehat{Q}_{\lambda^p}(x;\theta^p)\right]-Q_{\lambda^p}(x;\theta^p)\right|\leq\overline{M}\cdot h_n^M,\\
&\sup_{\substack{(x,y^p)\in\mathcal{W}\\ \theta^p\in\Theta}}\left|E\left[\widehat{Q}_{\mu^p}(y^p|x;\theta^p)\right]-Q_{\mu^p}(y^p|x;\theta^p)\right|\leq\overline{M}\cdot h_n^M.
\end{aligned}\tag{B-4}$$

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

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

is *Euclidean*[12] with respect to the constant envelope $\overline{K}$. Lemma 2.4 in Pakes and Pollard (1989) also implies that the class of functions

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

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

$$\begin{aligned}
&\mathscr{F}_1=\left\{f:f(y^{-p},x)=\eta^p(y^{-p};u|\theta^p)\cdot\mathcal{H}\left(x-u;h\right)\ \text{ for some } u\in\mathcal{X} \text{ and } \theta^p\in\Theta\right\},\\
&\mathscr{F}_2=\left\{f:f(y,x)=\mathbb{1}\left\{y^p\leq v\right\}\cdot\eta^p(y^{-p};u|\theta^p)\cdot\mathcal{H}\left(x-u;h\right)\ \text{ for some } v\in\mathbb{R}, u\in\mathcal{X} \text{ and } \theta^p\in\Theta\right\}
\end{aligned}$$

are Euclidean with respect to the envelope $\overline{K}\cdot\overline{\eta}^p(\cdot)$. 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

[12] See Definition 2.7 in Pakes and Pollard (1989).

that for any $\delta > 0$,

$$
\begin{aligned}
&Pr\left(\sup_{x\in\mathcal{X}}\left|\widehat{f}_X(x)-f_X(x)\right|\geq\delta\right)\leq A_1\cdot\exp\left\{-\left(\sqrt{n}\cdot h_n^q\left(A_2\cdot\delta-A_3\cdot h_n^M\right)\right)^2\right\},\\
&Pr\left(\sup_{(x,y^p)\in\mathcal{W}}\left|\widehat{Q}_{F_{Y^p}}(y^p|x)-Q_{F_{Y^p}}(y^p|x)\right|\geq\delta\right)\leq A_1\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(A_2\cdot\delta-A_3\cdot h_n^M\right)\right\},\\
&Pr\left(\sup_{x\in\mathcal{X},\theta^p\in\Theta}\left|\widehat{Q}_{\lambda^p}(x;\theta^p)-Q_{\lambda^p}(x;\theta^p)\right|\geq\delta\right)\leq A_1\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(A_2\cdot\delta-A_3\cdot h_n^M\right)\right\},\\
&Pr\left(\sup_{\substack{(x,y^p)\in\mathcal{W}\\ \theta^p\in\Theta}}\left|\widehat{Q}_{\mu^p}(y^p|x;\theta^p)-Q_{\mu^p}(y^p|x;\theta^p)\right|\geq\delta\right)\leq A_1\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(A_2\cdot\delta-A_3\cdot h_n^M\right)\right\}.
\end{aligned}
\tag{B-5}
$$

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

$$
\begin{aligned}
\psi_{F_{Y^p}}\left(Y_i^p,X_i,y^p,x;h\right)&=\frac{\left(\mathbb{1}\left\{Y_i^p\leq y^p\right\}-F_{Y^p}(y^p|x)\right)}{f_X(x)}\cdot\mathcal{H}\left(X_i-x;h\right),\\
\psi_{\lambda^p}\left(Y_i^{-p},X_i,x,\theta^p;h\right)&=\frac{\left(\eta^p\left(Y_i^{-p};x\middle|\theta^p\right)-\lambda^p(x;\theta^p)\right)}{f_X(x)}\cdot\mathcal{H}\left(X_i-x;h\right),\\
\psi_{\mu^p}\left(Y_i,X_i,y^p,x,\theta^p;h\right)&=\frac{\left(\mathbb{1}\left\{Y_i^p\leq y^p\right\}\cdot\eta^p\left(Y_i^{-p};x\middle|\theta^p\right)-\mu^p\left(y^p\middle|x;\theta^p\right)\right)}{f_X(x)}\cdot\mathcal{H}\left(X_i-x;h\right).
\end{aligned}
\tag{B-6}
$$

And let

$$
\begin{aligned}
\widehat{\zeta}_{F_{Y^p}}(y^p,x)&=\left(\left[\widehat{Q}_{F_{Y^p}}(y^p|x)-Q_{F_{Y^p}}(y^p|x)\right]\quad\left[\widehat{f}_X(x)-f_X(x)\right]\right)',\\
\widehat{\zeta}_{\lambda^p}(x,\theta^p)&=\left(\left[\widehat{Q}_{\lambda^p}(x;\theta^p)-Q_{\lambda^p}(x;\theta^p)\right]\quad\left[\widehat{f}_X(x)-f_X(x)\right]\right)',\\
\widehat{\zeta}_{\mu^p}(y^p,x,\theta^p)&=\left(\left[\widehat{Q}_{\mu^p}(y^p|x;\theta^p)-Q_{\mu^p}(y^p|x;\theta^p)\right]\quad\left[\widehat{f}_X(x)-f_X(x)\right]\right)'.
\end{aligned}
$$

Note that (B-5) implies that for any $\delta > 0$,

$$
\begin{aligned}
&Pr\left(\sup_{(x,y^p)\in\mathcal{W}}\left|\widehat{\zeta}_{F_{Y^p}}(y^p,x)\right|\geq\delta\right)\leq Pr\left(\sup_{(x,y^p)\in\mathcal{W}}\left|\widehat{Q}_{F_{Y^p}}(y^p|x)-Q_{F_{Y^p}}(y^p|x)\right|\geq\frac{\delta}{\sqrt{2}}\right)\\
&+Pr\left(\sup_{x\in\mathcal{X}}\left|\widehat{f}_X(x)-f_X(x)\right|\geq\frac{\delta}{\sqrt{2}}\right)\\
&\leq A_1\cdot\exp\left\{-\left(\sqrt{n}\cdot h_n^q\left(A_2\cdot\frac{\delta}{\sqrt{2}}-A_3\cdot h_n^M\right)\right)^2\right\}+A_1\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(A_2\cdot\frac{\delta}{\sqrt{2}}-A_3\cdot h_n^M\right)\right\}\\
&\leq 2\cdot A_1\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(A_2\cdot\frac{\delta}{\sqrt{2}}-A_3\cdot h_n^M\right)\right\}
\end{aligned}
$$

Similarly (B-5) yields

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

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

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

$$\widehat{F}_{Y^p}(y^p|x)-F_{Y^p}(y^p|x)=\frac{1}{nh_n^q}\sum_{i=1}^{n}\psi_{F_{Y^p}}\left(Y_i^p,X_i,y^p,x;h_n\right)+\xi_n^{F_{Y^p}}(y^p,x),$$

$$\text{where}\quad \xi_n^{F_{Y^p}}(y^p,x)=\frac{1}{2}\widehat{\zeta}_{F_{Y^p}}(y^p,x)'\begin{pmatrix}0 & -\frac{1}{\widetilde{f}_X^2(x)}\\ -\frac{1}{\widetilde{f}_X^2(x)} & \frac{2\widetilde{Q}_{F_{Y^p}}(y^p|x)}{\widetilde{f}_X^3(x)}\end{pmatrix}\widehat{\zeta}_{F_{Y^p}}(y^p,x)$$

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

$$\widehat{\lambda}^p(x;\theta^p)-\lambda^p(x;\theta^p)=\frac{1}{nh_n^q}\sum_{i=1}^{n}\psi_{\lambda^p}\left(Y_i^{-p},X_i,x,\theta^p;h_n\right)+\xi_n^{\lambda^p}(x,\theta^p),$$

$$\text{where}\quad \xi_n^{\lambda^p}(x,\theta^p)=\frac{1}{2}\widehat{\zeta}_{\lambda^p}(x,\theta^p)'\begin{pmatrix}0 & -\frac{1}{\breve{f}_X^2(x)}\\ -\frac{1}{\breve{f}_X^2(x)} & \frac{2\breve{Q}_{\lambda^p}(x;\theta^p)}{\breve{f}_X^3(x)}\end{pmatrix}\widehat{\zeta}_{\lambda^p}(x,\theta^p)$$

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

$$\widehat{\mu}^p(y^p|x;\theta^p)-\mu^p(y^p|x;\theta^p)=\frac{1}{nh_n^q}\sum_{i=1}^{n}\psi_{\mu^p}\left(Y_i,X_i,y^p,x,\theta^p;h_n\right)+\xi_n^{\mu^p}(y^p,x,\theta^p),$$

$$\text{where}\quad \xi_n^{\mu^p}(y^p,x,\theta^p)=\frac{1}{2}\widehat{\zeta}_{\mu^p}(y^p,x,\theta^p)'\begin{pmatrix}0 & -\frac{1}{\ddot{f}_X^2(x)}\\ -\frac{1}{\ddot{f}_X^2(x)} & \frac{2\ddot{Q}_{\mu^p}(y^p|x;\theta^p)}{\ddot{f}_X^3(x)}\end{pmatrix}\widehat{\zeta}_{\mu^p}(y^p,x,\theta^p)$$

where $\left(\ddot{f}_X(x),\ddot{Q}_{\mu^p}(y^p|x;\theta^p)\right)$ belongs in the line segment connecting $\left(\widehat{f}_X(x),\widehat{Q}_{\mu^p}(y^p|x;\theta^p)\right)$ and $\left(f_X(x),Q_{\mu^p}(y^p|x;\theta^p)\right)$. Let $\overline{Q}$ be as described in Assumption B1. For any $0<\underline{f}^*<\underline{f}$, define

$$D(\underline{f}^*) = \left\| \begin{matrix} 0 & -\frac{1}{(\underline{f}^*)^2} \\ -\frac{1}{(\underline{f}^*)^2} & \frac{3\overline{Q}}{(\underline{f}^*)^3} \end{matrix} \right\|. \tag{B-7}$$

Let $0 < \underline{f}^* < \underline{f}$ and $D(\underline{f}^*)$ be as described in (B-7). Combining our previous results, for any $\delta > 0$,

$$\begin{aligned} &Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup} \left|\xi_n^{F_{Y^p}}(y^p,x)\right| \geq \delta \right) \leq Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup} \left|\widehat{Q}_{F_{Y^p}}(y^p|x) - Q_{F_{Y^p}}(y^p|x)\right| \geq \overline{Q} \right) \\ &+ Pr\left(\underset{x\in\mathcal{X}}{sup} \left|\widehat{f}_X(x) - f_X(x)\right| \geq \underline{f} - \underline{f}^* \right) + Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup} \left|\widehat{\zeta}_{F_{Y^p}}(y^p,x)\right| \geq \sqrt{\frac{2\delta}{D(\underline{f}^*)}} \right) \\ &\leq 4A_1 \cdot \exp\left\{ -\sqrt{n}\cdot h_n^q \left(A_2 \cdot \min\left\{ \sqrt{\frac{\delta}{D(\underline{f}^*)}}, \overline{Q}, \underline{f} - \underline{f}^* \right\} - A_3 \cdot h_n^M \right) \right\}. \end{aligned}$$

And the same bound holds for

$$Pr\left(\underset{x\in\mathcal{X},\theta^p\in\Theta}{sup} \left|\xi_n^{\lambda^p}(x,\theta^p)\right| \geq \delta \right) \quad \text{and} \quad Pr\left(\underset{\substack{(x,y^p)\in\mathcal{W} \\ \theta^p\in\Theta}}{sup} \left|\xi_n^{\mu^p}(y^p,x,\theta^p)\right| \geq \delta \right).$$

Assumption B4 and Lemma 2.14 in Pakes and Pollard (1989) we have that the classes of functions

$$\begin{aligned} \mathscr{G}_1 &= \{g\colon g(y^p,x) = \psi_{F_{Y^p}}(y^p,x,v^p,u;h)\colon (v^p,u)\in\mathcal{W},\, h>0\}, \\ \mathscr{G}_2 &= \{g\colon g(y^{-p},x) = \psi_{\lambda^p}\left(y^{-p},x,u,\theta^p\right)\colon u\in\mathcal{X},\, \theta^p\in\Theta,\, h>0\}, \\ \mathscr{G}_3 &= \{g\colon g(y,x) = \psi_{\mu^p}(y,x,v^p,u,\theta^p;h)\colon (v^p,u)\in\mathcal{W},\, \theta^p\in\Theta,\, h>0\} \end{aligned}$$

are Euclidean with respect to envelopes $\frac{2\overline{K}}{\underline{f}}$, $\frac{2\overline{K}\overline{\eta}^p(\cdot)}{\underline{f}}$ and $\frac{2\overline{K}\overline{\eta}^p(\cdot)}{\underline{f}}$, respectively. The existence of moments feature of $\overline{\eta}^p(\cdot)$ 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 A_1', A_2' and A_3' such that for any $\delta > 0$, the probabilities

$$\begin{aligned} &Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{F_{Y^p}}(Y_i^p,X_i,y^p,x;h_n) \right| \geq \delta \right), \\ &Pr\left(\underset{x\in\mathcal{X},\theta^p\in\Theta}{sup} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\lambda^p}\left(Y_i^{-p},X_i,x,\theta^p;h_n\right) \right| \geq \delta \right), \\ &Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\mu^p}(Y_i,X_i,y^p,x,\theta^p;h_n) \right| \geq \delta \right), \end{aligned}$$

are bounded above by

$$A_1' \cdot \exp\left\{-\sqrt{n} \cdot h_n^q \left(A_2' \cdot \delta - A_3' \cdot h_n^M\right)\right\}.$$

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

$$\begin{aligned}
&Pr\left(\sup_{(x,y^p)\in\mathcal{W}} \left|\widehat{F}_{Y^p}\left(y^p|x\right) - F_{Y^p}\left(y^p|x\right)\right| \geq \delta\right) \leq Pr\left(\sup_{x\in\mathcal{X}} \left|\widehat{f}_X(x) - f_X(x)\right| \geq \underline{f} - \underline{f}^*\right) \\
&+ Pr\left(\sup_{(x,y^p)\in\mathcal{W}} \left|\frac{1}{nh_n^q}\sum_{i=1}^n \psi_{F_{Y^p}}\left(Y_i^p, X_i, y^p, x; h_n\right)\right| \geq \frac{\delta}{2}\right) \\
&+ Pr\left(\sup_{(x,y^p)\in\mathcal{W}} \left|\xi_n^{F_{Y^p}}(y^p, x)\right| \geq \frac{\delta}{2}\right) \\
&\leq A_1 \cdot \exp\left\{-\left(\sqrt{n} \cdot h_n^q \left(A_2 \cdot (\underline{f} - \underline{f}^*) - A_3 \cdot h_n^M\right)\right)^2\right\} \\
&+ A_1' \cdot \exp\left\{-\sqrt{n} \cdot h_n^q \left(A_2' \cdot \frac{\delta}{2} - A_3' \cdot h_n^M\right)\right\} \\
&+ 4A_1 \cdot \exp\left\{-\sqrt{n} \cdot h_n^q \left(A_2 \cdot \min\left\{\sqrt{\frac{\delta}{2D}}, \overline{Q}, \underline{f} - \underline{f}^*\right\} - A_3 \cdot h_n^M\right)\right\} \\
&\leq B_1 \cdot \exp\left\{-\sqrt{n} \cdot h_n^q \left(B_2 \cdot \min\left\{\frac{\delta}{2}, \sqrt{\frac{\delta}{2D}}, \overline{Q}, \underline{f} - \underline{f}^*\right\} - B_3 \cdot h_n^M\right)\right\}
\end{aligned}$$

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

$$\begin{aligned}
&Pr\left(\sup_{x\in\mathcal{X},\theta^p\in\Theta} \left|\widehat{\lambda}^p(x;\theta^p) - \lambda^p(x;\theta^p)\right| \geq \delta\right), \\
&Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta} \left|\widehat{\mu}^p(y^p|x;\theta^p) - \mu^p(y^p|x;\theta^p)\right| \geq \delta\right).
\end{aligned}$$

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

$$\begin{aligned}
&\psi_{\tau^p}\left(Y_i, X_i, y^p, x, \theta^p; h\right) \\
&= \lambda^p(x;\theta^p) \cdot \psi_{F_{Y^p}}\left(Y_i^p, X_i, y^p, x; h\right) + F_{Y^p}(y^p|x) \cdot \psi_{\lambda^p}\left(Y_i^{-p}, X_i, x, \theta^p; h\right) - \psi_{\mu^p}\left(Y_i, X_i, y^p, x, \theta^p; h\right) \\
&= \Bigg[\lambda^p(x;\theta^p) \cdot \left(\mathbb{1}\left\{Y_i^p \leq y^p\right\} - F_{Y^p}(y^p|x)\right) + F_{Y^p}(y^p|x) \cdot \left(\eta^p(Y_i^p; x|\theta^p) - \lambda^p(x;\theta^p)\right) \\
&\quad - \left(\mathbb{1}\left\{Y_i^p \leq y^p\right\} \cdot \eta^p(Y_i^p; x|\theta^p) - \mu^p(y^p|x;\theta^p)\right)\Bigg] \cdot \frac{\mathcal{H}(X_i - x; h)}{f_X(x)}
\end{aligned} \tag{B-8}$$

From our previous results we have

$$\widehat{\tau}^p\left(y^p|x;\theta^p\right)-\tau^p\left(y^p|x;\theta^p\right)=\frac{1}{nh_n^q}\sum_{i=1}^{n}\psi_{\tau^p}\left(Y_i,X_i,y^p,x,\theta^p;h_n\right)+\xi_n^{\tau^p}(y^p,x,\theta^p), \tag{B-9}$$

where

$$\begin{aligned}\xi_n^{\tau^p}(y^p,x,\theta^p)&=\lambda^p(x;\theta^p)\cdot\xi^{F_{Y^p}}(y^p,x)+F_{Y^p}(y^p|x)\cdot\xi^{\lambda^p}(x,\theta^p)-\xi_n^{\mu^p}(y^p,x,\theta^p)\\&+\left(\widehat{F}_{Y^p}(y^p|x)-F_{Y^p}(y^p|x)\right)\cdot\left(\widehat{\lambda}^p(x;\theta^p)-\lambda^p(x;\theta^p)\right).\end{aligned}$$

Let

$$\underset{x\in\mathcal{X},\theta^p\in\Theta}{sup}|\lambda^p(x;\theta^p)|=\overline{\lambda}^p.$$

For any $\delta>0$,

$$\begin{aligned}&Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}\left|\xi_n^{\tau^p}(y^p,x,\theta^p)\right|\geq\delta\right)\leq Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup}\left|\xi_n^{F_{Y^p}}(y^p,x)\right|\geq\frac{\delta}{4\overline{\lambda}^p}\right)\\&+Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}\left|\xi_n^{\lambda^p}(x,\theta^p)\right|\geq\frac{\delta}{4}\right)+Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}\left|\xi_n^{\mu^p}(y^p,x,\theta^p)\right|\geq\frac{\delta}{4}\right)\\&+Pr\left(\underset{(x,y^p)\in\mathcal{W}}{sup}\left|\widehat{F}_{Y^p}(y^p|x)-F_{Y^p}(y^p|x)\right|\geq\frac{\sqrt{\delta}}{2}\right)\\&+Pr\left(\underset{x\in\mathcal{X},\theta^p\in\Theta}{sup}\left|\widehat{\lambda}^p(x;\theta^p)-\lambda^p(x;\theta^p)\right|\geq\frac{\sqrt{\delta}}{2}\right)\end{aligned}$$

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

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

Let $\underline{B}=\frac{1}{2}\cdot\min\left\{\frac{1}{\sqrt{D\overline{\lambda}^p}},\frac{1}{\sqrt{D}},1,2\overline{Q},2(\underline{f}-\underline{f}^*)\right\}$ and define $C_1\equiv 4\cdot B_1$, $C_2\equiv B_2\cdot\underline{B}$, $C_3\equiv B_3$.

We have

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

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

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

is Euclidean with respect to the envelope $\frac{2\overline{\lambda}^p\overline{K}}{\underline{f}}+\frac{4\overline{K}\overline{\eta}^p(\cdot)}{\underline{f}}$. The existence of moments feature of $\overline{\eta}^p(\cdot)$ 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 $\delta>0$,

$$Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\frac{1}{nh_n^q}\sum_{i=1}^n\psi_{\tau^p}\left(Y_i,X_i,y^p,x,\theta^p;h_n\right)\right|\geq\delta\right)$$
$$\leq C_1'\cdot\exp\left\{-\sqrt{n}\cdot h_n^q\left(C_2'\cdot\delta-C_3'\cdot h_n^M\right)\right\}.$$

As before, if we let $0<\underline{f}^*<\underline{f}$ be as described in (B-7)

$$\begin{aligned}&Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\widehat{\tau}^p(y^p|x;\theta^p)-\tau^p(y^p|x;\theta^p)\right|\geq\delta\right)\\&\leq Pr\left(\sup_{x\in\mathcal{X}}\left|\widehat{f}_X(x)-f_X(x)\right|\geq\underline{f}-\underline{f}^*\right)\\&+Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\frac{1}{nh_n^q}\sum_{i=1}^n\psi_{\tau^p}\left(Y_i,X_i,y^p,x,\theta^p;h_n\right)\right|\geq\frac{\delta}{2}\right)\\&+Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\xi_n^{\tau^p}(y^p,x,\theta^p)\right|\geq\frac{\delta}{2}\right).\end{aligned}$$

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

$$\begin{aligned}&Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\widehat{\tau}^p(y^p|x;\theta^p)-\tau^p(y^p|x;\theta^p)\right|\geq\delta\right)\\&\leq D_1\exp\left\{-\sqrt{n}h_n^q\left(D_2\cdot\min\left\{\delta,\delta^{1/2},\delta^{1/4},1\right\}-D_3\cdot h_n^M\right)\right\},\end{aligned}\tag{B-10}$$

where $D_1=3\cdot\max\{A_1,C_1',C_1\}$, $D_2=\frac{1}{2}\cdot\min\left\{C_2',C_2,2A_2(\underline{f}-\underline{f}^*)\right\}$, $D_3=\max\{A_3,C_3,C_3'\}$.

Our results also imply

$$
\begin{aligned}
&\widehat{\tau}^p\left(y^p|x;\theta^p\right)-\tau^p\left(y^p|x;\theta^p\right)=\frac{1}{nh_n^q}\sum_{i=1}^n\psi_{\tau^p}\left(Y_i,X_i,y^p,x,\theta^p;h_n\right)+\xi_n^{\tau^p}(y^p,x,\theta^p),\\
&\text{where}\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\xi_n^{\tau^p}(y^p,x,\theta^p)\right|=O_p\left(\frac{\log(n)^2}{nh_n^q}\right),\\
&\text{and}\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\widehat{\tau}^p(y^p|x;\theta^p)-\tau^p(y^p|x;\theta^p)\right|=O_p\left(\frac{\log(n)}{\sqrt{nh_n^q}}\right).
\end{aligned}
\tag{B-11}
$$

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

$$
\begin{aligned}
&Pr\left(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}\left|\widehat{\tau}^p(y^p|x;\theta^p)-\tau^p(y^p|x;\theta^p)\right|\geq b_n\right)\\
&\leq D_1\exp\left\{-\sqrt{n}h_n^q\left(D_2\cdot b_n-D_3\cdot h_n^M\right)\right\},
\end{aligned}
\tag{B-12}
$$

This concludes Step 1 of our proof.

Step 2

Here we use the results from Step 1 to show that

$$
\begin{aligned}
&\widehat{T}_{\mathcal{X}}^p(\theta^p)=\frac{1}{n}\sum_{i=1}^n\widehat{\tau}^p(Y_i^p|X_i;\theta^p)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)+\varphi_n^p(\theta^p),\\
&\text{where}\quad\sup_{\theta^p\in\Theta}\left|\varphi_n^p(\theta^p)\right|=O_p\left(n^{-1/2-\epsilon}\right)\quad\text{for some }\epsilon>0.
\end{aligned}
$$

We begin by noting that we can express

$$
\widehat{T}_{\mathcal{X}}^p(\theta^p)=\frac{1}{n}\sum_{i=1}^n\widehat{\tau}^p(Y_i^p|X_i;\theta^p)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)+\varphi_n^p(\theta^p),
$$

where

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

$$+\underbrace{\left|\frac{2}{n}\sum_{i=1}^{n}\widehat{\tau}^p(Y_i^p|X_i;\theta^p)\cdot \mathbb{1}\left\{|\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)| \geq b_n\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)\right|}_{\equiv\left|\varphi_n^{p,2}(\theta^p)\right|}.$$

We begin by examining $\varphi_n^{p,2}$. Using (B-11), $\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}\widehat{\tau}^p(y^p|x;\theta^p) = O_p(1)$. Therefore,

$$\underset{\theta^p\in\Theta}{sup}\left|\varphi_n^{p,2}(\theta^p)\right| \leq O_p(1)\cdot \underset{\theta^p\in\Theta}{sup}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{|\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)| \geq b_n\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)\right|$$

Take any $\alpha > 0$ and any $\varepsilon > 0$. Then,

$$\begin{aligned}
&Pr\left(n^{\alpha}\cdot \underset{\theta^p\in\Theta}{sup}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{|\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)| \geq b_n\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)\right| > \varepsilon\right)\\
&\leq Pr\left(\mathbb{1}\left\{\underset{\theta^p\in\Theta}{sup}|\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)| \geq b_n\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i) \neq 0 \text{ for some } i = 1,\ldots,n\right)\\
&\leq \sum_{i=1}^{n} Pr\left(\mathbb{1}\left\{\underset{\theta^p\in\Theta}{sup}|\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)| \geq b_n\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i) \neq 0\right)\\
&\leq n\cdot Pr\left(\underset{(x,y^p)\in\mathcal{W},\theta^p\in\Theta}{sup}|\widehat{\tau}^p(y^p|x;\theta^p) - \tau^p(y^p|x;\theta^p)| \geq b_n\right)\\
&\leq n\cdot D_1\exp\left\{-\frac{1}{2}\sqrt{n}h_n^q\left(D_2\cdot b_n - D_3\cdot h_n^M\right)\right\} = D_1\exp\left\{-\frac{1}{2}\sqrt{n}h_n^q\left(D_2\cdot b_n - D_3\cdot h_n^M\right) + \log(n)\right\} \longrightarrow 0
\end{aligned}$$

Therefore, $\underset{\theta^p\in\Theta}{sup}\left|\varphi_n^{p,2}(\theta^p)\right| = o_p\left(n^{-\alpha}\right)$. In particular, the following much weaker (but useful for our purposes) result holds,

$$\underset{\theta^p\in\Theta}{sup}\left|\varphi_n^{p,2}(\theta^p)\right| = O_p\left(n^{-1/2-\epsilon}\right) \text{ for some } \epsilon > 0.$$

We move on to $\varphi_n^{p,1}(\theta^p)$. Note that

$$\widehat{\tau}^p(Y_i^p|X_i;\theta^p) = \sum_{j=0}^{1}\left(\tau^p(Y_i^p|X_i;\theta^p)\right)^{1-j}\cdot\left(\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)\right)^{j}.$$

Therefore,

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

Using (B-11) we have

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

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

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

For a given $b>0$ denote

$$g_i^{p,1}(\theta^p,b)=\mathbb{1}\left\{-b\leq\tau^p(Y_i^p|X_i;\theta^p)<0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i).$$

And let

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

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

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

where

$$\underset{\theta^p\in\Theta}{sup}\left|\xi_n^{p,1}(\theta^p)\right|=2Ab_n=O\left(b_n\right)\quad\text{and}\quad\underset{\theta^p\in\Theta}{sup}\,Var\left(\mathbb{1}\left\{-2b_n\leq\tau^p(Y_i^p|X_i;\theta^p)<0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)\right)=O(b_n).$$

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

$$\underset{\theta^p \in \Theta}{sup} \left|\nu_n^{p,1}(\theta^p)\right| = O_p\left(\sqrt{\frac{b_n}{n}}\right) = O_p\left(b_n\right).$$

Combining these results, we have

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

And therefore

$$\underset{\theta^p \in \Theta}{sup} \left|\varphi_n^{p,1}(\theta^p)\right| \leq O(b_n) \times O_p\left(b_n\right) = O_p\left(b_n^2\right) = O_p\left(n^{-1/2-\epsilon}\right) \quad \text{for some } \epsilon > 0.$$

Where the last line follows from the bandwidth convergence restrictions in Assumption B2. Combining the results for $\varphi_n^{p,1}$ and $\varphi_n^{p,2}$,

$$\begin{aligned}
&\widehat{T}_{\mathcal{X}}^p(\theta^p) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\tau}^p(Y_i^p|X_i;\theta^p) \cdot \mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p) \geq 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) + \varphi_n^p(\theta^p), \\
&\text{where} \quad \underset{\theta^p \in \Theta}{sup} \left|\varphi_n^p(\theta^p)\right| = O_p\left(n^{-1/2-\epsilon}\right) \quad \text{for some } \epsilon > 0.
\end{aligned} \tag{B-13}$$

Step 3

This is the last step in the proof. We take the results from Step 2 to show that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \left(\widehat{\tau}^p(Y_i^p|X_i;\theta^p) - \tau^p(Y_i^p|X_i;\theta^p)\right) \cdot \mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p) \geq 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\
=&\frac{1}{n^2}\sum_{j \neq i}\sum_{i=1}^{n} g_{\tau^p}\left(X_i, Y_i, X_j, Y_j; \theta^p, h_n\right) + \varrho_n^{p,1}(\theta^p), \\
&\text{where} \quad \underset{\theta^p \in \Theta}{sup} \left|\varrho_n^{p,1}(\theta^p)\right| = O_p\left(n^{-1/2-\epsilon}\right) \quad \text{for some } \epsilon > 0.
\end{aligned}$$

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

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{\tau}^{p}(Y_i^p|X_i;\theta^p)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)= \\
&\frac{1}{n}\sum_{i=1}^{n}\max\left\{\tau^p(Y_i^p|X_i;\theta^p),0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\tau}^{p}(Y_i^p|X_i;\theta^p)-\tau^p(Y_i^p|X_i;\theta^p)\right)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)
\end{aligned}\tag{B-14}$$

Let ψ_{τ^p} be as defined in (B-8). For any pair of observations i,j in $1,\ldots,n$ and $h>0$ let

$$g_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h\right)=\frac{1}{h^q}\cdot\psi_{\tau^p}\left(Y_j,X_j,Y_i^p,X_i,\theta^p;h\right)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i).\tag{B-15}$$

Note that

$$\underset{\theta^p\in\Theta}{sup}\left|\frac{1}{n^2}\sum_{i=1}^{n}g_{\tau^p}\left(X_i,Y_i,X_i,Y_i;\theta^p,h_n\right)\right|=O_p\left(\frac{1}{nh_n^q}\right)=O_p\left(n^{-1/2-\epsilon}\right)\ \text{ for some } \epsilon>0.$$

Combined with (B-11), this yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\tau}^{p}(Y_i^p|X_i;\theta^p)-\tau^p(Y_i^p|X_i;\theta^p)\right)\cdot\mathbb{1}\left\{\tau^p(Y_i^p|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i) \\
=&\frac{1}{n^2}\sum_{j\neq i}\sum_{i=1}^{n}g_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h_n\right)+\varrho_n^{p,1}(\theta^p), \\
&\text{where }\ \underset{\theta^p\in\Theta}{sup}\left|\varrho_n^{p,1}(\theta^p)\right|=O_p\left(\frac{\log(n)^2}{nh_n^q}\right)+O_p\left(\frac{1}{nh_n^q}\right)=O_p\left(n^{-1/2-\epsilon}\right)\ \text{ for some } \epsilon>0.
\end{aligned}\tag{B-16}$$

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

$$g_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h\right)=g^a_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h\right)+g^b_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h\right)+g^c_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h\right),$$

where

$$g^a_{\tau^p}\left(X_i, Y_i, X_j, Y_j; \theta^p, h\right) =$$
$$\frac{1}{h^q} \cdot \lambda^p(X_i;\theta^p) \cdot \left(\mathbb{1}\left\{Y^p_j \leq Y^p_i\right\} - F_{Y^p}(Y^p_i|X_i)\right) \cdot \mathbb{1}\left\{\tau^p(Y^p_i|X_i;\theta^p) \geq 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},$$
$$g^b_{\tau^p}\left(X_i, Y_i, X_j, Y_j; \theta^p, h\right) =$$
$$\frac{1}{h^q} \cdot F_{Y^p}(Y^p_i|X_i) \cdot \left(\eta^p(Y^{-p}_j; X_i|\theta^p) - \lambda^p(X_i;\theta^p)\right) \cdot \mathbb{1}\left\{\tau^p(Y^p_i|X_i;\theta^p) \geq 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},$$
$$g^c_{\tau^p}\left(X_i, Y_i, X_j, Y_j; \theta^p, h\right) =$$
$$\frac{1}{h^q} \cdot \left(\mathbb{1}\left\{Y^p_j \leq Y^p_i\right\} \cdot \eta^p(Y^{-p}_j; X_i|\theta^p) - \mu^p(Y^p_i|X_i;\theta^p)\right) \cdot \mathbb{1}\left\{\tau^p(Y^p_i|X_i;\theta^p) \geq 0\right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)},$$

Let γ^I_p, γ^{II}_p and γ^{III}_p be as defined in Assumption B1. By the smoothness conditions in Assumption B1, there exists a $C < \infty$ such that

$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|E\left[g^a_{\tau^p}\left(x, y, X, Y; \theta^p, h\right)\right]\right| \leq C \cdot h^M,$$
$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|E\left[g^b_{\tau^p}\left(x, y, X, Y; \theta^p, h\right)\right]\right| \leq C \cdot h^M,$$
$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|E\left[g^c_{\tau^p}\left(x, y, X, Y; \theta^p, h\right)\right]\right| \leq C \cdot h^M.$$

And

$$E\left[g^a_{\tau^p}\left(X, Y, x, y; \theta^p, h\right)\right] = \left(\gamma^I_p(y^p, x; \theta^p) - \gamma^{II}_p(x;\theta^p)\right) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma^a_p(y, x; \theta^p, h),$$
$$E\left[g^b_{\tau^p}\left(X, Y, x, y; \theta^p, h\right)\right] = \left(\eta^p(y^{-p}; x|\theta^p) - \lambda^p(x;\theta^p)\right) \cdot \gamma^{II}_p(x;\theta^p) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma^b_p(y, x; \theta^p, h),$$
$$E\left[g^c_{\tau^p}\left(X, Y, x, y; \theta^p, h\right)\right] = \left(\gamma^I_p(y^p, x; \theta^p) \cdot \eta^p(y^{-p}; x|\theta^p) - \gamma^{III}_p(x;\theta^p)\right) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma^c_p(y, x; \theta^p, h),$$

where

$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|\varsigma^a_p(y, x; \theta^p, h)\right| \leq C \cdot h^M,$$
$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|\varsigma^b_p(y, x; \theta^p, h)\right| \leq C \cdot h^M,$$
$$\underset{(x,y)\in\mathcal{W},\theta^p\in\Theta}{sup} \left|\varsigma^c_p(y, x; \theta^p, h)\right| \leq C \cdot h^M,$$

In particular, this implies that

$$\underset{\theta^p\in\Theta}{sup}\left|E\left[g_{\tau^p}\left(X_i, Y_i, X_j, Y_j; \theta^p, h_n\right)\big|X_i, Y_i\right]\right| \leq C \cdot h^M_n,$$

and if we define

$$
\begin{aligned}
&\psi^p_U\left(Y,X;\theta^p\right) = \\
&\Big[\left(\gamma^I_p(Y^p,X;\theta^p)-\gamma^{II}_p(X;\theta^p)\right)\cdot\lambda^p(X;\theta^p)+\left(\eta^p(Y^{-p};X|\theta^p)-\lambda^p(X;\theta^p)\right)\cdot\gamma^{II}_p(X;\theta^p) \\
&+\left(\gamma^I_p(Y^p,X;\theta^p)\cdot\eta^p(Y^{-p};X|\theta^p)-\gamma^{III}_p(X;\theta^p)\right)\Big]\cdot\mathbb{I}_{\mathcal{X}}(X),
\end{aligned}
\tag{B-17}
$$

then

$$
E\left[g_{\tau^p}\left(X_i,Y_i,X_j,Y_j;\theta^p,h_n\right)\big|X_j,Y_j\right]=\psi^p_U\left(Y_j,X_j;\theta^p\right)+\varsigma_{p,n}(\theta^p),\quad \text{where} \quad \underset{\theta^p\in\Theta}{sup}\left|\varsigma_{p,n}(\theta^p)\right|=O_p\left(h_n^M\right)
$$

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

$$
\mathscr{F}=\{f:\mathcal{W}\times\mathcal{W}\to\mathbb{R}:f(x_1,y_1,x_2,y_2)=g_{\tau^p}(x_1,y_1,x_2,y_2;\theta^p,h)\ \text{ for some } \theta^p\in\Theta \text{ and some } h>0\}
$$

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

$$
\frac{1}{n}\sum_{i=1}^n\left(\widehat{\tau}^p(Y^p_i|X_i;\theta^p)-\tau^p(Y^p_i|X_i;\theta^p)\right)\cdot\mathbb{1}\left\{\tau^p(Y^p_i|X_i;\theta^p)\geq 0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)=\frac{1}{n}\sum_{i=1}^n\psi^p_U(Y_i,X_i;\theta^p)+\vartheta_{p,n}(\theta^p),
$$

where

$$
\underset{\theta^p\in\Theta}{sup}\left|\vartheta_{p,n}(\theta^p)\right|=O_p\left(\frac{\log(n)^2}{nh_n^q}\right)+O_p\left(\frac{1}{nh_n^q}\right)+O_p\left(h_n^M\right)=O_p\left(n^{-1/2-\epsilon}\right)\ \text{ for some } \epsilon>0,
$$

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

$$
\begin{aligned}
&\widehat{T}^p_{\mathcal{X}}(\theta^p)=T^p_{\mathcal{X}}(\theta^p)+\frac{1}{n}\sum_{i=1}^n\psi^p(Y_i,X_i;\theta^p)+\varepsilon_{p,n}(\theta^p),\\
&\text{where}\quad \psi^p(Y_i,X_i;\theta^p)=\left(\max\left\{\tau^p(Y^p_i|X_i;\theta^p),0\right\}\cdot\mathbb{I}_{\mathcal{X}}(X_i)-T^p_{\mathcal{X}}(\theta^p)\right)+\psi^p_U(Y_i,X_i;\theta^p),\\
&\text{and}\quad \underset{\theta^p\in\Theta}{sup}\left|\varepsilon_{p,n}(\theta^p)\right|=O_p\left(n^{-1/2-\epsilon}\right)\ \text{ for some } \epsilon>0.
\end{aligned}
\tag{B-18}
$$

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

B.4.2 Two key properties of ψ^p

The "influence function" ψ^p has two key properties:

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

(ii) $\psi^p(Y_i, X_i; \theta^p) = 0 \ \forall\, \theta^p : \tau^p(Y^p|X; \theta^p) < 0$ 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 ψ^p. Note first that by definition,

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

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

$$\begin{aligned}
&E\left[\left(\gamma_p^I(Y_i^p, X_i; \theta^p) - \gamma_p^{II}(X_i; \theta^p)\right) \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\
&= E_{X_i}\Bigg[E_{Y_j|X_j}\left[E_{Y_i|X_i}\left[\left(\mathbb{1}\left\{Y_i^p \leq Y_j^p\right\} - F_{Y^p}(Y_j^p|X_i)\right)\big|X_i, Y_j, X_j\right] \cdot \mathbb{1}\left\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\right\}\big|X_j = X_i, X_i\right] \\
&\quad \times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\Bigg] \\
&= E_{X_i}\Bigg[E_{Y_j|X_j}\left[E_{Y_i|X_i}\left[\left(F_{Y^p}(Y_j^p|X_i) - F_{Y^p}(Y_j^p|X_i)\right)\big|X_i, Y_j, X_j\right] \cdot \mathbb{1}\left\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\right\}\big|X_j = X_i, X_i\right] \\
&\quad \times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\Bigg] = 0
\end{aligned}$$

For the second term we have

$$\begin{aligned}
&E\left[\left(\eta^p(Y_i^{-p}; X_i|\theta^p) - \lambda^p(X_i; \theta^p)\right) \cdot \gamma_p^{II}(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\
&\qquad = E_{X_i}\left[\left(\lambda^p(X_i; \theta^p) - \lambda^p(X_i; \theta^p)\right) \cdot \gamma_p^{II}(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] = 0,
\end{aligned}$$

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

$$\begin{aligned}
&E\left[\left(\gamma_p^I(Y_i^p, X_i; \theta^p) \cdot \eta^p(Y_i^{-p}; X_i|\theta^p) - \gamma_p^{III}(X_i; \theta^p)\right) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\
&= E_{X_i}\Bigg[E_{Y_j|X_j}\Big[E_{Y_i|X_i}\left[\left(\mathbb{1}\left\{Y_i^p \leq Y_j^p\right\} \cdot \eta^p(Y_i^{-p}; X_i|\theta^p) - \mu^p(Y_j|X_i; \theta^p)\right)\big|X_i, Y_j, X_j\right] \\
&\qquad \times \mathbb{1}\left\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\right\}\Big|X_j = X_i, X_i\Big] \times \mathbb{I}_{\mathcal{X}}(X_i)\Bigg] \\
&= E_{X_i}\left[E_{Y_j|X_j}\left[\left(\mu^p(Y_j|X_i; \theta^p) - \mu^p(Y_j|X_i; \theta^p)\right) \times \mathbb{1}\left\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\right\}\Big|X_j = X_i, X_i\right] \times \mathbb{I}_{\mathcal{X}}(X_i)\right] = 0
\end{aligned}$$

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

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

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

$$\underset{\theta \in \Theta}{sup}\ |\varsigma_n(\theta)| = \underset{\theta \in \Theta}{sup} \left| \frac{\sqrt{n} \cdot \varepsilon_n(\theta)}{\max\{\kappa_n, \sigma(\theta)\}} \right| = O_p\left(\frac{1}{n^\epsilon \cdot \kappa_n}\right) \text{ for some } \epsilon > 0,$$

and $n^\epsilon \kappa_n \to \infty$ for any $\epsilon > 0$. Let

$$\overline{\Theta}_{\mathcal{X}}^{I} = \{\theta \in \Theta : \tau^p(Y^p | X; \theta^p) < 0 \ \text{ w.p.1. } \forall\, p = 1, \ldots, P.\}$$

$\overline{\Theta}_{\mathcal{X}}^{I}$ 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 $\forall\, \theta \in \overline{\Theta}_{\mathcal{X}}^{I}$. On the other hand, inspecting the terms that comprise $\psi_U^p(Y, X; \theta^p)$ we can verify that $P\left(\psi_U^p(Y, X; \theta^p) \neq 0\right) > 0$ for any $\theta \in \Theta_{\mathcal{X}}^{I} \setminus \overline{\Theta}_{\mathcal{X}}^{I}$ and therefore $\sigma^2(\theta) > 0$ for any such θ. Therefore,

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

(ii) If $\theta \in \overline{\Theta}_{\mathcal{X}}^{I}$, then $t_n(\theta) = o_p(1)$.

(iii) If $\theta \in \Theta_{\mathcal{X}}^{I} \setminus \overline{\Theta}_{\mathcal{X}}^{I}$, then $t_n(\theta) \xrightarrow{d} \mathcal{N}(0, 1)$.

$t_n(\theta)$ is unfeasible because $\sigma^2(\theta)$ is unknown. However it can be estimated, we use $\widehat{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \widehat{\sigma}(\theta)\}}$, where

$$
\begin{aligned}
\widehat{\psi}^p_U\left(Y_i, X_i; \theta^p\right) &= \frac{1}{(n-1)}\sum_{j\neq i}\widehat{g}_{\tau^p}\left(X_j, Y_j, X_i, Y_i; \theta^p, h_n\right), \\
\widehat{\psi}^p(Y_i, X_i; \theta^p) &= \widehat{\tau}^p\left(Y^p_i|X_i; \theta^p\right)\cdot \mathbb{1}\left\{\widehat{\tau}^p\left(Y^p_i|X_i; \theta^p\right)\geq -b_n\right\}\cdot \mathbb{I}_{\mathcal{X}}(X_i) - \widehat{T}^p_{\mathcal{X}}(\theta^p) + \widehat{\psi}^p_U\left(Y_i, X_i; \theta^p\right), \\
\widehat{\sigma}^2(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}(Y_i, X_i; \theta)^2.
\end{aligned}
\tag{B-19}
$$

g_{τ^p} is as described in (B-15). Under our assumptions we have $\widehat{\sigma}^2(\theta) \xrightarrow{p} \sigma^2(\theta)$ for each $\theta \in \Theta$.

Confidence set and pointwise asymptotic properties

For a desired coverage probability $1-\alpha$, our confidence set (CS) for θ_0 is of the form

$$
CS_n\left(1-\alpha\right) = \left\{\theta \in \Theta \colon \widehat{t}_n(\theta) \leq c_{1-\alpha}\right\}, \tag{B-20}
$$

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,

$$
\underset{\theta\in\Theta:\theta=\theta_0}{inf}\ \underset{n\to\infty}{\lim inf}\ P\left(\theta \in CS_n(1-\alpha)\right) \geq 1-\alpha
$$

And if $\Theta^I_{\mathcal{X}} \setminus \overline{\Theta}^I_{\mathcal{X}} \neq \emptyset$, then

$$
\underset{\theta\in\Theta:\theta=\theta_0}{inf}\ \underset{n\to\infty}{\lim inf}\ P\left(\theta \in CS_n(1-\alpha)\right) = 1-\alpha
$$

Our CS will also satisfy

$$
\lim_{n\to\infty} P\left(\theta \in CS_n(1-\alpha)\right) = 0 \quad \forall\, \theta \in \Theta \setminus \Theta^I_{\mathcal{X}}.
$$

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 $\left\{\left((Y_i^p)_{p=1}^P, X_i\right) : 1 \leq i \leq n, n \geq 1\right\}$ 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 will denote $\psi(Y, X; \theta, F)$, $\sigma^2(\theta, F)$, $\Theta^I_{\mathcal{X}}(F)$, $\overline{\Theta}^I_{\mathcal{X}}(F)$, and so on. We assume the following conditions about $\mathcal{F}$.

Assumption B5. *The space of distributions $\mathcal{F}$ has common support and satisfies $P_F(X \in \mathcal{X}) \geq \underline{p} > 0$ 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 $b \leq \infty$,

$$\sup_{\substack{\theta \in \Theta \setminus \overline{\Theta}^I_{\mathcal{X}}(F) \\ F \in \mathcal{F}}} E_F\left[\frac{\left|\psi(Y, X; \theta, F)\right|^{2+\delta}}{\sigma^{2+\delta}(\theta, F)}\right] \leq b$$

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,

$$\lim_{\lambda \to \infty} \sup_{\substack{\theta \in \Theta \setminus \overline{\Theta}^I_{\mathcal{X}}(F) \\ F \in \mathcal{F}}} E_F\left[\frac{\left|\psi(Y, X; \theta, F)\right|^2}{\sigma^2(\theta, F)} \cdot \mathbb{1}\left\{\frac{\left|\psi(Y, X; \theta, F)\right|}{\sigma(\theta, F)} > \lambda\right\}\right] = 0.$$

To see why, note that for any $\widetilde{\lambda} > 0$ and $\delta > 0$, $\widetilde{\lambda}^\delta \cdot \psi(Y, X; \theta, F)^2 \cdot \mathbb{1}\left\{\left|\psi(Y, X; \theta, F)\right| > \widetilde{\lambda}\right\} \leq \left|\psi(Y, X; \theta, F)\right|^{2+\delta}$. Therefore $E\left[\psi(Y, X; \theta, F)^2 \cdot \mathbb{1}\left\{\left|\psi(Y, X; \theta, F)\right| > \widetilde{\lambda}\right\}\right] \leq \frac{E\left[\left|\psi(Y, X; \theta, F)\right|^{2+\delta}\right]}{\widetilde{\lambda}^\delta}$. The Lindeberg condition follows by using the δ described in Assumption B5, letting $\widetilde{\lambda} = \sigma(\theta, F)$ 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, θ_n) such that $F_n \in \mathcal{F}$ and $\theta_n \in \Theta^I_{\mathcal{X}}(F_n) \setminus \overline{\Theta}^I_{\mathcal{X}}(F_n)$,

$$\frac{\sqrt{n} \cdot \widehat{T}_{\mathcal{X}}(\theta_n)}{\sigma(\theta_n, F_n)} \xrightarrow{d} \mathcal{N}(0, 1).$$

And for any sequence (F_n, θ_n) such that $F_n \in \mathcal{F}$ and $\theta_n \in \overline{\Theta}^I_{\mathcal{X}}(F_n)$,

$$\frac{\sqrt{n} \cdot \widehat{T}_{\mathcal{X}}(\theta_n)}{\max\{\kappa_n, \sigma(\theta_n, F_n)\}} \xrightarrow{p} 0.$$

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

$$\liminf_{n\to\infty} \inf_{\substack{\theta\in\Theta:\theta=\theta_0 \\ F\in\mathcal{F}}} P_F\left(t_n(\theta) \leq c_{1-\alpha}\right) \geq 1-\alpha, \tag{B-21}$$

with

$$\liminf_{n\to\infty} \inf_{\substack{\theta\in\Theta:\theta=\theta_0 \\ F\in\mathcal{F}}} P_F\left(t_n(\theta) \leq c_{1-\alpha}\right) = 1-\alpha \quad \text{if } \Theta^I_{\mathcal{X}}(F) \setminus \overline{\Theta}^I_{\mathcal{X}}(F) \neq \emptyset \text{ for some } F \in \mathcal{F}.$$

Of course, our CS is based on $\widehat{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \widehat{\sigma}(\theta)\}}$, where $\widehat{\sigma}^2(\theta)$ 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 $\left|\widehat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n)\right| \xrightarrow{p} 0$ 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 $\psi_{F_{Y^p}}$, ψ_{λ^p}, ψ_{μ^p} and ψ_{τ^p} and g_{τ^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 $p = 1, \ldots, P$,

$$\sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ h > 0}} E_F \left[\left| \frac{1}{h^q} \psi_{F_{Y^p}} \left(Y_i^p, X_i, y^p, x; h, F \right) - E \left[\frac{1}{h^q} \psi_{F_{Y^p}} \left(Y_i^p, X_i, y^p, x; h, F \right) \right] \right|^{1+\delta} \right] \leq b,$$

$$\sup_{\substack{F \in \mathcal{F} \\ x \in \mathcal{X} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[\left| \frac{1}{h^q} \psi_{\lambda^p} \left(Y_i^{-p}, X_i, x, \theta^p; h \right) - E \left[\frac{1}{h^q} \psi_{\lambda^p} \left(Y_i^{-p}, X_i, x, \theta^p; h \right) \right] \right|^{1+\delta} \right] \leq b,$$

$$\sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[\left| \frac{1}{h^q} \psi_{\mu^p} \left(Y_i, X_i, y^p, x, \theta^p; h \right) - E \left[\frac{1}{h^q} \psi_{\mu^p} \left(Y_i, X_i, y^p, x, \theta^p; h \right) \right] \right|^{1+\delta} \right] \leq b,$$

$$\sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[\left| \frac{1}{h^q} \psi_{\tau^p} (Y, X, y^p, x, \theta^p; h, F) - E \left[\frac{1}{h^q} \psi_{\tau^p} (Y, X, y^p, x, \theta^p; h, F) \right] \right|^{1+\delta} \right] \leq b,$$

$$\sup_{\substack{F \in \mathcal{F} \\ (y, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[\left| g_{\tau^p} (Y, X, x, y; \theta^p, h, F) - E \left[g_{\tau^p} (Y, X, x, y; \theta^p, h, F) \right] \right|^{1+\delta} \right] \leq b,$$

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 $(F_n, \theta_n) \in \mathcal{F} \times \Theta$,

$$\left| \widehat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n) \right| \xrightarrow{p} 0.$$

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

$$\liminf_{n \to \infty} \inf_{\substack{\theta \in \Theta : \theta = \theta_0 \\ F \in \mathcal{F}}} P_F \left(\theta \in CS_n(1 - \alpha) \right) \geq 1 - \alpha,$$

with

$$\liminf_{n \to \infty} \inf_{\substack{\theta \in \Theta : \theta = \theta_0 \\ F \in \mathcal{F}}} P_F \left(\theta \in CS_n(1 - \alpha) \right) = 1 - \alpha \quad \text{if } \; \Theta^I_{\mathcal{X}}(F) \setminus \overline{\Theta}^I_{\mathcal{X}}(F) \neq \emptyset \; \text{ for some } \; F \in \mathcal{F}.$$

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, θ_n) such that $F_n \in \mathcal{F}$ and $\theta_n \in \Theta \setminus \Theta^I_{\mathcal{X}}(F_n)$. 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 $\sigma^2(\theta_n, F_n) = Var_{F_n}\left(\psi(Y, X; \theta_n, F_n)\right)$. Recall that $T_{\mathcal{X}}(\theta_n, F_n) = \sum_{p=1}^{P} E\left[\max\left\{\tau^p(Y^p|X; \theta^p_n, F_n), 0\right\}\cdot \mathbb{I}_{\mathcal{X}}(X)\right]$. By Assumption B5, $\lim_{n\to\infty} P_{F_n}(X \in \mathcal{X}) \geq \underline{p} > 0$ for any sequence $F_n \in \mathcal{F}$. Therefore we will have $T_{\mathcal{X}}(\theta_n, F_n) \longrightarrow 0$ if and only if $P_{F_n}\left(\tau^p(Y^p|X; \theta_n, F_n) > 0 | X \in \mathcal{X}\right) \longrightarrow 0$ 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

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

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

(i) If $T_{\mathcal{X}}(\theta_n, F_n) \to 0$ and $\Delta_{\mathcal{X}}(\theta_n, F_n) \to 0$, then $\sigma(\theta_n, F_n) \to 0$.

(ii) If $T_{\mathcal{X}}(\theta_n, F_n) \to 0$ but $\Delta_{\mathcal{X}}(\theta_n, F_n) \nrightarrow 0$, then $\sigma(\theta_n, F_n) \nrightarrow 0$.

(iii) If $T_{\mathcal{X}}(\theta_n, F_n) \nrightarrow 0$, then $\sigma(\theta_n, F_n) \nrightarrow 0$.

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

$$s_{1,n}(\theta_n, F_n) = \frac{\max\left\{\kappa_n, \sigma(\theta_n, F_n)\right\}}{\sigma(\theta_n, F_n)}, \quad \text{and} \quad s_{2,n}(\theta_n, F_n) = \frac{\sqrt{n}\cdot T_{\mathcal{X}}(\theta_n, F_n)}{\max\left\{\kappa_n, \sigma(\theta_n, F_n)\right\}}$$

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

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

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 $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$ 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 κ_n.

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

Part (a) shows that our procedure will have asymptotic power of 1 whenever $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$ for some $\alpha < 1/2$. Suppose $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$ for some $\alpha > 1/2$. Then we will have $s_2 = 0$ by the bandwidth convergence restrictions of κ_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 α. This will be the case, for example, for any sequence such that $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$ for some $\alpha > 1/2$ but $\lim_{n\to\infty} \Delta_{\mathcal{X}}(\theta_n, F_n) > 0$. On the other hand, our asymptotic power would be zero if $\lim_{n\to\infty} \Delta_{\mathcal{X}}(\theta_n, F_n) = 0$. If $\sigma(\theta_n, F_n) \propto \kappa_n$, the power will be bounded between zero and α. Finally, suppose $T_{\mathcal{X}}(\theta_n, F_n) = O\left(n^{-1/2}\right)$. Our procedure will have asymptotic power of 1 for any such sequence as long as $\lim_{n\to\infty} \Delta_{\mathcal{X}}(\theta_n, F_n) = 0$, as this would yield $s_2 = \infty$. If $\lim_{n\to\infty} \Delta_{\mathcal{X}}(\theta_n, F_n) \neq 0$, 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 $T_{\mathcal{X}}(\theta_n, F_n) = O\left(n^{-1/2}\right)$ will be determined by the limit of the sequence $\Delta_{\mathcal{X}}(\theta_n, F_n)$. Note that –as one should expect– choosing the maximum rate of convergence for κ_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 $T_{\mathcal{X}}(\theta_n, F_n) \to 0$.

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 $K(\psi_1, \ldots, \psi_8) = k(\psi_1) \cdot k(\psi_2) \cdots k(\psi_8)$, where each $k(\cdot)$ is a bias-reducing Biweight-type kernel of order $M = 18$ of the form,

$$k(u) = \sum_{j=1}^{9} c_j \cdot \left(1 - u^2\right)^{2j} \cdot \mathbb{1}\{|u| \leq s\},$$

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 $h_n = c \cdot \widehat{\sigma}(X) \cdot n^{-\alpha_h}$ (note that each X has its own bandwidth), where $\alpha_h = \frac{1}{2M} + \overline{\epsilon}$ and $\overline{\epsilon} = 10^{-}5$. 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

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

This yielded $c \approx 0.2$ and therefore $h_n \approx 0.16 \cdot \widehat{\sigma}(X)$ (for our sample size $n = 954$). Let $\overline{\Omega} = \underset{\theta \in \Theta}{max} \left|\widehat{\sigma}(\theta)\right|$. We used $b_n = c_b \cdot \overline{\Omega} \cdot n^{-\alpha_b}$ where $\alpha_b = \frac{1}{4} + \overline{\epsilon}$ and $\kappa_n = c_\kappa \cdot \overline{\Omega} \cdot \log(n)^{-1}$ with $c_b = 10^{-6}$ and $c_\kappa = 10^{-8}$. We chose these tuning parameters proportional to $\overline{\Omega}$ 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

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

where $\widehat{f}_X^{(0.15)}$ denotes the estimated 15th percentile of the density $\widehat{f}_X$. 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_κ, α_h and α_b used to construct our bandwidths.

References

Andrews, D. and X. Shi (2011a). Inference based on conditional moment inequality models. Working Paper. Yale University.

Andrews, D. and X. Shi (2011b). Nonparametric inference based on conditional moment inequalities. Working Paper. Yale University.

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

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

Aradillas-López, A., A. Gandhi, and D. Quint (2013). Identification and inference in ascending auctions with correlated private values. Forthcoming, Econometrica.

Armstrong, T. (2011a). Asymptotically exact inference in conditional moment inequality models. Working paper. Stanford University.

Armstrong, T. (2011b). Weighted ks statistics for inference on conditional moment inequalities. Working paper. Stanford University.

Asparouhova, E., R. Golanski, K. Kasprzyk, R. Sherman, and T. Asparouhov (2002). Rank estimators for a transformation model. *Econometric Theory 18*, 1099–1120.

Bajari, P., H. Hong, J. Kreiner, and D. Nekipelov (2009). Estimating static models of strategic interaction. *Journal of Business and Economic Statistics, forthcoming*.

Bajari, P., H. Hong, and S. Ryan (2005). Identification and estimation of discrete games of complete information. Working Paper.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*, 1785–1821.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*(4), 889–917.

Berry, S. and E. Tamer (2006). Identification in models of oligopoly entry. In R. Blundell, W. Newey, and T. Persson (Eds.), *Advances in Econometrics. Theory and Applications, Ninth World Congress*, Volume 2, pp. 46–85. Econometric Society Monographs: Cambridge University Press.

Bjorn, P. A. and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with an application to labor force participation. SSWP No. 537, Caltech.

Bresnahan, T. and P. Reiss (1991a). Empirical models of discrete games. *Journal of Econometrics 48*, 57—81.

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

Chernozhukov, V., S. Lee, and A. Rosen (2011). Intersection bounds: Estimation and inference. Cemmap Working Papers, CWP34/11. Institute for Fiscal Studies and Department of Economics, UCL.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*, 1791–1828.

Davis, P. (2006). Estimation of quantity games in the presence of indivisibilities and heterogeneous firms. *Journal of Econometrics 134*, 187–214.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Galichon, A. and M. Henry (2011). Set identification in models with multiple equilibria. *Review of Economic Studies, forthcoming*.

Gowrisankaran, G. and J. Krainer (2011). Entry and pricing in a differentiated products industry: evidence from the atm market. *Rand Journal of Economics 42*(1), 1–22.

Grieco, P. (2012). Discrete choice games with flexible information structure: An application to local grocery markets. Working Paper. Penn State University.

Harsanyi, J. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory 2*(1), 1–23.

Kline, B. and E. Tamer (2010). Bounds for best response functions in binary games. Working paper, Northwestern University, Department of Economics.

Lee, S., K. Song, and Y.-J. Whang (2011). Testing functional inequalities. Cemmap Working Papers, CWP12/11. Institute for Fiscal Studies and Department of Economics, UCL.

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

Lewbel, A. and X. Tang (2012). Identification and estimation of games with incomplete information using excluded regressors. Boston College, mimeo.

Milgrom, P. and R. Weber (1985). Distributional strategies for games with incomplete information. *Mathematics of Operation research 10*(4), 619–632.

Morris, S. (2008). Purification. *The New Palgrave Dictionary of Economics*, 779–782.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*, 780–799.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Pesendorfer, M. and P. Schmidt-Dengler (2008). Asymptotic least squares estimators for dynamic games. *Review of Economic Studies 75*, 901–928.

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

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

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics 37*(3).

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

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

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

Sweeting, A. (2009). The strategic timing of radio commercials: An empirical analysis using multiple equilibria. *RAND Journal of Economics 40*(4), 710–724.

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

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

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