Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities

Kyoo il Kim
University of Minnesota

August 16, 2008

Abstract

This paper studies set estimation and inference of models characterized by conditional moment inequalities. A consistent set estimator and a confidence set are suggested in this paper. Previous studies on set estimation and inference with conditional moment inequalities typically use only a finite set of moment inequalities implied by the conditional moment inequalities.

We consider an alternative strategy that preserves all the information from the conditional moment inequalities. Potentially, this will enable us to obtain a smaller set estimator and a tighter confidence region than those from other methods in the current literature, which lose some information by using only a subset of moment inequalties.

Keywords: Set Estimation, Set Inference, Conditional Moment Inequalities, Partial Identification, U-statistics

JEL Classification: C13, C14

1 Introduction

Initiated by Manski, econometric analyses of incomplete or partially identified models have been of substantial interest over the last decade. Incomplete models can arise in many contexts such as interval-censored observations, sample selection with missing counterfactuals, and games with multiple equilibria. Several estimation, inference methods, and/or specification testing for these models have been proposed including Manski (1990), Horowitz and Manski (1995), Manski and Tamer (2002), Chernozhukov, Hong, and Tamer (2007), Andrews, Berry, and Jia (2004), Rosen (2006), Romano and Shaikh (2006a,b), Guggenberger, Hahn, and Kim (2006), Beresteanu and Molinari (2006), and Andrews and Guggenberger (2007) among others. Existing literature, however, is focused on parametric models or models characterized by unconditional moment inequalities with few exceptions.

Chernozhukov, Hong, and, Tamer (2007) propose an inference method on parameter sets defined as minima of an econometric criterion function in a general setup. They apply their methods to regressions with interval observed data and partially identified method of moments problems but nonlinear models characterized by conditional moment inequalities are not covered.

For game models with multiple equilibria, the model probabilities of taking actions or strategies, implied by necessary conditions are larger than or equal to the corresponding true probabilities. From this relationship, conditional moment inequalities naturally arise. Andrews, Berry, and Jia (2004) deal with these conditional moment inequalities but restrict the number of moments considered to be essentially finite and thus the parameter set of being estimated could be potentially larger than the true parameter set from the original conditional moment inequalities. In other cases, researchers tend to convert conditional moment inequalities to a finite number of unconditional moment inequalities losing some information in the model (for example, see Rosen, 2006). Using all the information from the conditional moments is a secondary issue regarding efficiency in a point-identified model but it is a primary issue in a set-identified model because more restrictions will potentially produce a smaller identified set.

In the extreme, we may lose a point identification simply because we approximate the conditional moment inequalities while the usage of the full information indeed can give us a point identification. This is also true for a model that is characterized by conditional moment equalities whose point-identification is not guaranteed.

Partly motivated from this concern and to fill the gap in the literature, this paper considers a set estimation for models characterized by conditional moment inequalities. In our approach, we still convert the conditional moment inequalities into unconditional ones but preserving all the information in the model in the spirit of Dominguez and Lobato (2004). Our estimator is similar with the minimum distance estimator proposed by Khan and Tamer (2006) in the context of randomly censored regression models. Their identification strategy and estimation are mainly focused on conditional moment inequalities that yields a point identification. On the contrary, we are interested in the models where a set of conditional moment inequalities may be incomplete for a point identification. We also intend to provide asymptotic theories that justify our proposed set estimator and confidence set. Our proposal can also be applied to the case that a model is characterized by conditional moment equalities that do not necessarily produce a point-identification. We also consider the case that the identified set is indeed a singleton.

This paper's objectives are to provide a consistent set estimator and a confidence set for models characterized by conditional moment inequalities and to provide an asymptotic justification of Chernozhukov, Hong, and Tamer (2007, CHT) or Romano and Shaikh (2006, 2008, RS)'s approach for such models. Differently from CHT or RS, it turns out that we need to expand our sample criterion function into a sum of U-processes and obtain uniform convergences and uniform convergences with rates using the limit theorems of several degenerate U-processes. We establish the asymptotic validity of our set estimator and the confidence set using functional limit theorems of U-processes and uniform bounds in non-Gaussian limit distribution of degenerate U-processes due to Nolan and Pollard (1987), Sherman (1994a), Bentkus and Gtöze (1999).

This paper is organized as follows. In Section 2, we introduce the model and the set estimator. Several examples that fit into our model are discussed in Section 3. Section 4 derives the consistency and the convergence rate of the proposed estimator. In Section 5, we provide a consistent confidence set. In Section 6, we conclude. Technical details and mathematical proofs are presented in the appendix.

2 Model and Estimation

We consider econometric models characterized by the following conditional moment equalities and inequalities:

$$E\left[m_l(y_i, x_{1i}; \theta)|x_i\right] \leq 0 \text{ a.s. } l = 1, \ldots, d_{m_1} \tag{1}$$

$$E\left[m_l(y_i, x_{1i}; \theta)|x_i\right] = 0 \text{ a.s. } l = d_{m_1}+1, \ldots, d_{m_1}+d_{m_2} \tag{2}$$

where $y_i \in \mathcal{Y} \subset \mathbb{R}^{d_y}$, $x_{1i} \subset x_i \in \mathcal{X} \subset \mathbb{R}^{d_x}$, $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$, m_j's are $1 \times d_m (= d_{m_1} + d_{m_2})$ vector of known functions such that $m(\cdot) = (m_1(\cdot), \ldots, m_{d_m}(\cdot))'$. $\mathcal{Y}$ and $\mathcal{X}$ denote the support of the distribution of y_i and x_i, respectively. The support $\mathcal{X}$ can be discrete or continuous. Here we are more interested in the continuous support since it generates infinite number of moment inequalities and equalities implied by (1) and (2), respectively.

We let y_i denote endogenous variables including dependent variables and endogenous regressors, x_{1i} and x_i denote exogenous variables. Therefore, our model can nest IV models where $x_i \setminus x_{1i}$ becomes excluded instrumental variables. In particular, the moment inequality conditions of (1) arise in many cases including models with interval measured y_i and game theoretic models where (1) denotes a set of necessary conditions that characterize equilibria of the game. However, models characterized by inequality moment conditions do not necessarily result in set identification. Indeed, Khan and Tamer (2006) obtain conditions under which the moment inequality conditions of (1) induce a point-identification. We also note that in the models of equality constraints, we typically consider the case $d_m \geq d_\theta$ but here we allow for $d_m < d_\theta$ since θ is not necessarily point-identified. The inequality "$\geq$" is taken componentwise throughout this paper. We will use $E^{x_j,x_k}[\cdot]$ to denote the expectation of $[\cdot]$ fixing x_j and x_k in $[\cdot]$. Similarly $Var^{x_j,x_k}[\cdot]$ denotes the variance of $[\cdot]$ fixing x_j and x_k in $[\cdot]$.

The parameters of interest will be defined as a collection of parameters that satisfy the condition of (1) and (2), denoted by

$$\Theta_0 = \{\theta \in \Theta : (1) \text{ and } (2) \text{ hold for } \theta, \text{ a.e. } x \in \mathcal{X}\}. \tag{3}$$

Now define

$$H_l(\theta, t_1, t_2) = E\left[m_l(y_i, x_{1i}; \theta)\mathbf{1}[t_1 < x_i \leq t_2]\right] \text{ with } t_1 < t_2 \text{ for } l = 1, \ldots, d_m$$

and let $H(\theta, t_1, t_2) = (H_1(\cdot), \ldots, H_{d_m}(\cdot))'$. Form an extended measure result from Billingsley (1995, Theorem 11.3), we note that the conditional moment condition (1) and (2) are equivalent to the following set of unconditional moment conditions indexed by t_1 and t_2,

$$H_l(\theta, t_1, t_2) \leq 0 \text{ for almost all } t_1 < t_2 \in \mathcal{X} \times \mathcal{X},\ l = 1, \ldots, d_{m_1} \tag{4}$$

$$H_l(\theta, t_1, t_2) = 0 \text{ for almost all } t_1 < t_2 \in \mathcal{X} \times \mathcal{X},\ l = d_{m_1} + 1, \ldots, d_m \tag{5}$$

In other words, the alternative characterization of (4) and (5) does not lose any information about θ_0, contained in (1) and (2). This idea was proposed in several interesting works including Dominguez and Lobato (2004) for the conditional moment equality models. Therefore, we can use the above moment equalities and inequalities instead of (1) and (2) in constructing the population criterion function that determines the identified set. This equivalence is formally proved in Andrews and Shi (2008).

The alternative characterization also implies that when a moment equality or a moment inequality of (1)-(2) is violated, we should be able to find nonnegligible mass of (t_1, t_2)'s that make (4) or (5) violated. One may think we can construct the moment equations using one-sided indicator functions (say $H_l(\theta, t) = E\left[m_l(y_i, x_{1i}; \theta)\mathbf{1}[x_i \leq t]\right]$) as they originally appear in Dominguez and Lobato (2004) but here we indeed require the two-sided indicator functions due to the inequality conditions. The reason is that the alternative moment conditions based on the one-sided indicator functions cannot detect the violation of the moment inequalities when the moment inequalities hold for significant mass of $x_i \leq t$. To fix the idea, suppose that for $\widetilde{t} < t$, $H_l(\theta, \widetilde{t}) = E\left[m_l(y_i, x_{1i}; \theta)\mathbf{1}[x_i \leq \widetilde{t}]\right] = -3$ but $E\left[m_l(y_i, x_{1i}; \theta)\mathbf{1}[\widetilde{t} < x_i \leq t]\right] = 1$. Then, one would obtain

$$H_l(\theta, t) = H_l(\theta, \widetilde{t}) + E\left[m_l(y_i, x_{1i}; \theta)\mathbf{1}[\widetilde{t} < x_i \leq t]\right] = -2$$

and conclude the moment inequality holds for t. But this case obviously violates the moment inequality for some x_i's inside the interval $[\widetilde{t}, t]$. Therefore, we should use the two-sided indicator functions such that we can find $H_l(\theta, \widetilde{t}, t)$ is positive and conclude the moment inequality is violated.

We further define

$$\sigma_l^2(\theta, t_1, t_2) = Var[m_l(y_i, x_{1i}; \theta)\mathbf{1}[t_1 < x_i \leq t_2]] \text{ with } t_1 < t_2 \in \mathcal{X} \times \mathcal{X},\ l = 1, \ldots, d_m$$

and assume that $0 < c_L(x_j, x_k) < \sigma_l^2(\theta, x_j, x_k) < c_U(x_j, x_k)$, $\forall\theta \in \Theta$ a.e. $x_j \times x_k \in \mathcal{X} \times \mathcal{X}$ with $x_j < x_k$ for some bounded constant $c_L(x_j, x_k)$ from below and $c_U(x_j, x_k)$ from above. Then, we define the population criterion function as

$$Q(\theta) = E\left[\sum_{l=1}^{d_{m_1}} \left(\frac{H_l(\theta, x_j, x_k)}{\sigma_l(\theta, x_j, x_k)}\right)^2 \mathbf{1}[H_l(\theta, x_j, x_k) \geq 0] + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{H_l(\theta, x_j, x_k)}{\sigma_l(\theta, x_j, x_k)}\right)^2\right]. \tag{6}$$

Normalizing the moment functions by their individual standard errors, we can mitigate possible excess influences by a small subset of moment functions. This normalization also makes the resulting statistic as a sample analogue of (6) invariant to rescaling of the moment conditions. Our criterion function corresponds to a class of criterion function named *modified method of moments test function* in Andrews and Guggenberger (2007) and Andrews and Soares (2007).

Then, by construction, the identified set Θ_0 defined in (3) is equivalent to

$$\Theta_0 = \left\{\theta \in \Theta : Q(\theta) = \inf_{\theta \in \Theta} Q(\theta)\right\}.$$

From this observation, we obtain our set estimator as a collection of θ that minimizes a sample analogue of $Q(\theta)$:

$$\widehat{Q}_n(\theta) = \frac{1}{n(n-1)} \sum_{j \neq k} \left\{\sum_{l=1}^{d_{m_1}} \frac{\widehat{H}_l(\theta, x_j, x_k)^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2} \mathbf{1}[\widehat{H}_l(\theta, x_j, x_k) \geq 0] + \sum_{l=1+d_{m_1}}^{d_m} \frac{\widehat{H}_l(\theta, x_j, x_k)^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2}\right\} \tag{7}$$

where we use $\widehat{H}_l(\theta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^n m_l(y_i, x_{1i}; \theta)\mathbf{1}[x_j < x_i \leq x_k]$ and

$$\widehat{\sigma}_l^2(\theta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^n (m_l(\theta, y_i, x_{1i})\mathbf{1}[x_j < x_i \leq x_k])^2 - \widehat{H}_l(\theta, x_j, x_k)^2.$$

Note that a constraint of (1) is violated when $E[m_l(y_i, x_{1i}; \theta)|x_i] > 0$ for some $l = 1, \ldots, d_{m_1}$ and some x_i and a constraint of (2) is violated when $E[m_l(y_i, x_{1i}; \theta)|x_i] \neq 0$ for some $l = d_{m_1} + 1, \ldots, d_m$ and x_i. Thus, $\widehat{Q}_n(\theta)$ is the squared magnitude of the constraint violations summed over all constraints. It is worthwhile to add comments regarding the criterion function. For models with moment inequalities only, one may use different criterion functions than that of (6). For example, one may let $Q^{(1)}(\theta) = E\left[\sum_{l=1}^{d_m} H_l(\theta, x_j, x_k)\mathbf{1}[H_l(\theta, x_j, x_k) \geq 0]\right]$. This $Q^{(1)}(\theta)$ is in line with the criterion function chosen by Andrews, Berry, Jia (2004). It turns out that a convergence rate of a set estimator[1] depends on the choice of a criterion function. For example, if we use $Q^{(1)}(\theta)$ instead of $Q(\theta)$, we do not obtain the virtual $n^{-1/2}$ convergence rate obtained in Section 4. We also note

[1] So does the rate of slackness variable tending to zero and making sure that the resulting set estimator containes the true identified set with probability approaching to one.

that an estimator obtained from (7) does not achieve the semiparametric efficiency bound when it is point-identified as noted by Dominguez and Lobato (2004).

As an alternative to exploit all the information contained in (1), one may use a sequence of nested moment conditions implied by (1) such that the number of moment conditions goes to infinity as the sample size goes to infinity. This sort of idea has been popularly used in point-identification problem of models characterized by conditional moment models. The difference is that we use such a sequence to achieve the semiparametric efficiency bound in the point-identified case but in the set-identified case, we use such a sequence to reduce the identified set. It is obvious that for any finite number of moment conditions, the identified set from such moment conditions can not be smaller than that from the original conditional moments.

One can also let

$$\widetilde{Q}_n(\theta) = \widehat{Q}_n(\theta) - \inf_{\theta\in\Theta} \widehat{Q}_n(\theta).$$

This improves powers of tests based on such criterion function as discussed in CHT and Mikusheva (2006) when $\inf_{\theta\in\Theta} \widehat{Q}_n(\theta) \neq 0$ in a finite sample.

Now we are ready to define our set estimator. We proceed under two different scenarios regarding the identified set. For models with moment inequalities only, when Assumption 2.1 below holds, we often require less stringent conditions that governs the large sample properties of a set estimator. Andrews, Berry, and Jia (2004) utilizes this Assumption 2.1 and thus they can avoid using a slackness condition (originally devised by Manski and Tamer (2002) in a set estimation context, so called, ϵ_n-maximization) in their construction of the set estimator and the confidence intervals. Now let int(A), $\partial(A)$, and cl(A) denote the interior, the boundary, and the closure of a set A, respectively.

Assumption 2.1 *Either (i)* $\Theta_0 = \{\theta_0\}$ *or (ii) (a)* $\Theta_0 = cl(int(\Theta_0))$ *and (b) for all* $\theta_0 \in int(\Theta_0)$, $E[m_j(y_i, x_{1i}; \theta_0)|x_i] < 0$ *a.s.,* $j = 1, \ldots, d_{m_1}$

Assumption 2.1 (ii) implies that Θ_0 has a non-empty interior and does not contain isolated points, lines, or hyperplanes. Assumption 2.1 (ii) is one of cases where the degeneracy condition noted by CHT is satisfied. We, therefore, present two different versions of set estimator depending on whether we satisfy Assumption 2.1 or not.

Our set estimator is given by

$$\widehat{\Theta}_{n,0} = \left\{\theta \in \Theta : \widehat{Q}_n(\theta) = \inf_{\theta\in\Theta} \widehat{Q}_n(\theta)\right\} \tag{8}$$

when Assumption 2.1 holds. When Assumption 2.1 does not hold and thus the model allows the true parameter set to include some disjoint points, lines, or hyperplanes, we define our set estimator as

$$\widehat{\Theta}_n(\widehat{c}_n) = \left\{\theta \in \Theta : n\widehat{Q}_n(\theta) \leq \widehat{c}_n\right\} \tag{9}$$

where $\widehat{c}_n \to \infty$ but $\widehat{c}_n/n$ goes to zero but with a rate *slower* than the uniform convergence rate of $\widehat{Q}_n(\theta)$ to $Q(\theta)$ as Manski and Tamer (2002). Typically we let $\widehat{c}_n$ go to infinity as slow as possible while preserving desirable asymptotic behavior of $\widehat{\Theta}_n(\widehat{c}_n)$ since a larger $\widehat{c}_n$ results in a larger estimated set or a larger confidence set. CHT establishes that we can let $\widehat{c}_n = O(\log(n))$. We also achieve the same rate with CHT. This slackness condition makes sure that the estimated set covers the true set with probability tending to one.

3 Examples

Here we discuss two examples of interest that fit into our model. The interval observed outcome example by Manski and Tamer (2002) is nested in our model where we do not observe y_i but we observe its lower and upper bounds as $y_{Li} \leq y_i \leq y_{Ui}$. We can let

$$m_1(\theta, y_{Li}, x_i|x_i) = E[y_{Li}|x_i] - x_i'\theta \leq 0 \text{ and } m_2(\theta, y_{Ui}, x_i|x_i) = x_i'\theta - E[y_{Ui}|x_i] \leq 0.$$

We want to stress the difference between our model and that of CHT or Romano and Shaikh (2006, 2008) for this example. Since CHT only allows for unconditional moment inequalities, their estimator uses a finite number of moment conditions that are implied by the conditional moment inequalities. Similarly Romano and Shaikh (2006, 2008) assume that x_i follows a multinomial distribution so that we have a finite number of moment inequality conditions. We conjecture that our approach will produce less conservative set estimator and confidence set (CS) than those of CHT and Romano and Shaikh (2006, 2008).

Our model also can handle game theoretic models with multiple equilibria. Multiple equilibria arise often in discrete games such as entry-exit games (Bresnahan and Reiss (1990, 1991), Tamer (2003)). In particular, we consider the model where some asymptotic inequalities may define a region of parameters rather than a single point in the parameter space. By definition, when there are multiple equilibria, there exist regions of unobservables that are consistent with the necessary conditions for more than one equilibrium. Therefore, the probability implied by the necessary condition for a given event is greater than or equal to the true probability of the event. The set estimation and inference of this type of model utilizes these inequality conditions. Suppose y_i denote the observed outcomes of a discrete game. Then, under the possibility of multiple equilibria, we have

$$P[y_i = y|x_i, \theta] - P[y_i = y|x_i] \geq 0 \text{ for } \forall y \in \mathcal{Y} \tag{10}$$

where $\mathcal{Y}$ denotes the set of all possible equilibrium outcomes of the game. $P(y_i = y|x_i, \theta)$ denotes the choice probability implied by a parameterized game model and $P[y_i = y|x_i]$ denotes the true choice probability. We can rewrite (10) as

$$E[m(y_i, x_i; \theta)|x_i] = E[y_i - P(y_i|x_i, \theta)|x_i] \leq 0$$

and so it fits into our model. Andrews, Berry, and Jia (2004) propose a set estimator based on (10) but consider weaker moment inequalities implied by (10) through grouping y_i and x_i into a finite set and so we have a finite number of moment conditions. We again conjecture that our proposal will produce less conservative set estimator and CS than those of Andrews, Berry, and Jia (2004). Ciliberto and Tamer (2007) and Beresteanu, Molchanov, and Molinari (2008, BMM) exploits more inequalities implied by equilibrium conditions of discrete games and obtain sharper bounds. Heuristically speaking, BMM strengthens the moment inequality conditions of (10) to

$$P[y_i \in K|x_i, \theta] - P[y_i \in K|x_i] \geq 0 \text{ for } \forall K \subset \mathcal{Y}. \tag{11}$$

Since the set of all possible subsets of $\mathcal{Y}$ include all the y's in $\mathcal{Y}$, the resulting identified set from (11) is sharper than that from (10). In particular BMM shows that the identified set using their moment inequalities is the *sharp* one described by Berry and Tamer (2007) in the sense that the set of identified parameters are consistent with the data and the model. Interestingly the sharp identified set is also given by a set of conditional moment inequalities. Therefore, the sharp identified set is obtained only when we utilize all the moment inequalities implied by the conditional moment inequalities. In other words, even though one obtains an identified set from (11), the identified set may not be sharp if he discretizes X and uses only a finite number of moment inequalities implied by (11).

4 Consistency and Convergence Rate

We derive the consistency and the convergence rate of our set estimator defined in (9) extending CHT to the conditional moments inequalities models. We use the Hausdorff metric as the distance measure between two sets. The Hausdorff metric is defined for two sets A and B whose elements are in $\mathbb{R}^{d_\theta}$:

$$\begin{aligned} d(A, B) &= \max\{\rho(A|B), \rho(B|A)\}, \\ \text{where } \rho(A|B) &= \sup_{a \in A} \inf_{b \in B} \|a - b\| \text{ and } \rho(B|A) = \sup_{b \in B} \inf_{a \in A} \|a - b\|. \end{aligned}$$

To show the consistency, we need to prove two conditions: one is $\rho(\widehat{\Theta}_n(\widehat{c}_n)|\Theta_0) \to 0$, which means the estimated set is included in the true set w.p.a.1 and the other condition is $\rho(\Theta_0|\widehat{\Theta}_n(\widehat{c}_n)) \to 0$, which means the true set is included in the estimated set w.p.a.1. In proving the second part, we typically rely on the slackness condition.

When Assumption 2.1 (i) holds, we note that the slackness condition is not needed since $\rho(\Theta_0|\widehat{\Theta}_{n,0}) = \rho(\{\theta_0\}|\widehat{\Theta}_{n,0}) \le \rho(\widehat{\Theta}_{n,0}|\{\theta_0\}) \to 0$, i.e., showing the estimated set is included in the true set w.p.a.1 is sufficient for consistency. When Assumption 2.1 (ii) holds, the slackness condition is not necessary, either. In particular, Assumption 2.1 (ii) implies the degeneracy condition in CHT.

Now we present the consistency and the convergence rate of the set estimator. It turns out that we can expand our sample criterion function into a sum of U-processes and obtain uniform convergences and uniform convergence rates using the limit theorems of several degenerate U-processes, which are useful to obtain the consistency and convergence rate of our set estimator. As we require some regularities on the complexity of a functional space to obtain the uniform convergence result in the empirical process, we also need a similar regularity in U-processes. Such a condition can be found in Nolan and Pollard (1987), Pakes and Pollard (1988), and Sherman (1994a). We, therefore, introduce the following two concepts, *Envelope* and *Euclidean*:

Definition 4.1 *For a class of functions indexed by $\theta \in \Theta$, $\mathcal{F} = \{f(\cdot,\theta) : \theta \in \Theta\}$, we say $\mathcal{F}$ has an* envelope $\mathbb{F}$ *if*

$$\sup_{\theta\in\Theta} \|f(\cdot,\theta)\| \le \mathbb{F}(\cdot)$$

Definition 4.2 *Let $D(\varepsilon, d_\mu, \mathcal{F}, \mathbb{F})$ be the packing number of $\mathcal{F}$ with radius ε and a pseudometric d_μ where μ denotes a measure on $\mathcal{X}^k = \mathcal{X} \times \ldots \times \mathcal{X}$. $\mathcal{F}$ is called Euclidean for the envelope $\mathbb{F}$ if there exist constants A and V such that for a measure μ satisfying $0 < \mu\mathbb{F}^2 < \infty$,*

$$D(\varepsilon, d_\mu, \mathcal{F}, \mathbb{F}) \le A\varepsilon^{-V}, \quad \text{for } 0 < \varepsilon \le 1,$$

where, for $f, g \in \mathcal{F}$,

$$d_\mu(f,g) = [\mu\,|f-g|^2/\mu\mathbb{F}^2]^{1/2}.$$

The following assumptions are needed to obtain the consistency and the convergence rate of our set estimator. We let $\mathbf{1}_{jik} = \mathbf{1}[x_j < x_i \le x_k]$ and '$\Longrightarrow$' denotes a weak convergence. We also define $\mathbf{1}_{l,jk}(\theta) = \mathbf{1}[H_l(\theta, x_j, x_k) \ge 0]$. We start with these higher level assumptions and show later that how we can verify these conditions using more primitive conditions.

Assumption 4.1 *(a) Θ is nonempty compact subset of $\mathbb{R}^{d_\theta}$; (b) $\{y_i, x_i\}_{i=1}^n$ are iid;*
(c) For each $l = 1, \ldots, d_m$, the functional space

$$\begin{aligned}\mathcal{H}_l &\equiv \left\{H_l(\theta,x_j,x_k)^2 - E[H_l(\theta,x_j,x_k)^2] : \theta\in\Theta\right\} \text{ and} \\ \mathcal{H}_l^\sigma &\equiv \left\{\left(\frac{H_l(\theta,x_j,x_k)}{\sigma_l(\theta,x_j,x_k)}\right)^2 - E\left[\left(\frac{H_l(\theta,x_j,x_k)}{\sigma_l(\theta,x_j,x_k)}\right)^2\right] : \theta\in\Theta\right\}\end{aligned}$$

indexed by $\theta \in \Theta$ *are Euclidean with envelopes* $\mathbb{H}_l$ *and* $\mathbb{H}_l^\sigma$*, respectively;*
(d) For each $l = 1, \ldots, d_m$*, the functional spaces*

$$\mathcal{M}_l \equiv \{m_l(y_i, x_{1i}; \theta) m_l(y_t, x_{1t}; \theta) \mathbf{1}_{jik} \mathbf{1}_{jtk}] : \theta \in \Theta\} \ \ and$$

$$\mathcal{M}_l^\sigma \equiv \left\{ \frac{m_l(y_i, x_{1i}; \theta)}{\sigma_l(\theta, x_j, x_k)} \frac{m_l(y_t, x_{1t}; \theta)}{\sigma_l(\theta, x_j, x_k)} \mathbf{1}_{jik} \mathbf{1}_{jtk} - \frac{H_l(\theta, x_j, x_k)^2}{\sigma_l(\theta, x_j, x_k)^2}] : \theta \in \Theta \right\}$$

indexed by $\theta \in \Theta$ *are Euclidean with envelopes* $\mathbb{M}_l$ *and* $\mathbb{M}_l^\sigma$*, respectively;*
(e) (P-Donsker property) For all $l = 1, \ldots, d_m$*,*

$$\mathbb{G}_{l,n}(\theta, t_1, t_2) \equiv \sqrt{n} \left(\widehat{H}_l(\theta, t_1, t_2) - E[H_l(\theta, t_1, t_2)] \right) \Longrightarrow \mathcal{G}_l(\theta, t_1, t_2)$$

and $\mathcal{G}_l(\theta, t_1, t_2)$ *is a mean zero Gaussian process with a.s. continuous paths,* $Var[\mathcal{G}_l(\theta, t_1, t_2)] > 0$ *for each* $\theta \in \Theta$ *and* $t_1 < t_2$*;*
(f) $\widehat{\sigma}_l^2(\theta, x_j, x_k) = \sigma_l^2(\theta, x_j, x_k) + O_P(n^{-1/2})$ *for all* $\theta \in \Theta$ *and* $0 < c_L(x_j, x_k) < \sigma_l^2(\theta, x_j, x_k) < c_U(x_j, x_k)$*,* $\forall \theta \in \Theta$ *for some bounded constant* $c_L(x_j, x_k)$ *and* $c_U(x_j, x_k)$*, a.e.* $x_j < x_k \in \mathcal{X} \times \mathcal{X}$*;*
(g) There exist positive constants C *and* δ*, and a subset* $\widetilde{B} \subset \mathcal{X} \times \mathcal{X}$ *such that for all* $\theta \in \Theta$*,*

$$\sum_{l=1}^{d_{m_1}} H_l(\theta, \widetilde{x}_j, \widetilde{x}_k)^2 \mathbf{1}[H_l(\theta, \widetilde{x}_j, \widetilde{x}_k) \geq 0] + \sum_{l=d_{m_1}+1}^{d_m} H_l(\theta, \widetilde{x}_j, \widetilde{x}_k)^2 \geq C \cdot (d(\theta, \Theta_0) \wedge \delta)^2$$

for any $(\widetilde{x}_j, \widetilde{x}_k) \in \widetilde{B}$ *and* $\Pr((x_j, x_k) \in \widetilde{B}) > 0$*.*

Assumption 4.1 (c) and (d) are useful to derive the uniform convergence of the degenerated U-processes using the maximal inequalities obtained by Nolan and Pollard (1987) and Sherman (1994a) and it is not difficult to show these assumptions hold for particular models due to Nolan and Pollard (1987) and Pakes and Pollard (1989). In some cases, the moment conditions are given by indicator functions. Then, the Euclidean conditions hold trivially as in Khan and Tamer (2006). For general moment functions, we can use the result in Lemma 2.13 in Pakes and Pollard (1989).

Lemma 4.1 (Pakes and Pollard (1989)) *For a class of functions* $\mathcal{F}$*, if there exists an* $\alpha > 0$ *and a nonnegative function* $b(\cdot)$ *such that*

$$\left| f(\cdot, \theta) - f(\cdot, \theta') \right| \leq b(\cdot) \left\| \theta - \theta' \right\|^\alpha \qquad for\ \theta\ and\ \theta' \in \Theta,$$

then $\mathcal{F}$ *is Euclidean for the envelope* $|f(\cdot, \underline{\theta})| + Mb(\cdot)$*, where* $\underline{\theta}$ *is an arbitrary point in* Θ *and* $M = (2\sqrt{d_\theta} \sup_\Theta \left\| \theta - \theta' \right\|^\alpha)$*.*

The above lemma is quite useful and easy to verify for many class of moment functions. One can use the above Lemma 4.1 to show the Euclidean conditions hold for examples in Section

3. For instance, in the interval observed outcome example, we have $m_1(\theta, \cdot) = y_{Li} - x_i'\theta$ and $m_2(\theta, \cdot) = x_i'\theta - y_{Ui}$ and they are continuous in θ. Note that

$$H_1(\theta, x_j, x_k)^2 = \left\{E^{x_j,x_k}\left[(y_{Li} - x_i'\theta)\mathbf{1}_{jik}\right]\right\}^2$$

and that

$$\left|H_1(\theta, x_j, x_k)^2 - H_1(\theta', x_j, x_k)^2\right| \leq \left|H_1(\theta, x_j, x_k) + H_1(\theta', x_j, x_k)\right| \|x_i\| \left\|\theta - \theta'\right\|.$$

We, thus, conclude $\left\{H_1(\theta, \cdot)^2 - E[H_1(\theta, \cdot)^2] : \theta \in \Theta\right\}$ is Euclidean when $E\left[(y_{Li} - x_i'\theta)^2\right] < \infty$ uniformly over $\theta \in \Theta$.

In the example of the games with multiple equilibria, we have

$$H_l(\theta, x_j, x_k)^2 = \{E^{x_j,x_k}[(y_{li} - P(y_{li}|x_i, \theta))\, \mathbf{1}_{jik}]\}^2$$

and thus the Euclidean property obviously holds with a constant envelope since $H_l(\theta, x_j, x_k)^2 \leq 1$ uniformly.

Assumption 4.1 (e) is standard and sufficient conditions for a P-Donsker class are well-known in the literature. Assumption 4.1 (g) is used to obtain the convergence rate result. This condition means that when θ is bounded away from Θ_0, the moment equations tend to be bounded below proportional to the distance of θ from the identified set for nonnegligible mass of (x_j, x_k)'s. A similar regularity condition is often required in the point-identification case to obtain a convergence rate of an estimator. Here C and δ can depend on x_j and x_k but we can let $C \equiv \min_{x_j, x_k \in \widetilde{B}} C(x_j, x_k) > 0$ and $\delta \equiv \min_{x_j, x_{k \in \widetilde{B}}} \delta(x_j, x_k) > 0$ w.l.o.g.

Now we provide our main theorems. We first derive the consistency and the convergence rate result when Assumption 2.1 does not hold.

Theorem 4.1 *Suppose Assumption 4.1 (a)-(f) hold. Now let* $\widehat{c}_n \geq \sup_{\theta \in \Theta_0} n\widehat{Q}_n(\theta)$ *with probability approaching to one but* $\frac{\widehat{c}_n}{n} \longrightarrow_p 0$. *Then w.p.a.1, we have* $d(\widehat{\Theta}_n(\widehat{c}_n), \Theta_0) = o_P(1)$. *Further suppose Assumption 4.1 (g) holds, then* $d(\widehat{\Theta}_n(\widehat{c}_n), \Theta_0) = O_P\left(n^{-1/2} \vee (\widehat{c}_n/n)^{-1/2}\right)$.

Proof. See Appendix A.1. ■

Now we consider the case that Assumption 2.1 holds. Since Assumption 2.1 (ii) is satisfied for models characterized by moment inequalities only, we will let $d_{m_2} = 0$.

Theorem 4.2 *Suppose Assumption 2.1 holds and Assumption 4.1 (a)-(f) hold. Then,* $d(\widehat{\Theta}_{n,0}, \Theta_0) = o_P(1)$. *Further suppose Assumption 4.1 (g) holds. Then, we have* $d(\widehat{\Theta}_{n,0}, \Theta_0) = O_P(n^{-1/2})$.

Proof. See Appendix A.2. ■

5 Confidence Set

Now we study a confidence set (CS) for the parameters of interest for models characterized by conditional moment inequalities of (1) and (2). Current literature has taken one of two approaches or both. First we can construct a confidence set for the identified set, Θ_0 as in CHT and Romano and Shaikh (2008). We can also construct a confidence set for a true parameter $\theta_0 \in \Theta_0$ with a specified coverage probability. The true parameter may not be a singleton. The second type of confidence set is in spirit of Imbens and Manski (2004). Typically the latter type of CS provides less conservative inference when the true parameter is still of our interest. Imbens and Manski (2004) show that a confidence interval (CI) for a true parameter is shorter than that of the identified interval. The former corresponds to two-sided CI while the former does to two-sided CI. We focus on CS's for the true value θ_0 as Andrews and Guggenberger (2007) and Andrews and Soares (2007). We take this approach since an empirical researcher is more interested in the true value rather than the identified set when she answers to policy questions from the model.

We construct the confidence set by inverting Anderson-Rubin type test statistic with the critical value obtained by the subsampling procedure as in Romano and Shaikh (2008) and Andrews and Guggenberger (2007). Then, we provide a set of conditions under which such CS is justified asymptotically for models characterized by (1)-(2) and estimated by (8) or (9). We will use the notation $\Theta_0(P)$ and $Q(\theta, P)$ to explicitly denote their dependence on the distribution of the observed data, P.

In this paper, we first provide the pointwise consistency of the CS. Then we note that showing the consistency of the CS uniformly over the true distribution generating the data is potentially important in the models characterized by moment inequalities. Its importance has been well discussed in Romano and Shaikh (2006, 2008) and Andrews and Guggenberger (2007 and their other works). This is because the pointwise asymptotics does not account for the inherent discontinuity in these models where the asymptotic distribution is determined only by the binding moment conditions under a particular data-generating distribution, P. To be precise, let $\mathcal{I}$ denote the set of l indices for which the moment condition (1) holds with equality under a particular P.

Then, the asymptotic behavior of the CS is determined by the asymptotic behavior of

$$\frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \sum_{l \in \mathcal{I}} \left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)} \right)^2 \mathbf{1}[\widehat{H}_l(\theta, x_j, x_k) \geq 0] + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)} \right)^2 \right\}.$$

Note that the index set $\mathcal{I}$ can depend on P as well as θ. Given θ, $\mathcal{I}$ includes the index l if the l-th moment of (1)-(1) holds with equality for some $x_i \in \mathcal{X}$, i.e., if the l-th moment of (1)-(1) holds with equality for some $x_j < x_k \in \mathcal{X} \times \mathcal{X}$.

We will let $\mathcal{I} = \mathcal{I}(P)$. Evidently, $\mathcal{I}(P)$ is not continuous in P. Therefore, not having the uniform consistency means the usual pointwise asymptotic approximation can lead to very poor testing or inference result even when the sample size is very large.

Romano and Shaikh (2006,2008) provide the uniform asymptotic validity of subsampling methods under higher level assumptions and provide primitive conditions for several interesting examples including one-sided, two-sided mean, regression with interval outcomes, unconditional moment inequalities while Andrews and Guggenberger (2007) provides more general results. However, a valid inference for the conditional moment inequalities has not been fully provided yet either pointwise in P or uniformly in P.

The $1-\alpha$ CS for the true parameter is given by

$$CS_{\theta_0,n}(1-\alpha) = \{\theta \in \Theta : n\widehat{Q}_n(\theta) \leq \widehat{d}_n(\theta, 1-\alpha)\}$$

with a subsampled critical value $\widehat{d}_n(\theta, 1-\alpha)$.

Let $N_n = \binom{n}{b}$ and let $\widehat{Q}_{n,b,i}(\theta)$ denote the statistic $\widehat{Q}_n(\theta)$ evaluated at the i-th subsample of size b. The critical value is obtained as the $(1-\alpha)$ quantile of the N_n numbers of subsampled criterion functions such that

$$\widehat{d}_n(\theta, 1-\alpha) = \inf\left\{x : \frac{1}{N_n}\sum_{1\leq i\leq N_n} \mathbf{1}[b\widehat{Q}_{n,b,i}(\theta) \leq x] \geq 1-\alpha\right\}. \tag{12}$$

The CS for the true parameter will provide a less conservative inference than the CS for the identified set. However, it is potentially more costly in terms of computation. This is because obtaining the critical values $\widehat{d}_n(\theta, 1-\alpha)$, we need to run the subsampling procedure for each trial value of $\theta \in \Theta$.

We will provide the consistency of these confidence sets both pointwise in P and uniformly in P in the following sections.

To show the asymptotic validity of the CS using the subsampling method, we need to obtain the asymptotic distribution of $n\widehat{Q}_n(\theta)$. In that, we first expand $n\widehat{Q}_n(\theta)$ to the fourth-order U-process and decompose the U-process into a sum of degenerate U-processes up to the order of two, following Sherman (1994a) and Serfling (1980). We let $\overline{\mathcal{I}} = \mathcal{I} \cup \{l : 1 + d_{m_1} \leq l \leq d_m\}$. We find that for $\theta \in \partial(\Theta_0(P))$, $n\widehat{Q}_n(\theta)$ follows a nonstandard distribution. The distribution is approximated by an infinite mixture of independent, centered chi-square distributions such that

$$n\widehat{Q}_n(\theta) = \textstyle\sum_{a=1}^{\infty} \overline{q}_a(\theta)\left(\chi_a^2 - 1\right) + o_P(1) \tag{13}$$

where $\overline{q}_1(\theta), \overline{q}_2(\theta), \ldots$ denote eigenvalues of the Hilbert-Schmidt operator associated with the kernel function, $\phi(w_u, w_v; \theta)$, defined below in (15), which constitutes the U-process, satisfying $\sum_{a=1}^{\infty} \overline{q}_a^2(\theta) < \infty$ (see Gregory (1977), Neuhaus (1977), Nolan and Pollard (1988), and Bentkus and Gtöze (1999)).

To prove the pointwise consistency of the subsampled CS, we need to establish (i) the uniform convergence of the distribution in θ and also (ii) the uniform convergence of the distribution function itself. The uniform consistency in P will be obtained by restricting P to a set $\boldsymbol{P}$ satisfying additional regularity conditions.

5.1 Uniform convergence of distribution in θ

We first obtain the uniform convergence in distribution uniformly over $\theta \in \Theta$ due to the functional limit theorem by Nolan and Pollard (1988), which is an U-process analog of the limit theorem for empirical processes. We start with introducing additional definitions and notations. Let $W = Y \times X$. Define the Hilbert-Schmidt operator $\mathbb{H}$ as

$$(\mathbb{H}f)(w_1) = E[\phi(w_1, W_2) f(W_2)] \tag{14}$$

for any square integrable functions $f(\cdot)$. Let $\{e_j : j \geq 1\}$ denote an orthonormal complete system of eigenfunctions of $\mathbb{H}$ ordered by decreasing absolute values of the corresponding eigenvalues (without loss of generality) $q_1, q_2, \ldots$ such that $|q_1| \geq |q_2| \geq \ldots$. Now we extend this to the case where the kernel $\phi(\cdot)$ is indexed by the parameter $\theta \in \Theta$. We let

$$\begin{aligned} \phi(w_u, w_v; \theta) &= \Sigma_{l \in \overline{\mathcal{I}}} \phi_l(w_u, w_v; \theta) \text{ and} \\ \phi_l(w_u, w_v; \theta) &= m_l(y_u, x_{1u}; \theta) m_l(y_v, x_{1v}; \theta) E\left[\mathbf{1}_{l,jk}(\theta) \frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)} \frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)} | x_u, x_v\right] \end{aligned} \tag{15}$$

where $w_u = (y_u, x_u)$ and $w_v = (y_v, x_v)$. Also let $\overline{q}_1(\theta), \overline{q}_2(\theta), \ldots$ are the ordered eigenvalues of the operator $\mathbb{H}$ in (14) with $\phi(w_u, w_v; \theta)$ such that $|\overline{q}_1(\theta)| \geq |\overline{q}_2(\theta)| \geq \ldots$ for all $\theta \in \partial(\Theta_0(P))$. The kernel function $\phi_l(w_u, w_v; \theta)$ will play a critical role when we derive the asymptotic distribution of (13).

Then, now we need to add two additional assumptions that regulate the functional space $\phi_l(w_u, w_v; \theta)$. We let $N(\varepsilon, \mu, \mathcal{F}, \mathbb{F})$ denote the covering number of radius ε for the functional space $\mathcal{F}$ with envelope $\mathbb{F}$ where μ is a measure on $\{(x_j, x_k) \in \mathcal{X} \times \mathcal{X} : x_j < x_k\}$. We also let T_n be the measure that place mass 1 on each of pairs (x_j, x_k) for $1 \leq j, k \leq 2n$ with the exception of the $4n$ pairs for which $j = k$ (for $1 \leq j \leq 2n$), $j = k - n$ (for $1 \leq j \leq n$), $j = k + n$ (for $n + 1 \leq j \leq 2n$). Finally, define the functional space of $\phi_l(w_u, w_v; \theta)$, indexed by $\theta \in \Theta$, as $\mathcal{F}_{\phi_l} = \{\phi_l(w_u, w_v; \theta) : \theta \in \Theta\}$ with an envelope $\mathbb{F}_{\phi_l}$.

Assumption 5.1 *(i) For all $l \in \overline{\mathcal{I}}$, $\sup_n P\left[\int_0^1 \log N(\varepsilon, T_n, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l}) d\varepsilon\right] < \infty$; (ii) For each $\alpha > 0$ and $\epsilon > 0$, there exists a $\gamma > 0$ such that*

$$\limsup P\left[\int_0^\gamma \log N(\varepsilon, T_n, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l}) d\varepsilon > \alpha\right] < \epsilon$$

; (iii) $\log N(\varepsilon, P_n \otimes P, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l}) = o_P(n)$ for each $\varepsilon > 0$

Assumption 5.1 is the modified conditions from the Theorem 7 in Nolan and Pollard (1988), which justify the functional limit theorem of (13) with the kernel given by (15). Note that when $\mathcal{F}_{\phi_l}$ is Euclidean for the envelope F_{ϕ_l}, we have $N(\varepsilon, \mu, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l}) \leq A\varepsilon^{-V}$ for some constants A and V for any μ such that $0 < \mu F_{\phi_l} < \infty$. It follows that $\log N(\varepsilon, \mu_n, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l}) \longrightarrow_{(\mu_n \longrightarrow \mu)} \log N(\varepsilon, \mu, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l})$ and therefore Assumption 5.1 is trivially satisfied for such classes of functions. Many interesting classes of $\mathcal{F}_{\phi_l}$ satisfy the Euclidean condition including examples in Section 3. For other examples, one can verify Assumption 5.1 using the Lemma 2.13 in Pakes and Pollard (1989).

Assumption 5.2 *(i) For all $l \in \overline{\mathcal{I}}$, $E[\phi_l(W_u, W_v; \theta)^2] < \infty$ for all $\theta \in \partial(\Theta_0(P))$; (ii) For all $l \in \overline{\mathcal{I}}$, $E[\phi_l(W_u, W_v; \theta)^4] < \infty$ and $\overline{q}_9(\theta) \neq 0$ for all $\theta \in \partial(\Theta_0(P))$.*

Note that Assumption 5.2 (i) implies $\sum_{a=1}^{\infty} q_{l,a}^2(\theta) < \infty$ since $E[\phi_l(W_u, W_v; \theta)^2] = \sum_{a=1}^{\infty} q_{l,a}^2(\theta)$ and $\phi_l(W_u, W_v; \theta) = \sum_{a=1}^{\infty} q_{l,a}(\theta) e_{l,a}(W_u; \theta) e_{l,a}(W_v; \theta)$ where $\{e_{l,a}(\cdot; \theta) : j \geq 1\}$ is an orthonormal complete system of eigenfunctions of $\mathbb{H}$ with the kernel function $\phi_l(w_u, w_v; \theta)$ and $q_{l,1}(\theta), q_{l,2}(\theta), \ldots$are the corresponding eigenvalues. This also implies the desirable condition $\sum_{a=1}^{\infty} \overline{q}_a^2(\theta) < \infty$ because we have

$$
\begin{aligned}
E[\phi(W_u, W_v; \theta)^2] &= \textstyle\sum_{a=1}^{\infty} \overline{q}_a^2(\theta) \text{ and} \\
E[\phi(W_u, W_v; \theta)^2] &\leq 2\Sigma_{l \in \overline{\mathcal{I}}} E[\phi_l(W_u, W_v; \theta)^2] = 2\Sigma_{l \in \overline{\mathcal{I}}} \sum_{a=1}^{\infty} q_{l,a}^2(\theta) < \infty.
\end{aligned}
$$

Therefore, we conclude

Lemma 5.1 *Suppose Assumption 4.1 (a)-(b) and (e)-(f) hold. Further suppose Assumption 5.1 and 5.2 (i) hold. Then, uniformly over $\theta \in \partial(\Theta_0(P))$, we have*

$$
n\widehat{Q}_n(\theta) = \textstyle\sum_{a=1}^{\infty} \overline{q}_a(\theta) \left(\chi_a^2 - 1\right) + o_P(1)
$$

where $\overline{q}_a(\theta)$, $a = 1, \ldots, \infty$ denote eigenvalues of the Hilbert-Schmidt operator associated with the kernel function, $\phi(w_u, w_v; \theta)$ in (15).

5.2 Uniform convergence of distribution function

Next we consider the uniform convergence of the distribution function itself. Assumption 5.2 (ii) is an extension of the condition in Theorem 1.1 of Bentkus and Gtöze (1999). We need Assumption 5.2 (ii) to obtain the uniform convergence of the distribution function itself. Let $F_n(\cdot, \theta, P)$ be the distribution function of $n\widehat{Q}_n(\theta)$ and let $F_0(\cdot, \theta, P)$ be the distribution function of $\sum_{a=1}^{\infty} \overline{q}_a(\theta) \left(\chi_a^2 - 1\right)$. We obtain the uniform convergence as

$$
\sup_x |F_n(x, \theta, P) - F_0(x, \theta, P)| = o_P(1) \tag{16}
$$

by resorting to Theorem 1.1 and 1.2 of Bentkus and Gtöze (1999).

Lemma 5.2 *Suppose Assumption 4.1 (a)-(b) and (e)-(f) hold. Further suppose Assumption 5.1 and 5.2 hold. Let $F_n(\cdot,\theta,P)$ be the distribution function of $n\widehat{Q}_n(\theta)$ and let $F_0(\cdot,\theta,P)$ be the distribution function of $\sum_{a=1}^{\infty}\overline{q}_a(\theta)\left(\chi_a^2-1\right)$.*
Then, for any $\theta\in\partial(\Theta_0(P))$ we have $\sup_x|F_n(x,\theta,P)-F_0(x,\theta,P)|=o_P(1)$.

5.3 Pointwise consistency of CS

We will show the consistency of the CS both in pointwise and uniformly in P. To show the pointwise consistency, we need to verify that

$$\liminf_{n\to\infty} P\left[\theta_0\in CR_{\theta_0,n}(1-\alpha)\right]\geq 1-\alpha. \tag{17}$$

We can establish (17) by showing that for every $\theta\in\Theta_0(P)$,

$$\limsup_{n\to\infty}\sup_x\left\{F_b(x,\theta,P)-F_n(x,\theta,P)\right\}\leq 0 \tag{18}$$

due to Theorem 3.1 and 3.2 of Romano and Shaikh (2008). We note that the condition (18) is satisfied from the uniform convergence result in (16) by applying the triangle inequality since

$$\begin{aligned}&\sup_x|F_b(x,\theta,P)-F_n(x,\theta,P)|\\ \leq\ &\sup_x|F_b(x,\theta,P)-F_0(x,\theta,P)|+\sup_x|F_n(x,\theta,P)-F_0(x,\theta,P)|.\end{aligned}$$

Theorem 5.1 *Suppose Assumption 4.1 (a)-(b) and (e)-(f) hold. Further suppose Assumption 5.1 and 5.2 hold. Then, for all $\theta\in\Theta_0(P)$, we have*

$$\liminf_{n\to\infty} P[\theta\in CR_n(1-\alpha)]\geq 1-\alpha.$$

Proof. It suffices to show that

$$\limsup_{n\to\infty}\sup_{x\geq 0}\left\{F_b(x,\theta,P)-F_n(x,\theta,P)\right\}\leq 0 \tag{19}$$

by Theorem 3.1 and Theorem 3.2 of Romano and Shaikh (2008).

We first consider $\theta\in int(\Theta_0(P))$ and $d_{m_2}=0$. For all $t_1<t_2$ such that the set $\{(y_i,x_i):t_1<x_i\leq t_2\}$ is not negligible, we have $H_{0l}(\theta,t_1,t_2)<0$ for all $1\leq l\leq d_{m_1}$. Therefore, $n\widehat{Q}_n(\theta)=0$ w.p.a.1. since we can let

$$\widehat{H}_l(\theta,x_j,x_k)<-\epsilon$$

w.p.a.1 for any arbitrary small $\epsilon > 0$, a.e. $x_j < x_k \in \mathcal{X} \times \mathcal{X}$ by Assumption 4.1 (e). Hence, (19) holds trivially.

Now we turn to $\theta \in \partial(\Theta_0(P))$ and allow for $d_{m_2} > 0$. In this case, let $\mathcal{I}$ denote the set of l ($1 \leq l \leq d_{m_1}$) indices for which the moment condition (1) holds with equality. By construction, $\mathcal{I}$ is not-empty. Then, by applying the similar argument with $\theta \in int(\Theta_0(P))$, w.p.a.1, we can write

$$\widehat{Q}_n(\theta) = \frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \sum_{l \in \mathcal{I}} \left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)} \right)^2 \widehat{\mathbf{1}}_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)} \right)^2 \right\}.$$

From Lemma 5.1, we have

$$n\widehat{Q}_n(\theta) = \textstyle\sum_{a=1}^{\infty} \overline{q}_a(\theta) \left(\chi_a^2 - 1\right) + o_P(1)$$

uniformly over $\theta \in \partial(\Theta_0(P))$ due to the functional limit theorem of Nolan and Pollard (1988).

Now let $F_n(\cdot, \theta, P)$ be the distribution function of $n\widehat{Q}_n(\theta)$ and let $F_0(\cdot, \theta, P)$ be the distribution function of $\sum_{a=1}^{\infty} \overline{q}_a(\theta) \left(\chi_a^2 - 1\right)$. Then, for any $\theta \in \partial(\Theta_0(P))$ we have (16) from Lemma 5.2. Finally, the condition (19) follows by applying the triangle inequality. ■

5.4 Uniform consistency of CS

Now we consider the uniform consistency CS in P. The condition (17) needs to be strengthened as

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P\left[\theta_0 \in CR_{\theta_0,n}(1-\alpha)\right] \geq 1 - \alpha.$$

This is achieved by restricting P to a set of P's that satisfy additional regularity conditions under which the condition (18) holds for any subsequence $\{n_k\}$ of $\{n\}$. To be precise, we desire

$$\limsup_{n \to \infty} \sup_x \left\{ F_{b_{n_k}}(x, \theta_{n_k}, P_{n_k}) - F_{n_k}(x, \theta_{n_k}, P_{n_k}) \right\} \leq 0 \tag{20}$$

for any subsequence n_k and a corresponding sequence $(\theta_{n_k}, P_{n_k}) \in \Theta \times \boldsymbol{P}$ such that $\theta_{n_k} \in \Theta_0(P_{n_k})$. This can be achieved by restricting P such that the strengthened limit theorem of (13) holds for any subsequence n_k and a corresponding sequence $(\theta_{n_k}, P_{n_k}) \in \Theta \times \boldsymbol{P}$ such that $\theta_{n_k} \in \Theta_0(P_{n_k})$. For this purpose, we strengthen Assumptions 5.1 and 5.2 as follows. We define

$$\phi_{P,l}(W_u, W_v; \theta) = m_{P,l}(y_u, x_{1u}; \theta) m_{P,l}(y_v, x_{1v}; \theta) E_P \left[\mathbf{1}_{P,l,jk}(\theta) \frac{\mathbf{1}_{juk}}{\sigma_{P,l}(\theta, x_j, x_k)} \frac{\mathbf{1}_{jvk}}{\sigma_{P,l}(\theta, x_j, x_k)} | x_u, x_v \right]$$

where $\mathbf{1}_{P,l,jk}(\theta)$= $\mathbf{1}[H_{P,l}(\theta, x_j, x_k) \geq 0]$. The subscript P denotes that the expectation is taken w.r.t. P.

Finally let $\boldsymbol{P}$ be a set of distributions that satisfy Assumptions 5.3 and 5.4 below.

Assumption 5.3 *(i) For all* $l = 1, \ldots, d_m$, $\sup_n P\left[\int_0^1 \log N(\varepsilon, T_{P,n}, \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}})d\varepsilon\right] < \infty$; *(ii) For each* $\alpha > 0$ *and* $\epsilon > 0$, *there exists a* $\gamma > 0$ *such that*

$$\limsup_n P\left[\int_0^\gamma \log N(\varepsilon, T_{P,n}, \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}})d\varepsilon > \alpha\right] < \epsilon$$

; *(iii)* $\log N(\varepsilon, P_n \otimes P, \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}}) = o_P(n)$ *for each* $\varepsilon > 0$

Note again that when $\mathcal{F}_{\phi_{P,l}}$ is Euclidean for the envelope $F_{\phi_{P,l}}$, we have $N(\varepsilon, \mu, \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}}) \leq A\varepsilon^{-V}$ for some constants A and V for any μ such that $0 < \mu(P)F_{\phi_{P,l}} < \infty$. It follows that $\log N(\varepsilon, \mu(P)_n, \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}}) \longrightarrow_{(\mu(P)_n \longrightarrow \mu(P))} \log N(\varepsilon, \mu(P), \mathcal{F}_{\phi_{P,l}}, \mathbb{F}_{\phi_{P,l}})$ and therefore Assumption 5.3 is trivially satisfied for such classes of functions.

Assumption 5.4 *(i) For all* $l = 1, \ldots, d_m$, $\sup_{P \in \boldsymbol{P}} E_P[\phi_{P,l}(W_u, W_v; \theta)^2] < \infty$ *for all* $\theta \in \partial(\Theta_0(P))$; *(ii)* $\sup_{P \in \boldsymbol{P}} E_P[\phi_{P,l}(W_u, W_v; \theta)^4] < \infty$ *and* $\overline{q}_9(\theta) \neq 0$ *for all* $\theta \in \partial(\Theta_0(P))$.

Now we conclude that

Theorem 5.2 *Suppose Assumption 4.1 (a)-(b) and (e)-(f) hold for* $P \in \boldsymbol{P}$. *Further suppose Assumption 5.3 and 5.4 hold for* $P \in \boldsymbol{P}$. *Then, for all* $\theta \in \Theta_0(P)$, *we have*

$$\liminf_{n\to\infty} \inf_{P \in \boldsymbol{P}} P[\theta \in CR_n(1-\alpha)] \geq 1 - \alpha.$$

Proof. See Appendix A.5. ■

5.5 Discussion: CS for the identified set

We can construct our CS for the identified set using subsampled criterion functions as in CHT or Romano and Shaikh (2006). First, for any set of parameters $\widetilde{\Theta}$ such that $\Theta_0(P) \subseteq \widetilde{\Theta} \subseteq \Theta$, the $1-\alpha$ CS of the identified set is obtained as a collection of parameters as

$$CS_{\Theta_0,n}(1-\alpha) = \{\theta \in \Theta : \sup_{\theta \in \widetilde{\Theta}} n\widehat{Q}_n(\theta) \leq \widehat{c}_n(\widetilde{\Theta}, 1-\alpha)\}.$$

So we collect all the parameter values that make the sample criterion function less than or equal to a critical value. The critical value is obtained as the $(1-\alpha)$ quantile of the subsampled distribution of the sup criterion function taken over a reference set $\widetilde{\Theta}$ as follows. Let $N_n = \binom{n}{b}$ and let $\widehat{Q}_{n,b,i}(\theta)$ denote the statistic $\widehat{Q}_n(\theta)$ evaluated at the i-th subsample of size b. The critical value is obtained as the $(1-\alpha)$ quantile of the N_n numbers of subsampled criterion functions such that

$$\widehat{c}_n(\widetilde{\Theta}, 1-\alpha) = \inf\left\{x : \frac{1}{N_n}\sum_{1\leq i\leq N_n} \mathbf{1}[\sup_{\theta\in\widetilde{\Theta}} b\widehat{Q}_{n,b,i}(\theta) \leq x] \geq 1-\alpha\right\}.$$

Here the reference parameter set $\widetilde{\Theta}$ can be data-dependent as CHT or can be constructed by an iterative procedure as Romano and Shaikh (2006).

6 Conclusion

This paper studies set estimation and inference of models characterized by conditional moment inequalities. A consistent set estimator and a consistent confidence set with subsampled critical values are suggested. In previous studies on set estimation and inference with conditional moment inequalities, only a finite set of moment inequalities implied by the original conditional moment inequalities are used. We note that while such a practice is well justified in the point-identified models since using all the information in the model is about efficiency of an estimator but it is not in the partially or non-identified models. The reason is that an identified set defined by a smaller set of inequality constraints cannot be tighter than an identified set obtained from a larger number of constraints. Since the conditional moment inequalities imply infinite number of moment inequalities, an inference based on only a finite moment inequalities is potentially worse than the same inference based on the original conditional moment inequalities. In the extreme, we may lose a point identification simply because we approximate the conditional moment inequalities while the usage of the full information indeed can produce a point identification.

In this paper, we consider an estimation and an inference strategy that preserves all the information from the conditional moment inequalities. Potentially, this will enable us to obtain a smaller set estimator and tighter confidence regions than those from methods in the current literature. Our proposal can also be used for a model characterized by conditional moment equalities but not necessarily point-identified.

Finally, we establish the asymptotic validity of our set estimator and the confidence set using functional limit theorems of U-processes and uniform bounds in non-Gaussian limit distribution of degenerate U-processes, which are not trivial.

Appendix

A Mathematical Proofs

We will use the following notation: $\mathbf{1}_{jik} = \mathbf{1}[x_j < x_i \leq x_k]]$, $\mathbf{1}_{l,jk}(\theta) = \mathbf{1}[H_l(\theta, x_j, x_k) \geq 0]$, $\mathbf{1}^0_{l,jk}(\theta) = \mathbf{1}[H_{0l}(\theta, x_j, x_k) \geq 0]$, and $\widehat{\mathbf{1}}_{l,jk}(\theta) = \mathbf{1}[\widehat{H}_l(\theta, x_j, x_k) \geq 0]$. We also let $E^{x_j,x_k}[\cdot]$ denote the expectation fixing $x_j < x_k$.

A.1 Proof of Theorem 4.1

We prove Theorem 4.1 by showing Condition C.1 and C.2 of CHT are satisfied under our model and assumptions.

Proof. (Condition C.1 of CHT) The condition C.1 (a)-(c) of CHT is satisfied by Assumption 4.1 (a) and our construction of $Q(\theta)$ and $\widehat{Q}_n(\theta)$. Define

$$Q_n(H, \sigma, \theta) = \frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \sum_{l=1}^{d_{m_1}} \left(\frac{H_l(\theta, x_j, x_k)}{\sigma_l(\theta, x_j, x_k)} \right)^2 \mathbf{1}_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{H_l(\theta, x_j, x_k)}{\sigma_l(\theta, x_j, x_k)} \right)^2 \right\}$$

and thus $\widehat{Q}_n(\theta) = Q_n(\widehat{H}, \widehat{\sigma}, \theta)$.

Now we verify the condition C.1 (d). Note that

$$\sup_{\Theta}(Q(\theta) - \widehat{Q}_n(\theta))_+ \leq \sup_{\Theta} |Q(\theta) - Q_n(H_0, \sigma_0, \theta)| + \sup_{\Theta} \left| Q_n(H_0, \sigma_0, \theta) - \widehat{Q}_n(\theta) \right|. \tag{21}$$

Due to Corollary 7 in Sherman (1994a), we bound the first term in the R.H.S by

$$\sup_{\Theta} |Q_n(H_0, \sigma_0, \theta) - Q(\theta)| = O_P(n^{-1/2}) \tag{22}$$

noting

$$\begin{aligned} & |Q_n(H_0, \sigma_0, \theta) - Q(\theta)| \\ = \; & \frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \sum_{l=1}^{d_{m_1}} \left(\frac{H_{0l}(\theta, x_j, x_k)}{\sigma_{0l}(\theta, x_j, x_k)} \right)^2 \mathbf{1}^0_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{H_{0l}(\theta, x_j, x_k)}{\sigma_{0l}(\theta, x_j, x_k)} \right)^2 \right\} \\ & -E \left[\sum_{l=1}^{d_{m_1}} \left(\frac{H_{0l}(\theta, x_j, x_k)}{\sigma_{0l}(\theta, x_j, x_k)} \right) \mathbf{1}^0_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m} \left(\frac{H_{0l}(\theta, x_j, x_k)}{\sigma_{0l}(\theta, x_j, x_k)} \right) \right] \end{aligned}$$

and the functional space $\left\{ \left(\frac{H_{0l}(\theta,\cdot)}{\sigma_{0l}(\theta,\cdot)} \right)^2 \mathbf{1}^0_{l,jk}(\theta) - E \left[\left(\frac{H_{0l}(\theta,\cdot)}{\sigma_{0l}(\theta,\cdot)} \right)^2 \mathbf{1}^0_{l,jk}(\theta) \right] \right\}$ is Euclidean with the same envelope $\mathbb{H}^\sigma_l$ for $\mathcal{H}^\sigma_l$ since $\mathbf{1}[H_{0l}(\theta, x_j, x_k) \geq 0] \leq 1$ by construction. Now we bound the second R.H.S term of (21). Note that

$$
\begin{aligned}
&\sup_{\Theta}\left|\widehat{Q}_n(\theta)-Q_n(H_0,\sigma_0,\theta)\right| \\
\leq\ &\sup_{\Theta}\left|\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l=1}^{d_{m_1}}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\widehat{\sigma}_l(\theta,x_j,x_k)^2}\widehat{\mathbf{1}}_{l,jk}(\theta)-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\mathbf{1}^0_{l,jk}(\theta)\right)\right\}\right| \qquad (23)\\
&+\sup_{\Theta}\left|\frac{1}{n(n-1)}\sum_{j\neq k}\sum_{l=1+d_{m_1}}^{d_m}\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\widehat{\sigma}_l(\theta,x_j,x_k)^2}-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\right|
\end{aligned}
$$

We will show the first term in (23) is bounded by $O_P(n^{-1/2})$ and the second term of (23) can be shown to be $O_P(n^{-1/2})$ similarly. Using the triangle inequality several times, we obtain

$$
\begin{aligned}
&\sup_{\Theta}\left|\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l=1}^{d_{m_1}}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\widehat{\sigma}_l(\theta,x_j,x_k)^2}\widehat{\mathbf{1}}_{l,jk}(\theta)-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\mathbf{1}^0_{l,jk}(\theta)\right)\right\}\right| \\
\leq\ &\sup_{\Theta}\left|\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l=1}^{d_m}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\right)\mathbf{1}^0_{l,jk}(\theta)\right\}\right| \qquad (24)\\
&+\sup_{\Theta}\left|\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l=1}^{d_m}\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma^2_{0l}(\theta,\cdot)}\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta)\right\}\right| \qquad (25)\\
&+\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l=1}^{d_{m_1}}\sup_{\Theta}\left|\widehat{\mathbf{1}}_{l,jk}(\theta)-\mathbf{1}^0_{l,jk}(\theta)\right|\left(\frac{\widehat{H}_{0l}(\theta,x_j,x_k)}{\widehat{\sigma}_{0l}(\theta,x_j,x_k)}\right)^2\right\}. \qquad (26)
\end{aligned}
$$

In (24), we expand $\widehat{H}_l(\theta,x_j,x_k)^2$ such that $\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\right)\mathbf{1}^0_{l,jk}(\theta)$ becomes

$$
\frac{1}{n(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq t}\left\{\frac{m_l(y_i,x_{1i};\theta)}{\sigma_{0l}(\theta,x_j,x_k)}\frac{m_l(y_t,x_{1t};\theta)}{\sigma_{0l}(\theta,x_j,x_k)}\mathbf{1}_{jik}\mathbf{1}_{jtk}-\frac{H_{0l}(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,x_j,x_k)^2}\right\}\mathbf{1}^0_{l,jk}(\theta)
$$

and notice this is a fourth order U-process with zero mean. Thus, by Corollary 7 in Sherman (1994a) and Assumption 4.1 (d), we bound the above term by $O_P(n^{-1/2})$. Note that we will have the same envelope $\mathbb{M}^\sigma_l$ for $\mathcal{M}^\sigma_l$ since $\mathbf{1}[H_{0l}(\theta,x_j,x_k)\geq 0]\leq 1$ by construction. We, therefore, bound (24) as $O_P(n^{-1/2})$.

Now we bound (25). It suffices to show that for each given l and uniformly over $\theta\in\Theta$,

$$
\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma^2_{0l}(\theta,\cdot)}\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta)\right\}=O_P(n^{-1/2}).
$$

Consider

$$\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma^2_{0l}(\theta,\cdot)}\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta)\right\}$$

$$= \frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma^2_{0l}(\theta,\cdot)}-E\left[\frac{H_l(\theta,\cdot)^2}{\sigma_{0l}(\theta,\cdot)^2}\right]\right)\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta) \quad (27)$$

$$+E\left[\frac{H_l(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,\cdot)^2}\right]\frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta).$$

Note that

$$\frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)^2}{\sigma^2_{0l}(\theta,\cdot)}-E\left[\frac{H_l(\theta,x_j,x_k)^2}{\sigma_{0l}(\theta,\cdot)^2}\right]\right) \quad (28)$$

$$= \frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{\widehat{H}_l(\theta,\cdot)^2}{\sigma^2_{0l}(\theta,\cdot)}-\frac{H_l(\theta,\cdot)^2}{\sigma_{0l}(\theta,\cdot)^2}\right)+\frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{H_l(\theta,\cdot)^2}{\sigma_{0l}(\theta,\cdot)^2}-E\left[\frac{H_l(\theta,\cdot)^2}{\sigma_{0l}(\theta,\cdot)^2}\right]\right)$$

Then, by Assumption 4.1 (d) and applying the Corollary 7 in Sherman (1994a), we bound the first term in (28) by $O_P(n^{-1/2})$ and similarly we bound the second term in (28) by Assumption 4.1 (c) and the Corollary 7 in Sherman (1994a). We, therefore, bound the first term in (27) by $O_P(n^{-1/2})$ applying the dominated convergence theorem and Assumption 4.1 (f).

Now we turn to the second term of (27). Note

$$\frac{1}{n(n-1)}\sum_{j\neq k}\left(\frac{\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2}{\widehat{\sigma}^2_l(\theta,\cdot)}\right)\mathbf{1}^0_{l,jk}(\theta)\leq (C+o_P(1))\frac{1}{n(n-1)}\sum_{j\neq k}\left|\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2\right|$$

since $\sup_\Theta \frac{1}{\widehat{\sigma}^2_l(\theta,x_j,x_k)} \leq \frac{1}{c_L(x_j,x_k)^2} < C < \infty$ w.p.a.1 by Assumption 4.1 (f). Now fixing x_j and x_k, we obtain $\left|\widehat{\sigma}_l(\theta,\cdot)^2-\sigma_{0l}(\theta,\cdot)^2\right| = O_P(n^{-1/2})$ by Assumption 4.1 (f). Then, we conclude the second term in (27) is bounded by $O_P(n^{-1/2})$ applying the dominated convergence theorem. This concludes (25) is bounded by $O_P(n^{-1/2})$.

To bound (26), we first show that

$$\max_{1\leq j\neq k\leq n}\sup_\Theta\left|\mathbf{1}[\widehat{H}_l(\theta,x_j,x_k)\geq 0]-\mathbf{1}[H_{0l}(\theta,x_j,x_k)\geq 0]\right| = O_P(n^{-1/2}) \quad (29)$$

using a similar argument with Lemma A.3 in Newey, Powell, and Vella (1999) such that

$$\max_{1\leq j\neq k\leq n}\sup_\Theta\left|\mathbf{1}[\widehat{H}_l(\theta,x_j,x_k)\geq 0]-\mathbf{1}[H_{0l}(\theta,x_j,x_k)\geq 0]\right| \quad (30)$$

$$= O_P\left(\max_{1\leq j\neq k\leq n}\sup_\Theta\left|\widehat{H}_l(\theta,x_j,x_k)-H_{0l}(\theta,x_j,x_k)\right|\right)$$

and conclude $\max_{1\le j\neq k\le n}\sup_\Theta \left|\widehat{H}_l(\theta,x_j,x_k)-H_{0l}(\theta,x_j,x_k)\right| = O_P(n^{-1/2})$ by Assumption 4.1 (e). Now for (26), consider

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sup_\Theta\left|\widehat{\mathbf{1}}_{l,jk}(\theta)-\mathbf{1}_{l,jk}(\theta)\right|\left(\frac{\widehat{H}_{0l}(\theta,x_j,x_k)}{\widehat{\sigma}_{0l}(\theta,x_j,x_k)}\right)^2\right\} \qquad (31)\\
\le\;& O_P(n^{-1/2})(C+o_P(1))\frac{1}{n(n-1)}\sum_{j\neq k}\left(\widehat{H}_{0l}(\theta,x_j,x_k)\right)^2\\
=\;& O_P(n^{-1/2})(C+o_P(1))\frac{1}{n(n-1)}\sum_{j\neq k}\left(\widehat{H}_{0l}(\theta,x_j,x_k)\right)^2\\
=\;& O_P(n^{-1/2})(C+o_P(1))\left(\frac{1}{n(n-1)}\sum_{j\neq k}\left(\widehat{H}_{0l}(\theta,x_j,x_k)^2-H_{0l}(\theta,x_j,x_k)^2\right)\right) \qquad (32)\\
&+O_P(n^{-1/2})(C+o_P(1))\frac{1}{n(n-1)}\sum_{j\neq k}\left(H_{0l}(\theta,x_j,x_k)\right)^2 \qquad (33)
\end{aligned}$$

where the first inequality is due to (30) and by Assumption 4.1 (f). By expanding the term in (32) to the fourth order U-process, we obtain

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{j\neq k}\left(\widehat{H}_{0l}(\theta,x_j,x_k)^2-\left(H_{0l}(\theta,x_j,x_k)\right)^2\right)\\
=\;&\frac{1}{n(n-1)(n-2)(n-3)}\sum_{i\neq j\neq k\neq t} m_l(y_i,x_{1i};\theta)m_l(y_t,x_{1t};\theta)\mathbf{1}_{jik}\mathbf{1}_{jtk}-H_{0l}(\theta,x_j,x_k)^2\\
=\;& O_P(n^{-1/2})
\end{aligned}$$

by Assumption 4.1 (d) and Corollary 7 of Sherman (1994a). The term $\frac{1}{n(n-1)}\sum_{j\neq k}\left(H_{0l}(\theta,\cdot)\right)^2 = O_P(1)$ in (33) under $E\left[H_{0l}(\theta,x_j,x_k)^2\right]$ is bounded. Therefore, we conclude (31) is $O_P(n^{-1/2})$ and thus we bound (26) by $O_P(n^{-1/2})$ by the dominated convergence theorem with the dominating function $E[H_l(\theta,x_j,x_k)^2]$. Combining above results, we conclude

$$\sup_\Theta\left|\widehat{Q}_n(\theta)-Q_n(H,\theta)\right| = O_P(n^{-1/2}). \qquad (34)$$

Therefore, the condition C.1. (d) of CHT is satisfied. Now we show the condition C.1. (e) of CHT is satisfied under Assumption 4.1. Now note under $\theta \in \Theta_0$

$$\begin{aligned}
\widehat{Q}_n(\theta) &\leq \frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \sum_{l=1}^{d_{m_1}} \frac{\widehat{H}_l(\theta, x_j, x_k)^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2} \widehat{\mathbf{1}}_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m} \frac{\widehat{H}_l(\theta, x_j, x_k)^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2} \right\} \\
&\leq \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \left| \frac{\widehat{H}_l(\theta, x_j, x_k)^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2} \right| \\
&\leq \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \sup_{\Theta} \frac{1}{\widehat{\sigma}_l^2(\theta, x_j, x_k)} \left| \widehat{H}_l(\theta, x_j, x_k)^2 \right| \\
&= \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \sup_{\Theta} \frac{1}{\widehat{\sigma}_l^2(\theta, x_j, x_k)} \left(\frac{1}{n} \sum_{i=1}^{n} m_l(y_i, x_{1i}; \theta) \mathbf{1}[x_j < x_i \leq x_k] \right)^2 \\
&\leq (C + o_P(1)) \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \left(\frac{1}{n} \sum_{i=1}^{n} m_l(y_i, x_{1i}; \theta) \mathbf{1}[x_j < x_i \leq x_k] \right)^2
\end{aligned}$$

where $0 < 1/C < c_L(x_j, x_k) < \sigma_l^2(\theta, x_j, x_k)$ for all $\theta \in \Theta'$ and $x_j < x_k \in \mathcal{X} \times \mathcal{X}$. Such C exists by Assumption 4.1 (f).

For each given l, now we expand $\frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \left(\frac{1}{n} \sum_{i=1}^{n} m_l(y_i, x_{1i}; \theta) \mathbf{1}[x_j < x_i \leq x_k] \right)^2$ to a fourth-order U-process and decompose it into a sum of degenerate U-processes up to the order of two following Sherman (1994a) and Serfling (1980). Following a similar step as in Section A.3, we obtain

$$\begin{aligned}
&\frac{1}{n(n-1)(n-2)(n-3)} \sum_{u \neq v \neq j \neq k} \{ m_l(y_u, x_{1u}; \theta) m_l(y_v, x_{1v}; \theta) \mathbf{1}_{juk} \mathbf{1}_{jvk} \} \qquad (35) \\
= \; &\frac{1}{n(n-1)} \sum_{u \neq v} m_l(y_u, x_{1u}; \theta) m_l(y_v, x_{1v}; \theta) E\left[\mathbf{1}_{juk} \mathbf{1}_{jvk} | x_u, x_v \right] + O_P(n^{-3/2})
\end{aligned}$$

where the leading term in the RHS of (35) is a degenerate second-order U-process. Now we obtain

$$\frac{1}{n(n-1)} \sum_{u \neq v} m_l(y_u, x_{1u}; \theta) m_l(y_v, x_{1v}; \theta) E\left[\mathbf{1}_{juk} \mathbf{1}_{jvk} | x_u, x_v \right] = O_P(n^{-1})$$

since $\frac{n}{n(n-1)} \sum_{u \neq v} m_l(y_u, x_{1u}; \theta) m_l(y_v, x_{1v}; \theta) E\left[\mathbf{1}_{juk} \mathbf{1}_{jvk} | x_u, x_v \right]$ has a limiting distribution given by a linear combination of independent centered χ_1^2 distributions, indexed by $\theta \in \partial(\Theta_0(P))$ for all $\theta \in \partial(\Theta_0(P))$ due to Gregory (1977), Neuhaus (1977), and Nolan and Pollard (1988). Thus, we bound (35) by $O_P(n^{-1})$ uniformly over $\theta \in \partial(\Theta_0(P))$. When $int(\Theta_0(P))$ is not empty and for all $\theta \in int(\Theta_0(P))$, for all $t_1 < t_2$ such that the set $\{(y_i, x_i) : t_1 < x_{1i} \leq t_2\}$ is not negligible, we have $H_{0l}(\theta, t_1, t_2) < 0$ for all $1 \leq l \leq d_{m_1}$. Therefore, $n\widehat{Q}_n(\theta) = 0$ w.p.a.1. since we can let

$$\widehat{H}_l(\theta, x_j, x_k) < -\epsilon$$

w.p.a.1 for any arbitrary small $\epsilon > 0$, a.e. $x_j < x_k \in \mathcal{X} \times \mathcal{X}$ by Assumption 4.1 (e). Therefore, we have $\widehat{Q}_n(\theta) = O_P(n^{-1})$ uniformly over $\theta \in \Theta_0(P)$.

Then, we conclude $\sup_{\theta \in \Theta_0} \widehat{Q}_n(\theta) = O_P(1/n)$ since the convergence rate result in (35) holds uniformly over $\theta \in \Theta$. ■

Remark 1 *Assumption 4.1 (f) is actually implied by Assumption 4.1 (e) and .the following addtiional assumption*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (m_l(\theta, y_i, x_{1i}) \mathbf{1}[t_1 \leq x_i \leq t_2])^2 - \sqrt{n} E[(m_l(\theta, y_i, x_{1i}) \mathbf{1}[t_1 \leq x_i \leq t_2])^2] \tag{36}$$

also satisfies the P-Donsker property. Note that

$$\begin{aligned}
& \widehat{\sigma}_l(\theta, x_j, x_k)^2 - \sigma_{0l}(\theta, x_j, x_k)^2 \\
= \; & \frac{1}{n} \sum_{i=1}^{n} (m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik})^2 - \widehat{m}_l(\theta, x_j, x_k)^2 \\
& - \left(E^{x_j, x_k}[(m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik})^2] - (E^{x_j, x_k}[m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik}])^2 \right) \\
= \; & \frac{1}{n} \sum_{i=1}^{n} (m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik})^2 - E^{x_j, x_k}[(m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik})^2] \\
& - \left\{ \widehat{m}_l(\theta, x_j, x_k)^2 - (E^{x_j, x_k}[m_l(\theta, y_i, x_{1i}) \mathbf{1}_{jik}])^2 \right\} \\
= \; & O_p(n^{-1/2})
\end{aligned}$$

where the last result holds by Assumption 4.1 (e) and by requiring (36) to satisfy the P-Donsker property.

Proof. (Condition C.2 of CHT) Now we turn to the condition C.2 of CHT. This is proved similarly with the proof of Theorem 4.2 in CHT. Without loss of generality, we will let $d_{m_2} = 0$.

Observe that w.p.a.1 for any $\theta \notin \Theta_0$, fixing x_j and x_k inside the outer summation,

$$\begin{aligned}
n\widehat{Q}_n \; = \; & \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \frac{(\mathbb{G}_{l,n}(\theta, x_j, x_k) + \sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}])^2}{\widehat{\sigma}_l(\theta, x_j, x_k)^2} \widehat{\mathbf{1}}_{l,jk}(\theta) \qquad (37) \\
\geq \; & \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \frac{(\mathbb{G}_{l,n}(\theta, x_j, x_k) + \sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}])^2 \widehat{\mathbf{1}}_{l,jk}(\theta)}{\sup_{\Theta} \widehat{\sigma}_l^2(\theta, x_j, x_k)} \\
= \; & \frac{1}{n(n-1)} \sum_{j \neq k} \sum_{l=1}^{d_m} \frac{1}{\sup_{\Theta} \widehat{\sigma}_l^2(\theta, x_j, x_k)} \begin{array}{c} (\sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}])^2 \mathbf{1}_{l,jk}(\theta) \\ \cdot \frac{(\mathbb{G}_{l,n}(\theta, x_j, x_k) + \sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}])^2}{(\sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}])^2 \mathbf{1}[H_l(\theta, x_j, x_k) \geq 0]} \widehat{\mathbf{1}}_{l,jk}(\theta) \end{array}
\end{aligned}$$

where $\mathbb{G}_{l,n}(\theta, x_j, x_k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik} - \sqrt{n} E^{x_j, x_k}[m_l(y_i, x_{1i}; \theta) \mathbf{1}_{jik}]$.

Now note that $\frac{1}{\sup_\Theta \widehat{\sigma}_l^2(\theta,x_j,x_k)} \geq \frac{1}{c_U(x_j,x_k)} \geq \frac{1}{c_U}$ w.p.a.1. by Assumption 4.1 (f). Further note that $\max_{1\leq j\neq k\leq n}\sup_{\theta\in\Theta}|\mathbb{G}_n(\theta,x_j,x_k)| = O_P(1)$ from the P-Donsker property by Assumption 4.1 (e). It follows that

$$\frac{\left(\mathbb{G}_{l,n}(\theta,x_j,x_k)+\sqrt{n}E^{x_j,x_k}[m_l(y_i,x_{1i};\theta)\mathbf{1}_{jik}]\right)^2}{\left(\sqrt{n}E^{x_j,x_k}[m_l(y_i,x_{1i};\theta)\mathbf{1}_{jik}]\right)^2\mathbf{1}[H_l(\theta,x_j,x_k)\geq 0]}\widehat{\mathbf{1}}_{l,jk}(\theta)\to 1.$$

Also by Assumption 4.1 (g), we have for all $(x_j,x_k)\in\widetilde{B}$

$$\sum_{l=1}^{d_m}\left(\sqrt{n}E^{x_j,x_k}[m_l(y_i,x_{1i};\theta)\mathbf{1}_{jik}]\right)^2\mathbf{1}[H_l(\theta,x_j,x_k)\geq 0]\geq C\cdot n\cdot\left(d\left(\theta,\Theta_0\right)\wedge\delta\right)^2$$

on Θ for some $C>0$ and $\delta>0$. Therefore, applying a similar argument in the proof of Theorem 4.2 in CHT, we can choose $(\kappa_\varepsilon,n_\varepsilon)$ such that for all $n\geq n_\varepsilon$ we have

$$\widehat{Q}_n\geq\frac{1}{2}\cdot\frac{1}{c_U}\Pr((x_j,x_k)\in\widetilde{B})\cdot C\cdot\left(d\left(\theta,\Theta_0\right)\wedge\delta\right)^2 \tag{38}$$

uniformly in $\{\theta\in\Theta: d\left(\theta,\Theta_0\right)\wedge\delta\geq(\kappa_\varepsilon/n)^{1/2}\}$ with probability at least $1-\varepsilon$. This completes the proof. ■

A.2 Proof of Theorem 4.2

Proof. Define ϵ-contraction of Θ_0 as $\Theta_0^{-\epsilon}=\{\theta\in\Theta_0: d(\theta,\Theta\backslash\Theta_0)\geq\epsilon\}$ and ϵ-expansion of Θ_0 as $\Theta_0^{\epsilon}=\{\theta\in\Theta: d(\theta,\Theta_0)\leq\epsilon\}$. First, note that we have $\widehat{Q}_n(\theta)=0$ for $\theta\in\Theta_0^{-\epsilon}$ by Assumption 2.1 (ii) and Assumption 2.1 (e) and (f). This can be shown as the proof of Theorem 4.2 in CHT.

It follows that w.p.a.1, $\Theta_0^{-\epsilon}\subset\widehat{\Theta}_{n,0}$ by definition of $\Theta_0^{-\epsilon}$ and $\widehat{\Theta}_{n,0}=\{\theta\in\Theta:\widehat{Q}_n=0\}$ and since $\widehat{\Theta}_{n,0}$ is not empty. We can also show that $\widehat{\Theta}_{n,0}\subset\Theta_0^{\epsilon}$ w.p.a.1, following the Step (b) in the proof of Theorem 3.1 in CHT under the Condition C.1, which was shown to be satisfied by Assumptions 4.1 (a)-(f) in Section (A.1). It follows that w.p.a.1,

$$\Theta_0^{-\epsilon}\subset\widehat{\Theta}_{n,0}\subset\Theta_0^{\epsilon}. \tag{39}$$

Therefore, it is obvious that $d(\widehat{\Theta}_{n,0},\Theta_0)\to_p 0$ by (39).

Now let $\epsilon_n=O(n^{-1/2})$. Then, $\Theta_0^{-\epsilon_n}\subset\widehat{\Theta}_{n,0}$ w.p.a.1 by construction of $\Theta_0^{-\epsilon_n}$ and $\widehat{\Theta}_{n,0}$ under Assumption 2.1 (ii). We can also show that $\widehat{\Theta}_{n,0}\subset\Theta_0^{\epsilon_n}$ w.p.a.1, following the Part (2) in the proof of Theorem 3.1 in CHT under the Condition C.1 and C.2 ,which were shown to be satisfied by Assumptions 4.1 (a)-(g) in Section (A.1). Therefore, we have w.p.a.1,

$$\Theta_0^{-\epsilon_n}\subset\widehat{\Theta}_{n,0}\subset\Theta_0^{\epsilon_n}. \tag{40}$$

Then, w.p.a.1 $\rho\left(\widehat{\Theta}_{n,0}|\Theta_0\right) \leq \rho\left(\Theta_0^{\epsilon_n}|\Theta_0\right) \leq \epsilon_n$ by (40). Also note w.p.a.1 $\rho\left(\Theta_0|\widehat{\Theta}_{n,0}\right) \leq \rho\left(\Theta_0|\Theta_0^{-\epsilon_n}\right) \leq \epsilon_n$ by (40). Therefore, we have $d(\widehat{\Theta}_{n,0}, \Theta_0) = O_P(\epsilon_n)$.

When Assumption 2.1 (i) holds, we have $\widehat{\Theta}_{n,0} \subset \Theta_0^{\epsilon_n}$. It follows that

$$\rho\left(\widehat{\Theta}_{n,0}|\theta_0\right) \leq \rho\left(\Theta_0^{\epsilon_n}|\theta_0\right) \leq \epsilon_n. \tag{41}$$

Also note that $\rho\left(\theta_0|\widehat{\Theta}_{n,0}\right) \leq \rho\left(\widehat{\Theta}_{n,0}|\theta_0\right) \leq \epsilon_n$ by (41) and since (i) the distance from a point to a non-empty set is less than or equal to the distance from the set to the point by definition of $\rho(\cdot|\cdot)$ and (ii) $\widehat{\Theta}_{n,0}$ is not empty. ■

A.3 Proof of Lemma 5.1

Proof. Consider $\theta \in \partial(\Theta_0(P))$ and let $\mathcal{I}$ denote the set of l $(1 \leq l \leq d_{m_1})$ indices for which the moment condition (1) holds with equality for some x_i. By construction, $\mathcal{I}$ is not-empty. Note that for all $t_1 < t_2$ such that the set $\{(y_i, x_i) : t_1 < x_i \leq t_2\}$ is not negligible, we have $H_{0l}(\theta, t_1, t_2) < 0$ for all $l \in \mathcal{I}^c \equiv \{l \notin \mathcal{I} : 1 \leq l \leq d_{m_1}\}$. Therefore, $\sum_{l\in\mathcal{I}^c}\left(\frac{\widehat{H}_l(\theta,x_j,x_k)}{\widehat{\sigma}_l(\theta,x_j,x_k)}\right)^2 \widehat{\mathbf{1}}_{l,jk}(\theta) = 0$ w.p.a.1. since we can let

$$\widehat{H}_l(\theta, x_j, x_k) < -\epsilon$$

w.p.a.1 for any arbitrary small $\epsilon > 0$, a.e. $x_j < x_k \in \mathcal{X} \times \mathcal{X}$ by Assumption 4.1 (e). Therefore, w.p.a.1, we can write

$$\widehat{Q}_n(\theta) = \frac{1}{n(n-1)}\sum_{j\neq k}\left\{\sum_{l\in\mathcal{I}}\left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)}\right)^2 \widehat{\mathbf{1}}_{l,jk}(\theta) + \sum_{l=1+d_{m_1}}^{d_m}\left(\frac{\widehat{H}_l(\theta, x_j, x_k)}{\widehat{\sigma}_l(\theta, x_j, x_k)}\right)^2\right\}.$$

Now to derive the asymptotic behavior of $\widehat{Q}_n(\theta)$, we expand $\widehat{Q}_n(\theta)$ to the fourth-order U-process by replacing $\mathbf{1}[\widehat{H}_l(\theta, x_j, x_k) \geq 0]$ with $\mathbf{1}[H_l(\theta, x_j, x_k) \geq 0]$ and such a replacement does not affect our asymptotic result due to (29). We also replace $\widehat{\sigma}_l(\theta, x_j, x_k)$ with $\sigma_l(\theta, x_j, x_k)$ since $\widehat{\sigma}_l(\theta, x_j, x_k) \to \sigma_l(\theta, x_j, x_k)$ uniformly over $\theta \in \Theta$ by Assumption 4.1 (f).

Therefore, for each $l \in \overline{\mathcal{I}}$, the U-process is given by

$$U_n^4 f_l(j,k,u,v) \equiv \frac{1}{n(n-1)(n-2)(n-3)}\sum_{j\neq k\neq u\neq v}\left\{\mathbf{1}_{l,jk}(\theta)m_l(y_u, x_{1u};\theta)m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,\cdot)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,\cdot)}\right\}$$

where $U_n^i[\cdot]$ denotes the i-th order U-process of $[\cdot]$.

Now we can decompose $U_n^4 f_l(j,k,u,v)$ into a sum of degenerate U-processes (up to the order of two) following Sherman (1994a) and Serfling (1980) such that

$$\begin{aligned}
& \left\{U_n^4 f_l(j,k,u,v) - Ef_l\right\} \\
= \; & P_n\left[f_l(P,P,P,v) - Ef_l\right] + P_n\left[f_l(P,P,u,P) - Ef_l\right] \\
& + P_n\left[f_l(j,P,P,P) - Ef_l\right] + P_n\left[f_l(P,k,P,P) - Ef_l\right] \\
& + U_n^2\left[f_l(j,k,P,P) - Ef_l\right] + U_n^2\left[f_l(j,P,P,v) - Ef_l\right] + U_n^2\left[f_l(j,P,u,P) - Ef_l\right] \\
& + U_n^2\left[f_l(P,P,u,v) - Ef_l\right] + U_n^2\left[f_l(P,k,P,v) - Ef_l\right] + U_n^2\left[f_l(P,k,u,P) - Ef_l\right] \\
& + O_P(n^{-3/2})
\end{aligned}$$

where P_n denotes empirical expectation (sample mean), $U_n^2[\cdot]$ denotes the second order U-process of $[\cdot]$, Ef_l denotes the unconditional expectation of $f_l(j,k,u,v)$, $f_l(P,P,P,v)$ denotes the conditional expectation of $f_l(j,k,u,v)$ on v, $f_l(j,k,P,P)$ denotes the conditional expectation of $f_l(j,k,u,v)$ conditional on j and k, and others are defined in a similar manner.

Applying the law of iterated expectation (IIE) twice, we have

$$\begin{aligned}
Ef_l &= E\left[\mathbf{1}_{l,jk}(\theta) m_l(y_u, x_{1u};\theta) m_l(y_v, x_{1v};\theta) \frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)} \frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)}\right] \\
&= E\left[\mathbf{1}_{l,jk}(\theta) E^{x_j,x_k}\left[m_l(y_u, x_{1u};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)}\right] E^{x_j,x_k}\left[m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)}\right]\right].
\end{aligned}$$

We consider two cases. First note that for those (x_j, x_k) which (4)-(5) are binding, we have $\mathbf{1}_{l,jk}(\theta) = 1$ and

$$E^{x_j,x_k}\left[m_l(y_u, x_{1u};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)}\right] = 0$$

by construction. Second for those (x_j, x_k) which (4) is not binding and $\theta \in \Theta_0(P)$, we have $\mathbf{1}_{l,jk}(\theta) = 0$. Therefore, we conclude

$$Ef_l = 0.$$

Next consider that

$$\begin{aligned}
f_l(P,P,P,v) &= E[\mathbf{1}_{l,jk}(\theta) m_l(y_u, x_{1u};\theta) m_l(y_v, x_{1v};\theta) \frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)} \frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)} |(y_v, x_v)] \\
&= E\left[\mathbf{1}_{l,jk}(\theta) E[m_l(y_u, x_{1u};\theta) m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,\cdot)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,\cdot)}|(y_v, x_v), x_j, x_k]|(y_v, x_v)\right]
\end{aligned}$$

and for the term inside the first expectation, observe that

$$\begin{aligned}
& E[m_l(y_u, x_{1u};\theta) m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)}|(y_v, x_v), x_j, x_k] \\
= \; & E[m_l(y_u, x_{1u};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)}|(y_v, x_v), x_j, x_k] m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)} \\
= \; & E^{x_j,x_k}\left[m_l(y_u, x_{1u};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta, x_j, x_k)}\right] m_l(y_v, x_{1v};\theta)\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta, x_j, x_k)}.
\end{aligned}$$

Therefore, by the similar argument, we conclude $f_l(P,P,P,v)=0$. Similarly, we can show that $f_l(P,P,u,P)=0$.

Next consider that

$$
\begin{aligned}
f_l(j,P,P,P) &= E[\mathbf{1}_{l,jk}(\theta)m_l(y_u,x_{1u};\theta)m_l(y_v,x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,x_j,x_k)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,x_j,x_k)}|x_j] \\
&= E\left[\mathbf{1}_{l,jk}(\theta)E[m_l(y_u,x_{1u};\theta)m_l(y_v,x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,\cdot)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,\cdot)}|x_j,x_k]|x_j\right]
\end{aligned}
$$

and for the term inside the first expectation, observe that

$$
\begin{aligned}
& E[m_l(y_u,x_{1u};\theta)m_l(y_v,x_{1v};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,x_j,x_k)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,x_j,x_k)}|x_j,x_k] \\
= \; & E^{x_j,x_k}\left[m_l(y_u,x_{1u};\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,x_j,x_k)}\right]E^{x_j,x_k}\left[m_l(y_v,x_{1v};\theta)\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,x_j,x_k)}\right].
\end{aligned}
$$

Therefore, by the similar argument, we conclude $f_l(j,P,P,P)=0$. Following similar arguments, we further conclude

$$
\begin{aligned}
f_l(P,k,P,P) &= 0,\ f_l(j,k,P,P)=0,\ f_l(j,P,P,v)=0,\ f_l(j,P,u,P)=0, \\
f_l(P,k,u,P) &= 0, \text{ and } f_l(P,k,P,v)=0.
\end{aligned}
$$

under $\overline{\mathcal{I}}$ by applying the law of iterated expectations. We, therefore, conclude

$$
\begin{aligned}
U_n^4 f_l(j,k,u,v) &= U_n^2 f_l(P,P,u,v)+O_P(n^{-3/2}) \qquad (42)\\
&= \frac{1}{n(n-1)}\sum_{u\neq v} m_l(y_u,x_{1u};\theta)m_l(y_v,x_{1v};\theta)E\left[\mathbf{1}_{l,jk}(\theta)\frac{\mathbf{1}_{juk}}{\sigma_l(\theta,\cdot)}\frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,\cdot)}|x_u,x_v\right] \\
&\quad +O_P(n^{-3/2})
\end{aligned}
$$

uniformly over $\theta\in\partial(\Theta_0(P))$.

Now note that the first term in $nU_n^4 f_l(j,k,u,v)$ has a limiting distribution of a linear combination of independent centered χ_1^2 distributions (and so $nU_n^4 f_l(j,k,u,v)$), indexed by $\theta\in\partial(\Theta_0(P))$. Therefore, uniformly over $\theta\in\partial(\Theta_0(P))$, we have

$$
n\widehat{Q}_n(\theta)=\sum_{l\in\overline{\mathcal{I}}}\left\{\sum_{a=1}^{\infty} q_{l,a}(\theta)\left(\chi_{l,a}^2-1\right)\right\}+o_P(1)
$$

due to the functional limit theorem for degenerated U-processes by Nolan and Pollard (1988) (see also Gregory (1997) and Neuhaus (1977)) and by Assumption 5.2 (i) where $q_{l,1}(\theta),q_{l,2}(\theta),\ldots$ are the ordered eigenvalues of the operator $\mathbb{H}$ in (14) such that $|q_{l,1}(\theta)|\geq|q_{l,2}(\theta)|\geq\ldots$ for all $\theta\in\partial(\Theta_0(P))$

with the kernel function $\phi_l(w_u, w_v; \theta)$ in (15). Note that from (42), we obtain

$$\begin{aligned}
&\sum_{l \in \overline{\mathcal{I}}} U_n^4 f_l(j,k,u,v) \\
= \ & \frac{1}{n(n-1)} \sum_{u \neq v} \sum_{l \in \overline{\mathcal{I}}} \left\{ m_l(w_u;\theta) m_l(w_v;\theta) E\left[\mathbf{1}_{l,jk}(\theta) \frac{\mathbf{1}_{juk}}{\sigma_l(\theta,\cdot)} \frac{\mathbf{1}_{jvk}}{\sigma_l(\theta,\cdot)} | x_u, x_v \right] \right\} + O_P(n^{-3/2}) \\
= \ & \frac{1}{n(n-1)} \sum_{u \neq v} \phi(w_u, w_v; \theta) + O_P(n^{-3/2}).
\end{aligned}$$

Now note that

$$|\phi(w_u, w_v; \theta)| \leq \sum_{l \in \overline{\mathcal{I}}} |\phi_l(w_u, w_v; \theta)|$$

from which it follows that

$$N(\varepsilon, \mu, \mathcal{F}_\phi, \mathbb{F}_\phi) \leq \prod_{l \in \overline{\mathcal{I}}} N(\varepsilon, \mu, \mathcal{F}_{\phi_l}, \mathbb{F}_{\phi_l})$$

where $N(\varepsilon, \mu, \mathcal{F}, \mathbb{F})$ denotes the covering number of radius ε for the functional space $\mathcal{F}$ with envelope $\mathbb{F}$ where μ is a measure on $\{(x_j, x_k) \in \mathcal{X} \times \mathcal{X} : x_j < x_k\}$. Also note that $\mathbb{F}_\phi = \sum_{l \in \overline{\mathcal{I}}} \mathbb{F}_{\phi_l}$. Therefore, by Assumption 5.1 and 5.2 (i), we have

$$n\widehat{Q}_n(\theta) = \sum_{a=1}^\infty \overline{q}_a(\theta) \left(\chi_a^2 - 1 \right) + o_P(1)$$

uniformly over $\theta \in \partial(\Theta_0(P))$ due to the functional limit theorem of Nolan and Pollard (1988). ■

A.4 Proof of Lemma 5.2

Proof. We have

$$F_n(x, \theta, P) = P\left[\sum_{l \in \overline{\mathcal{I}}} U_n^4 f_l(j,k,u,v) \leq x \right]$$

from the definition of $F_n(x, \theta, P)$ and $U_n^4 f_l(j,k,u,v)$. We also have

$$F_0(x, \theta, P) = P\left[\sum_{a=1}^\infty \overline{q}_a(\theta) \left(\chi_a^2 - 1 \right) \leq x \right]$$

from Lemma 5.1. Then, due to Theorem 1.1 and 1.2 of Bentkus and Götze (1999) and Assumption 5.2 (ii), we have

$$\sup_x |F_n(x, \theta, P) - F_0(x, \theta, P)| = O_P(n^{-1})$$

and the claim follows. ■

A.5 Proof of Theorem 5.2

Proof. Without loss of generality, we consider the model with moment inequalities only (i.e., $d_{m_2}=0$). We show that the condition (20) holds under Assumptions 5.3 and 5.4 by contradiction. When one or more of Assumptions 5.3 and 5.4 are violated, we can find a subsequence $\{n_\kappa\}$ and a corresponding sequence $(\theta_{n_\kappa}, P_{n_\kappa}) \in \Theta \times \boldsymbol{P}$ such that $\theta_{n_\kappa} \in \Theta_0(P_{n_\kappa})$ and

$$\sup_x \left\{F_{b_{n_\kappa}}(x, \theta_{n_\kappa}, P_{n_\kappa}) - F_{n_\kappa}(x, \theta_{n_\kappa}, P_{n_\kappa})\right\} \to \delta \tag{43}$$

for some $\delta > 0$. With some abuse of notation, also let n_κ denote the size of the subsequence $\{n_\kappa\}$. We also define a function $(A)_+ = A \cdot 1(A \geq 0)$. Then, the sample criterion function obtained from the subsequence $\{n_\kappa\}$ is written as

$$\widehat{Q}_{n_k}(\theta_{n_k}, P_{n_k}) = \frac{1}{n_k(n_k-1)} \sum_{j\neq k} \sum_{l=1}^{d_{m_1}} \left(\frac{\widehat{H}_{P_{n_\kappa},l}(\theta_{n_k}, x_j, x_k)}{\sigma_{P_{n_\kappa},l}(\theta_{n_k}, x_j, x_k)} \right)_+^2 .$$

To simplify the notation, we will replace $\widehat{\sigma}_{P,l}(\theta, x_j, x_k)$ with $\sigma_{P,l}(\theta, x_j, x_k)$ since $\widehat{\sigma}_{P,l}(\theta, x_j, x_k) \to \sigma_{P,l}(\theta, x_j, x_k)$ uniformly over $\theta \in \Theta$ and $\boldsymbol{P}$ by Assumption 4.1 (f) and this replacement does not change the asymptotic result. Define

$$\begin{aligned} G_{l,jik,n_\kappa}(\theta, P) &= \frac{1}{\sqrt{n_\kappa}} \sum_{i=1}^{n_\kappa} \frac{m_l(y_i, x_{1i};\theta)\mathbf{1}_{jik} - E_P\left[m_l(y_i, x_{1i};\theta)\mathbf{1}_{jik}\right]}{\sigma_{P,l}(\theta, x_j, x_k)} \text{ and} \\ T_{l,jik,n_\kappa}(\theta, P) &= \sqrt{n_\kappa} E_P\left[m_l(y_i, x_{1i};\theta)\mathbf{1}_{jik}\right] / \sigma_{P,l}(\theta, x_j, x_k) \end{aligned}$$

We can write

$$n_\kappa \widehat{Q}_{n_\kappa}(\theta_{n_k}, P_{n_k}) = \frac{1}{n_\kappa(n_\kappa - 1)} \sum_{j\neq k} \sum_{l=1}^{d_{m_1}} \left(G_{l,jik,n_\kappa}(\theta_{n_\kappa}, P_{n_\kappa}) + T_{l,jik,n_\kappa}(\theta_{n_\kappa}, P_{n_\kappa})\right)_+^2 .$$

Further define

$$b_{n_\kappa} \widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k}, P_{n_k}) = \frac{1}{b_{n_\kappa}(b_{n_\kappa} - 1)} \sum_{j\neq k} \sum_{l=1}^{d_{m_1}} \left(G_{l,jik,b_{n_\kappa}}(\theta_{n_\kappa}, P_{n_\kappa}) + T_{l,jik,n_\kappa}(\theta_{n_\kappa}, P_{n_\kappa})\right)_+^2$$

and define its distribution function as $\widetilde{F}_{b_{n_\kappa}}(x, \theta_{n_\kappa}, P_{n_\kappa})$.

Since we choose $b_{n_\kappa} < n_\kappa$, we have $T_{l,jik,b_{n_\kappa}}(\theta_{n_\kappa}, P_{n_\kappa}) \geq T_{l,jik,n_\kappa}(\theta_{n_\kappa}, P_{n_\kappa})$. It follows that

$$b_{n_\kappa} \widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k}, P_{n_k}) \leq b_{n_\kappa} \widehat{Q}_{b_{n_\kappa}}(\theta_{n_k}, P_{n_k}).$$

Applying the same arguments in the proof of Lemma 5.1, under Assumptions 5.3 and 5.4, we obtain

$$b_{n_\kappa} \widehat{Q}_{b_{n_\kappa}}(\theta_{n_k}, P_{n_k}) = \sum_{l \in \mathcal{I}(\theta_{n_k}, P_{n_k})} \left\{ \textstyle\sum_{a=1}^{\infty} q_{l,a}(\theta_{n_k}, P_{n_k}) \left(\chi_{l,a}^2 - 1\right) \right\} + o_{P_{n_k}}(1)$$

and so

$$\widetilde{F}_{b_{n_\kappa}}(x,\theta_{n_\kappa},P_{n_\kappa}) \geq F_{b_\kappa}(x,\theta_{n_\kappa},P_{n_\kappa})$$

by the stochastic dominance. Combining this result with (43), we further obtain

$$\sup_x \left\{ \widetilde{F}_{b_{n_\kappa}}(x,\theta_{n_\kappa},P_{n_\kappa}) - F_{n_\kappa}(x,\theta_{n_\kappa},P_{n_\kappa}) \right\} > 0. \tag{44}$$

Now we will derive the limit distribution of $\widetilde{F}_{b_{n_\kappa}}(x,\theta_{n_\kappa},P_{n_\kappa})$ and $F_{n_\kappa}(x,\theta_{n_\kappa},P_{n_\kappa})$ and obtain the contradiction. It is again obvious that

$$n_\kappa \widehat{Q}_{n_\kappa}(\theta_{n_k},P_{n_k}) = \sum_{l\in\mathcal{I}(\theta_{n_k},P_{n_k})} \left\{ \textstyle\sum_{a=1}^{\infty} q_{l,a}(\theta_{n_k},P_{n_k}) \left(\chi^2_{l,a}-1\right) \right\} + o_{P_{n_\kappa}}(1)$$

from Lemma 5.1. Now we consider

$$\begin{aligned} & b_{n_\kappa} \widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k},P_{n_k}) \\ = \ & \frac{1}{b_{n_\kappa}(b_{n_\kappa}-1)} \sum_{j\neq k} \sum_{l=1}^{d_{m_1}} \left(G_{l,jik,b_{n_\kappa}}(\theta_{n_\kappa},P_{n_\kappa}) + T_{l,jik,n_\kappa}(\theta_{n_\kappa},P_{n_\kappa}) \right)_+^2 \\ = \ & \frac{b_{n_\kappa}}{b_{n_\kappa}(b_{n_\kappa}-1)} \sum_{j\neq k} \sum_{l=1}^{d_{m_1}} \left(\frac{\widehat{H}_{P_{n_\kappa},l}(\theta_{n_k},x_j,x_k)}{\sigma_{P_{n_\kappa},l}(\theta_{n_k},x_j,x_k)} + \left(\frac{\sqrt{n_\kappa}-\sqrt{b_{n_\kappa}}}{\sqrt{b_{n_\kappa}}} \right) \frac{E_{P_{n_\kappa}}\left[m_l(y_i,x_{1i};\theta_{n_\kappa})\mathbf{1}_{jik}\right]}{\sigma_{P_{n_\kappa},l}(\theta_{n_\kappa},x_j,x_k)} \right)_+^2 . \end{aligned}$$

Here note that the term $\left(\frac{\sqrt{n_\kappa}-\sqrt{b_{n_\kappa}}}{\sqrt{b_{n_\kappa}}} \right) \frac{E_{P_{n_\kappa}}\left[m_l(y_i,x_{1i};\theta_{n_\kappa})\mathbf{1}_{jik}\right]}{\sigma_{P_{n_\kappa},l}(\theta_{n_\kappa},x_j,x_k)} \to_{n\to\infty} -\infty$ for those moment inequalities that do not bind and becomes zero for those binding moment inequalities. Therefore, for those (θ_{n_k},P_{n_k}) such that $\mathcal{I}(\theta_{n_k},P_{n_k})$ is not empty, we obtain

$$\begin{aligned} b_{n_\kappa} \widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k},P_{n_k}) &= \frac{b_{n_\kappa}}{b_{n_\kappa}(b_{n_\kappa}-1)} \sum_{j\neq k} \sum_{l\in\mathcal{I}\left(\theta_{n_k},P_{n_k}\right)} \left(\frac{\widehat{H}_{P_{n_\kappa},l}(\theta_{n_k},x_j,x_k)}{\sigma_{P_{n_\kappa},l}(\theta_{n_k},x_j,x_k)} \right)_+^2 + o_{P_{n_\kappa}}(1) \qquad (45) \\ &= \sum_{l\in\mathcal{I}(\theta_{n_k},P_{n_k})} \left\{ \textstyle\sum_{a=1}^{\infty} q_{l,a}(\theta_{n_k},P_{n_k}) \left(\chi^2_{l,a}-1\right) \right\} + o_{P_{n_\kappa}}(1). \end{aligned}$$

and for those (θ_{n_k},P_{n_k}) such that $\mathcal{I}(\theta_{n_k},P_{n_k})$ is empty, we obtain $b_{n_\kappa}\widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k},P_{n_k}) \to 0$. Therefore, we conclude $b_{n_\kappa}\widetilde{Q}_{b_{n_\kappa}}(\theta_{n_k},P_{n_k})$ has the same limit distribution with that of $n_\kappa\widehat{Q}_{n_\kappa}(\theta_{n_k},P_{n_k})$. Moreover, Assumption 5.4 ensures that uniformly over $\theta_{n_k} \in \Theta_0(P_{n_k})$,

$$\begin{aligned} \sup_x \left| \widetilde{F}_{b_{n_\kappa}}(x,\theta_{n_\kappa},P_{n_\kappa}) - F_0(x,\theta_{n_\kappa},P_{n_\kappa}) \right| &= o_P(1) \text{ and} \\ \sup_x \left| F_{n_\kappa}(x,\theta_{n_\kappa},P_{n_\kappa}) - F_0(x,\theta_{n_\kappa},P_{n_\kappa}) \right| &= o_P(1). \end{aligned}$$

Therefore, (44) cannot be true. This completes the proof. ■

References

[1] Andrews, D.W.K., S. Berry, and P. Jia (2004), "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working paper, Yale University.

[2] Andrews, D.W.K. and P. Guggenberger (2007), "Validity of Subsampling and "Plug-in" Asymptotic" Inference for Parameters Defined by Moment Inequalities," Working paper.

[3] Andrews, D.W.K. and X. Shi (2008), "Inference for Parameters Defined by Conditional Moment Inequalities," Working paper, Yale University.

[4] Andrews, D.W.K. and G. Soares (2007), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Working paper, Yale University.

[5] Bentkus, V. and F. Götze (1999), "Optimal Bounds in Non-Gaussian Limit Theorems for U-statistics," *Annals of Probability*, 27-1, 454-521.

[6] Beresteanu, A., I. Molchnov, and F. Molinari (2008), "Sharp Identification Regions in Games," Working paper, Duke University, University of Bern, and Cornell University.

[7] Berry, S. and E. Tamer (2007), "Identification in Models of Oligopoly Entry," in Advances in Economics and Econometrics: Theory and Application, ed. by R. Blundell, W.K. Newey, and T. Persson, II-2, 46-85, Cambridge University Press.

[8] Billingsley, P. (1995), *Probability and Measure*, New York: Wiley and Sons.

[9] Beresteanu, A. and F. Molinari (2006), "Asymptotic Properties for a Class of Partially Identified Models," working paper, Duke University and Cornell University.

[10] Bresnahan, T.F. and P.C. Reiss (1990), "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531-553.

[11] Bresnahan, T.F. and P.C. Reiss (1991), "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977-1009.

[12] Chernozhukov, V., H. Hong, and E. Tamer (2004), "Inference on Parameter Sets in Econometric Models," working paper.

[13] Ciliberto, F. and E. Tamer (2004), "Market Structure and Multiple Equilibria in Airline Markets," Working paper, University of Virginia and Northwestern University.

[14] Dominguez, M. and I. Lobato (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601-1615.

[15] Guggenberger, P., J. Hahn, and K. Kim (2006), "Specification Testing Under Moment Inequalities," forthcoming in *Economics Letters*.

[16] Gregory, G.G. (1977), "Large Sample Theory for U-statistics and Tests of Fit," *Annals of Statistics*, 5, 110-123.

[17] Hall, P. (1984), "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators", *Journal of Multivariate Analysis*, 14, 1-16.

[18] Horowitz, J. and Manski, C. (1995), "Identification and Robustness with Contaminated and Corrupted data," *Econometrica*, 63, 281-302.

[19] Imbens, G. and C. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

[20] Khan, S. and E. Tamer (2006), "Quantile Minimum Distance Estimation of Randomly Censored Regression Models," working paper.

[21] Manski, C.F. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 319-323.

[22] Manski, C.F. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

[23] Mikusheva, A. (2008), "Uniform Inferences in Autoregressive Models," forthcoming in *Econometrica*.

[24] Neuhaus, G. (1977), "Functional Limit Theorem for U-statistics in the Degenerate Case," *Journal of Multivariate Analysis*, 7, 424-439.

[25] Nolan, D. and D. Pollard (1987), "U-Processes: Rates of Convergence," *Annals of Statistics*, 15, 780-799.

[26] Nolan, D. and D. Pollard (1988), "Functional Limit Theorems for U-processes," *Annals of Statistics*, 16-3, 1291-1298.

[27] Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57-5, 1027-1057.

[28] Romano, J.P. and A.M. Shaikh, (2006), "Inference for Identified Set in Partially Identified Econometric Models," working paper.

[29] Romano, J.P. and A.M. Shaikh, (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models," forthcoming *Journal of Statistical Planning and Inference* (Special Issue in Honor of T. W. Anderson).

[30] Rosen, A.R., 2005, "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," working paper.

[31] Sherman, R.P. (1994a), "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators," *Annals of Statistics*, 22, 439-459.

[32] Sherman, R.P. (1994b), "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator", *Econometric Theory*, 10, 372-395.

[33] Tamer, E.T. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies* 70, p.147-165.