Inference for Partially Identified Models with Inequality Moment Constraints (Preliminary and Incomplete)

Gustavo B. Soares
Yale University

Mar, 2006

1 Introduction

This paper discusses inference on partially identified models that are specified in terms of a finite number of unconditional moment restrictions. By partially identified models we mean econometric models that are not restrictive enough to specify a one-to-one map between the parameter space and the probability distribution of the data. This paper proposes specification tests for such models. We also discuss how to construct confidence sets for identifiable parameters.

The empirical research on oligopoly games has for a long time addressed the difficulties in performing inference on discrete games with multiple equilibria.[1] In fact, many game theoretic models are not identified because, typically, there will exist a group of models which are observationally equivalent. In these models, the lack of identification often comes from the multiplicity of equilibria but it may also come from the informational structure of the game which may not be restrictive enough to yield point identification of the parameters. Tamer (2003) formally discusses identification in a 2×2 discrete game with complete information and states conditions under which one can obtain point identification of the parameters of interest.

In this paper, we assume the model belongs to a parametric family $\mathcal{P} := \left\{\mathbb{P}^{(\theta,\gamma)} : (\theta, \gamma) \in \Theta \times \Gamma\right\}$. We assume that $\Theta \subset \mathbb{R}^{p+q}$ but Γ is potentially

[1] See Bresnahan and Reiss (1990, 1991), Berry (1992) and Reiss (1996).

infinite dimensional. We assume there exists a pair $\left(\theta^0, \gamma^0\right) \in \Theta \times \Gamma$ which defines the true data generating process $\mathbb{P}^0$. The true probability law $\mathbb{P}^0$ specifies a set of moment constraints, which are satisfied by the true parameters $\left(\theta^0, \gamma^0\right)$, but can potentially be satisfied by other points in $\Theta \times \Gamma$ as well. For that reason the models are only partially identified. Here we focus on the particular class of partially identified models that are specified in terms of a finite number of equality as well as inequality unconditional moment restrictions. Many models discussed in the literature can be rewritten using this framework as we illustrate with the examples of Section 2.

Models with moment inequality constraints arise in many contexts. Tamer (2003) shows that a 2×2 discrete game with complete information specifies a finite number of conditional moment inequality constraints. Ciliberto and Tamer (2004) extend the discussion to the case with many players: they show how the inequality conditional moment constraints imply a finite set of unconditional moment inequality restrictions that can be used for inference. Andrews, Berry and Jia (2005), henceforth ABJ (2005), and Pakes, Porter, Ho, and Ishii (2005), henceforth PPHI (2005), discuss how many behavioral models naturally imply a finite set of unconditional moment inequality restrictions.

The main goal of this paper is to derive confidence sets for the true parameter θ^0 or, in some cases, for a sub-vector of the true parameter θ^0. We achieve this goal by utilizing the structure of the model to derive the asymptotic distributions of a function $M_n(\cdot)$, the Modified Method of Moments criterion function introduced by Manski and Tamer (2002), evaluated at fixed points $\theta \in \Theta$. This insight goes back to at least Anderson-Rubin (1949) and has been recently utilized by Kleibergen (2001) and Moreira (2003) to construct confidence intervals that are robust to weak instruments in the linear regression model.[2] Kleibergen (2005) and Guggenberger and Smith (2005) extend the same insight to models specified in terms of a finite set of equality moment constraints. Using the same type of insight, PPHI (2005), Rosen (2005) and Shaikh (2005) construct confidence intervals for θ^0 in partially identified models with inequality moment constraints.

In the literature on partially identified models, the set Θ_I^0 is typically defined to be the set of points in Θ that minimize some criterion function $Q(\cdot)$, i.e., $\Theta_I^0 = Arg\min_{\theta \in \Theta} Q(\theta)$. Manski and Tamer (2002), for instance, propose to estimate Θ_I^0 using a sample analogue of the criterion function

[2] See Andrews and Stock (2005) for a survey on weak intruments and further results.

$Q(\cdot)$, $Q_n(\cdot)$. The models we discuss in this paper fall into this framework and the methods developed in Manski and Tamer (2002) for estimating Θ_I^0 directly apply to the criterion function $M_n(\cdot)$ discussed in Section 3.

Chernozhukov, Hong and Tamer (2005) discuss how to use subsampling to construct confidence sets that asymptotically cover the identified set Θ_I^0 with probability at least $1-\alpha$, i.e. they show how to use subsampling to construct $\{CS_n : n \geq 1\}$ such that

$$\liminf_{n\to\infty} \mathbb{P}^0 \left\{\Theta_I^0 \subset CS_n\right\} \geq 1-\alpha.$$

Shaikh (2005) shows that the problem of constructing confidence sets with this property is equivalent to the problem in which one wants to test a family of hypotheses indexed by $\theta \in \Theta$, $H_\theta : \theta \in \Theta_I^0$, controlling the probability of even one false rejection. The results in Chernozhukov, Hong and Tamer (2005) and Shaikh (2005) directly apply to the criterion functions discussed in Section 3. Shaikh (2005) also establishes conditions under which the convergence of his method is uniform in a certain class of probability measures, which is an issue we address in Section 5.

After describing the general model and discussing some examples in Section 2, we proceed as described below:

In Section 3, we introduce a version of the Modified Method of Moments criterion function and derive some asymptotic results. As is sometimes the case with some statistics that utilize the structure of the model under a fixed point $\theta \in \Theta$, the asymptotic distribution will, in general, depend on the value of θ.[3] In our case, it will depend on a parameter $c^0(\theta)$ that identifies the binding inequalities among the set of inequality moment constraints evaluated at θ. The test statistics introduced by PPHI (2005) and Rosen (2005) also share this feature but their approach to obtaining critical values differ from the approach taken here. Shaikh (2005) also proposes a method for constructing confidence sets for the true parameter value θ^0, but he uses subsampling for obtaining the critical values. His method can be applied to the criterion function discussed here.

In section 4, we discuss how to utilize the results in Section 3 to construct a test of overidentifying restrictions and perform inference on the true value θ^0. The difficulty of estimating $c^0(\theta)$ correctly in finite samples motivated Rosen (2005) to take a conservative approach towards constructing confi-

[3] See Moreira (2003) and Kleibergen (2005) for instance.

dence sets. In section 4, we discuss two methods for estimating $c^0(\theta)$ and how they can be utilized for constructing less conservative confidence sets.

In section 5, we address the issue of uniform consistency in levels for the confidence sets constructed in the previous section. Uniform consistency is an important property for good finite sample behavior of these confidence sets. In Section 6, we investigate further the finite sample performance of $\{CS_n : n \geq 1\}$, as well as alternative methods available in the literature, using Monte Carlo simulations.. Section 7 concludes and the Appendix contains the proofs. We often refer to the Continuous Mapping Theorem such as in van der Vaart (1998, pg. 7) by CMT and to the Portmanteau Theorem such as in van der Vaart (1998, pg. 6) by PT. The indicator function is denoted by $1\{\cdot\}$. Unless stated otherwise, all probability limits are taken with respect to the true probability law $\mathbb{P}^0$ for $n \to \infty$. We denote "*with probability approaching one as* $n \to \infty$" by w.p.a.1.

2 Model

Let $\{X_i : 1 \leq i\}$ be a sequence of $\mathbb{R}^l$-valued random vectors drawn from an unknown joint probability law $\mathbb{P}^{(\theta^0,\gamma^0)}$ belonging to a parametric family $\mathcal{P} := \{\mathbb{P}^{(\theta,\gamma)} : (\theta,\gamma) \in \Theta \times \Gamma\}$. We assume that $\Theta \subset \mathbb{R}^{p+q}$ but γ is potentially infinite dimensional. Many quantities introduced below depend on the true distribution $\mathbb{P}^{(\theta^0,\gamma^0)}$, and they carry a superscript "0" to make this dependence explicit. Hereafter, we denote $\mathbb{P}^{(\theta^0,\gamma^0)}$ simply by $\mathbb{P}^0$.

Let g be a measurable function $g : \mathbb{R}^l \times \Theta \to \mathbb{R}^{d+s}$. Denote $g(X_i,\theta)$ by $g_i(\theta)$, and define the sequence of functions $\{\hat{g}_n : \Theta \to \mathbb{R}^{d+s} : n \geq 1\}$ and $\{\hat{\Omega}_n : \Theta \to \mathbb{R}^{(d+s)\times(d+s)} : n \geq 1\}$ as

$$\begin{aligned} \hat{g}_n(\theta) \; &: \; = n^{-1}\sum_{i=1}^{n} g_i(\theta) \text{ and} \\ \hat{\Omega}_n(\theta) \; &: \; = n^{-1}\sum_{i=1}^{n} g_i(\theta)\, g_i(\theta)' \end{aligned}$$

and the functions $g^0 : \Theta \to \mathbb{R}^{d+s}$ and $\Omega^0 : \Theta \to \mathbb{R}^{(d+s)\times(d+s)}$ as

$$g^0(\theta) \;:=\; \lim_{n\to\infty} E^0 n^{-1} \sum_{i=1}^{n} g_i(\theta) \text{ and}$$
$$\Omega^0(\theta) \;:=\; \lim_{n\to\infty} E^0 n^{-1} \sum_{i=1}^{n} g_i(\theta)\, g_i(\theta)'.$$

We now state a set of high-level assumptions on the model. We assume:

Assumption M: (i) Θ is convex and compact; (ii) The functions $g^0 : \Theta \to \mathbb{R}^{d+s}$ and $\Omega^0 : \Theta \to \mathbb{R}^{(d+s)\times(d+s)}$ are continuous and bounded on Θ, $\sup_{\theta\in\Theta} \|\hat{g}_n(\theta) - g^0(\theta)\| = o_p(1)$ and $\sup_{\theta\in\Theta} \left\|\hat{\Omega}_n(\theta) - \Omega^0(\theta)\right\| = o_p(1)$, and for every $\theta \in \Theta$, $\Omega^0(\theta)$ is a nonsingular positive definite matrix; (iii) For every point in $\theta \in \Theta$, $\Psi_n(\theta) := n^{1/2}(\hat{g}_n(\theta) - g^0(\theta)) \to_d \Psi(\theta)$, where $\Psi(\theta)$ is a $(d+s)$-dimensional normal with mean zero and covariance matrix given by $\Delta^0(\theta) := \Omega^0(\theta) - g^0(\theta)\, g^0(\theta)'$; (iv) For any sequence $\{\rho_n : n \geq 1\}$ converging to zero we have $\sup_{\theta,\theta^*\in\Theta:\|\theta-\theta^*\|<\rho_n} \|\Psi_n(\theta) - \Psi_n(\theta^*)\| = o_p(1)$.

The conditions of Assumption M are general enough to allow $\mathcal{P}$ to include distributions of sequences of $\mathbb{R}^l$-valued random vectors that are iid, inid, and stationary and ergodic. Also, they may hold when the function $g : \mathbb{R}^l \times \Theta \to \mathbb{R}^{d+s}$ is non-smooth (as in the case of simulation based estimators) and/or when it depends upon $\{X_i : 1 \leq i\}$ (as in the form of a preliminary estimator, which can potentially be infinite dimensional.)

We consider models specified by a finite number of moment restrictions which are not fully identified:

Assumption ID: There exists a nonempty subset $\Theta_I^0 \subset \Theta$ such that $\theta^0 \in \Theta_I^0$ and for every $\theta \in \Theta_I^0$ we have

$$g_k^0(\theta) = 0 \text{ for } k = 1, ..., d \text{ and}$$
$$g_k^0(\theta) \leq 0 \text{ for } k = d+1, ..., d+s,$$

where $g_k^0(\theta)$ denotes the k-th coordinate of the $\mathbb{R}^{d+s}$-vector $g^0(\theta)$.

On some occasions, we will split the function $g : \mathbb{R}^l \times \Theta \to \mathbb{R}^{d+s}$ into two functions $g_1 : \mathbb{R}^l \times \Theta \to \mathbb{R}^d$ and $g_2 : \mathbb{R}^l \times \Theta \to \mathbb{R}^s$ corresponding to the first d equality constraints and to the s inequality constraints, respectively.

Assumption ID above may be written as $g_1^0(\theta) = 0$ and $g_2^0(\theta) \leq 0$, while the vector $\hat{g}_n(\theta)$ may be split into two subvectors $\hat{g}_{1n}(\theta) \in \mathbb{R}^d$ and $\hat{g}_{2n}(\theta) \in \mathbb{R}^s$.

In partially identified models, it is common to specify equality moment restrictions as two inequalities. We choose not to take that approach because we require in Assumption M(ii) the matrix $\Omega^0(\theta)$ to be positive definite, which would not necessarily be true if there was perfect collinearity between the elements of $\{g_i(\theta) : 1 \leq i\}$.

The reader will notice that some of the results easily generalize to models specified only in terms of inequalities ($d = 0$). In some circumstances, it is useful to differentiate between equality and inequality restrictions and Assumption ID is designed to make the distinction explicit.

2.1 Examples

Example 1 - Imbens and Manski (2004) Let $\{X_i = (Y_i, W_i) : 1 \leq i\}$ be a sequence of iid bivariate random vectors with joint distribution $F^0 \in \mathcal{F}$, where Y_i takes values on a compact convex subset $V \subset \mathbb{R}$ and $W_i \in \{0, 1\}$, respectively. The reseacher has a sample of $\{(W_i, Y_i W_i) : 1 \leq i \leq n\}$, i.e., the reseacher does not observe the realization of the random variable Y_i when $W_i = 0$. Define the true parameter $\theta^0 := \left(\theta_1^0, \theta_2^0\right)$ where $\theta_1^0 = E^0[Y_i]$ and $\theta_2^0 := E^0[Y_i | W_i = 0]$. The distribution F^0 may be seen as being part of the infinite dimensional nuisance parameter γ^0. Define the function $g : \mathbb{R}^2 \times \Theta \to \mathbb{R}$,

$$g(X_i, \theta) := g(X_i, \theta_1, \theta_2) = \theta_1 - Y_i W_i - \theta_2 (1 - W_i)$$

where $\Theta := V \times V$ is a compact and convex subset of $\mathbb{R}^2$.

[Start of Incomplete Part]

Under certain regularity conditions this satisfies M ... show this later .... for

$$g^0(\theta) = \theta_1 - E^0[Y_i W_i] - \theta_2 E^0[(1 - W_i)],$$

which is such that $g^0\left(\theta^0\right) = 0$, but there are other points in Θ for which $g^0(\theta) = 0$, and that defines Θ_I^0.

[End of Incomplete Part]

Example 2 - Manski and Tamer (2002) Let $\{X_i = (Y_i, Z_i) : 1 \leq i\}$ be a sequence of bounded iid $\mathbb{R}^{1+l}$-valued random vectors with joint distribution $F^0 \in \mathcal{F}$, where Y_i takes values on $V_Y \subset \mathbb{R}$ and $Z_i \in V_Z \subset \mathbb{R}^l$, respectively. The reseacher does not observe the realization of the random variable Y_i when it follows below a certain threshold, say $Y_i \leq 0$, but it does observe the realization of Z_i. Define the true parameter θ^0 to be such that $E^0 \left[Y_i | Z_i = z\right] = z'\theta^0$. The distribution F^0 may be seen as being part of the infinite dimensional nuisance parameter γ^0. Define the function $g : \mathbb{R}^{1+l} \times \Theta \to \mathbb{R}^2$,

$$g\left(X_i, \theta\right) := \left[\begin{array}{c} g_1\left(X_i, \theta\right) \\ g_2\left(X_i, \theta\right) \end{array}\right] := \left[\begin{array}{c} \left(Y_i - Z_i'\theta\right) 1\left\{Y_i > 0\right\} \\ \left(Z_i'\theta\right) 1\left\{Y_i \leq 0\right\} \end{array}\right]$$

where Θ is a compact and convex subset of $\mathbb{R}^2$ containing θ^0.

[Start of Incomplete Part]

Under certain regularity conditions this satisfies M ... show this later .... with

$$g^0\left(\theta\right) = \left[\begin{array}{c} g_1^0\left(\theta\right) \\ g_2^0\left(\theta\right) \end{array}\right] = \left[\begin{array}{c} E^0\left[\left(Y_i - Z_i'\theta\right) 1\left\{Y_i > 0\right\}\right] \\ E^0\left[\left(Z_i'\theta\right) 1\left\{Y_i \leq 0\right\}\right] \end{array}\right]$$

with $g_1^0\left(\theta^0\right) = 0$ and $g_2^0\left(\theta^0\right) \leq 0$.

[End of Incomplete Part]

Example 3 - A Binary Simultaneous Game with Complete Information[4] In a binary simultaneous game with N players, each player $j = 1, ..., N$ decides which action $a_j \in \{0, 1\}$ to take. Let $z \in \mathbb{R}^{d_z}$ for each $j = 1, ..., N$ represent a vector of observed exogenous variables. The vector z may contain variables that describe characteristics of the individual players as well as variables that are common to all players. Let u be a random $\mathbb{R}^N$-vector of latent variables (not observed by the econometrician) known to the players (complete information), which is independent from z and has conditional distribution $F_u(\cdot)$. In this example, we assume $F_u(\cdot)$ is symmetric for simplicity.

[4] This example is a particular case of the binary simultaneous game with complete information discussed in Ciliberto and Tamer (2004) and ABJ (2005).

We assume that player j receives utility $\pi_j(a_j, a_{-j}, z, u)$ from taking action $a_j \in \{0,1\}$ given the other players' vector of actions $a_{-j} \in \{0,1\}^{N-1}$. For simplicity, we assume that when agent j^* takes action $a_{j^*} = 1$, everything else held constant, agent j's payoff decreases by $|\delta_j|$, regardless of her identity j^*. We normalize $\pi_j(0, a_{-j}, z, u) = 0$ for all (z,u) and $j = 1, ..., N$, and assume:

$$\pi_j(1, a_{-j}, z, u) = \delta_j \left(\sum_{j^* \neq j} a_{j^*} \right) + z\beta_j + u_j$$

The parameters $\theta := (\delta, \beta)$ are assumed to belong to some compact set $\Theta \subset \mathbb{R}_-^N \times \mathbb{R}^{Nd_z}$, i.e., we assume $\delta_j \leq 0$ for all $j = 1, ...N$.

In a Nash Equilibrium in Pure Strategies (NE), the vector of actions $a^{NE} \in \{0,1\}^N$ is such that

$$\pi_j\left(a_j^{NE}, a_{-j}^{NE}, z, u\right) \geq \pi_j\left(1 - a_j^{NE}, a_{-j}^{NE}, z, u\right)$$

for each player $j = 1, ..., N$. The Nash Equilibrium in Pure Strategies Assumption imposes that every observed vector of actions $a \in \{0,1\}^N$ in the data is the outcome of a NE of the underlying game.

Berry (1992) shows that under the assumptions above, for any vector (z, u), there exists a NE in Pure Strategies that may be constructed by allowing the more profitable players to choose before the less profitable ones. More formally, we may reorder the players' identities $j = 1, ..., N$ according to the order of $\left\{z\beta_j + u_j : 1 \leq j \leq N\right\}$ and set $a_j^{NE} = 1$ for all j such that $\pi_j\left(1, a_{-j}^{NE}, z, u\right) \geq 0$ and set $a_j^{NE} = 0$ otherwise, where a_{-j}^{NE} is the vector consisting of ones in its first $j-1$ coordinates and zeros in all other coordinates. This is a NE in Pure Strategies because no agent j has an incentive to change the value of a_j^{NE} given that the other players are playing a_{-j}^{NE}.

Using a standard revealed preference argument, we can see that we observe the vector of actions $a = (0, ..., 0)$ if and only if u is such that $\pi_j(1, (0, ..., 0), z, u) \leq 0$ for all $j = 1, ..., N$. Similarly, we observe the vector of actions $a = (1, ..., 1)$ if and only if the vector u is such that $\pi_j(1, (1, ..., 1), z, u) \geq 0$ for all $j = 1, ..., N$. The symmetry of $F_u(\cdot)$ then implies:

$$E\left[1\{a = (0, ..., 0)\} | Z = z\right] - F_u(-z\beta_1, ..., -z\beta_N) = 0 \text{ and}$$
$$E\left[1\{a = (1, ..., 1)\} | Z = z\right] - F_u(\delta_1(N-1) + z\beta_1, ..., \delta_N(N-1) + z\beta_N) = 0.$$

Whenever u falls in the region defined by the intervals

$$\prod_{j=1}^{N} \left[-\delta_j (N-1) - z\beta_j, -z\beta_j\right],$$

the model above fails to yield a full prediction for the vector of actions $a \in \{0,1\}^N$. The incompleteness of the econometric model is in fact a product of the multiplicity of NE associated with the underlying game.

However, the same revealed preference argument can help us generate an additional set of inequality restrictions on the parameters of the model. If we restrict ourselves to the analysis of pure strategy equilibria, the Nash Equilibrium in Pure Strategies Assumption implies that for any given $a_{-j} \in \{0,1\}^{N-1}$, player j is not willing to deviate from his/her chosen strategy $a_j \in \{0,1\}$. Hence, if we observe a vector of actions $a \in \{0,1\}^N$, it must be the case that $\pi_j (a_j, a_{-j}, z, u) \geq 0$ for every $j = 1, ..., N$. This implies that whenever we see a vector of actions $a \in \{0,1\}^N$, the vector u must be such that

$$u_j \geq -\delta_j \left(\sum_{j^* \neq j} a_{j^*}\right) - z\beta_j \text{ for every } j = 1, ..., N.$$

The discussion above suggests that we should observe in the data the vector $a \in \{0,1\}^N$ with a relative frequency that is at least as large as the probability that the vector u satisfies the inequality above, i.e., it implies the conditional moment inequality restriction:

$$E[a|Z=z] - F\left(\delta_1 \left(\sum_{j^* \neq 1} a_{j^*}\right) + z\beta_1, ..., \delta_N \left(\sum_{j^* \neq N} a_{j^*}\right) + z\beta_N\right) \leq 0.$$

Hence, in addition to the two equality constraints generated by the actions $(0, ..., 0)$ and $(1, ..., 1)$, the Nash Equilibrium in Pure Strategies assumption yields a set of at least $2^{N-1} - 2$ inequalities.

Of course, the restrictions above are not the only set of restrictions implied by the conditional moment constraints. In fact, since the variables Z are exogenous, an arbitrary number of equality constraints may be created by multiplying the conditional moment equality constraints by any deterministic function of Z. Similarly, one may create an arbitrary number of inequality constraints by multiplying the conditional moment equality constraints by any strictly positive deterministic function of Z.

Suppose we observe a cross-section of n markets, where the same game is being played in each market $i = 1, ..., n$. Define each $X_i := (a_i, Z_i)$ for $i = 1, ..., n$, and let $\mathbb{P}^0$ be its unknown probability law generated through the game structure, where $\theta^0 := \left(\delta^0, \beta^0\right)$ are the true parameters of the model, i.e., the parameters of the profit function $\pi\left(\cdot; \theta^0\right)$. When coupled with an equilibrium selection rule $\gamma^0 \in \Gamma$ that selects among the different equilibria for a given draw of X_i, the model is then fully specified by $\left(\theta^0, \gamma^0\right) \in \Theta \times \Gamma$. This imposes some additional structure on the model, i.e., it places a restriction over the mechanism that selects among the different possible equilibria in each market i. It imposes that, conditional on X_i, the equilibrium in each market is chosen by the same equilibrium selection rule γ^0 even if the rule itself is left unspecified (fully nonparametric). Additionally, as in Bajari, Hong and Ryan (2005), one may parametrize the equilibrium selection rule and treat Γ parametrically.[5]

Define $g : \mathbb{R}^{N+d_z} \times \Theta \to \mathbb{R}^{2+s}$ as

$$
\begin{aligned}
g_1(X_i, \theta) \quad &: \quad = 1\{a_i = (0, ..., 0)\} - F_u(-Z_i'\beta_1, ..., -Z_i'\beta_N) \\
g_2(X_i, \theta) \quad &: \quad = 1\{a_i = (1, ..., 1)\} - F_u(\delta_1(N-1) + Z_i'\beta_1, ..., \delta_N(N-1) + Z_i'\beta_N) \\
g_k(X_i, \theta) \quad &: \quad = 1\{a_i = a_k\} - F_u\left(\delta_1\left(\sum_{j^* \neq 1} a_{kj^*}\right) + Z_i'\beta_1, ..., \delta_N\left(\sum_{j^* \neq N} a_{kj^*}\right) + Z_i'\beta_N\right),
\end{aligned}
$$

where $a_k \in \{0,1\}^N$ for $1 \leq k \leq 2^N - 2$ belongs to the set of all possible outcome vectors a excluding $a = (0, ..., 0)$ and $a = (1, ..., 1)$.

Standard assumptions on the moments of the sequence $\{Z_i : 1 \leq i\}$ can be imposed in order to guarantee that Assumptions ID and M hold for this model. Typically, these assumptions are the assumptions necessary to prove the Uniform Law of Large Numbers and Weak Convergence of Empirical Processes.[6] In some cases, the sequence $\{Z_i : 1 \leq i\}$ may be allowed to have some degree of distributional heterogeneity as well as some time and/or spatial dependence.

Often, computing the function $F_u(\cdot)$ will involve multidimensional integration, which may be avoided by simulation such as in McFadden (1989) and Pakes and Pollard (1989). The use of simulation to compute $\hat{g}_n(\cdot)$ is common in applications of discrete choice decision. For that reason we avoid

[5] One may parametrize the distribution function $F_u(\cdot)$ as well and treat its parameters as being part of the nuisance parameter $\gamma \in \Gamma$.

[6] See Andrews (1994).

assuming differentiability of $\hat{g}_n(\cdot)$ by imposing only differentiability in the limit through Assumption G (below).[7]

In many cases, the researcher will choose to use a nonparametric estimator $\tau_n(Z_i)$ of the conditional expectation $E[a_i|Z_i]$ instead of simply using the indicator function as suggested above. In this case, the results in Andrews (1994) may be used to verify Assumptions ID and M. In some cases, such as in ABJ (2005), the researcher may choose to classify the data into cells $\left\{C_n^k : 1 \leq k \leq K\right\}$ and use the indicator function $1\left\{X_i \in C_n^k\right\}$ instead of the indicator function $1\{a_i = a_k\}$ used in the example above. The reader may use the arguments found in Pollard (1979) to verify Assumptions ID and M in the case of data-dependent cells.

3 Asymptotic Results

Inference on models defined in terms of inequalities typically depends crucially on which inequalities actually "matter". In our case, the inequalities that do "matter" are the inequalities in Assumption ID that are binding when evaluated at the true parameter $\theta^0 \in \Theta$ ($k \in \{d+1, ..., d+s\}$ such that $g_k^0\left(\theta^0\right) = 0$.)

We introduce a parameter, which may be called the *selection vector*:

Definition 1 *The selection vector is a function $c^0 : \Theta \to \{0,1\}^{d+s}$ such that $c^0(\theta)$ is a $(d+s)$-vector belonging to $\{0,1\}^{d+s}$ whose k-th coordinate, $c_k^0(\theta)$, is equal to one if $k = 1, ..., d$ and equal to $1\{g_k^0(\theta) \geq 0\}$ if $k = d+1, ..., d+s$.*

The vector $c^0(\theta)$ selects the binding constraints among the $d+s$ moment constraints imposed in Assumption ID and is generally unknown. For any selection vector $c \in \{0,1\}^{d+s}$, whose k-th coordinate is denoted by c_k, we define the number of selected constraints to be $|c| = \sum_{k=1}^{d+s} c_k$.

[7] See McFadden (1989) and Pakes and Pollard (1989) for a discussion of Assumption G in this case.

3.1 Continuous Updating Modified Method of Moments

The Modified Method of Moments criterion function was introduced by Manski and Tamer (2002) for $W_n(\theta)$ not depending on $\theta \in \Theta$. Hence, the criterion function $M_n(\theta)$ below is a continuous updating version of the Modified Method of Moments criterion function introduced by them.

Define the map $I : \mathbb{R}^{d+s} \to \mathbb{R}^{d+s}$, which maps the vector $x = (x_1, ..., x_{d+s})' \in \mathbb{R}^{d+s}$ into a vector $I(x) \in \mathbb{R}^{d+s}$ whose elements $i_k(x)$ are defined as

$$i_k(x) := \begin{cases} x_k \text{ if } k = 1, ..., d \\ x_k 1\{x_k > 0\} \text{ if } k = d+1, ..., d+s. \end{cases}$$

Denote the quadratic form $x'Ax$ as $\|x\|_A^2$ and define

$$M_n(\theta) := \left\| I\left(n^{1/2}\hat{g}_n(\theta)\right)\right\|_{W_n(\theta)}^2,$$

where $W_n : \Theta \to \mathbb{R}^{(d+s)\times(d+s)}$ is specified by the researcher and assumed to satisfy:

Assumption W: There exists a nonstochastic continuous function $W^0 : \Theta \to \mathbb{R}^{(d+s)\times(d+s)}$ such that $\sup_{\theta\in\Theta} \|W_n(\theta) - W^0(\theta)\| = o_p(1)$, and for every $\theta \in \Theta$, $W^0(\theta)$ is a nonsingular positive definite diagonal matrix;

We denote the dot product by $\cdot$, i.e., for $x := (x_1, ..., x_{d+s})'$ and $y := (y_1, ..., y_{d+s})' \in \mathbb{R}^{d+s}$, $x \cdot y$ denotes the vector $(x_1y_1, ..., x_{d+s}y_{d+s})' \in \mathbb{R}^{d+s}$. For any fixed $\theta \in \Theta$, define the random variable $\tilde{M}(\theta)$ as

$$\tilde{M}(\theta) := \left\| c^0(\theta) \cdot I(\Psi(\theta))\right\|_{W^0(\theta)}^2.$$

Theorem 2 *Under Assumptions ID, M(ii)-(iii), and W, for every* $\theta \in \Theta_I^0$ *we have* $M_n(\theta) \to_d \tilde{M}(\theta)$.

Remark: The proof of the theorem is given here. After rewriting $n^{1/2}\hat{g}_n(\theta)$ as $n^{1/2}\hat{g}_n(\theta) = \Psi_n(\theta) + n^{1/2}g^0(\theta)$, the result follows directly from the assumptions of the theorem and the CMT.

3.2 Identified Parameters

In this section, we decompose the parameter space Θ into two subspaces $\Theta := \Theta_1 \times \Theta_2$. We assume that the parameter of interest is the subvector $\theta_1 \in \Theta_1 \subset \mathbb{R}^p$ and that $\theta_2 \in \Theta_2 \subset \mathbb{R}^q$ is a vector of nuisance parameters. This decomposition is not arbitrary. We assume that these nuisance parameters are identified for fixed values of θ_1 and we define $\Theta_{1I}^0 := \{\theta_1 \in \Theta_1 : (\theta_1, \theta_2) \in \Theta_I^0 \text{ for some } \theta_2 \in \Theta_2\}$ as the set of identifiable subvectors θ_1.

Assumption ID2: For any $\theta_1 \in \Theta_{1I}^0$ there exists a unique $\theta_2(\theta_1) \in \Theta_2$ such that $\theta := (\theta_1, \theta_2(\theta_1)) \in \Theta_I^0$.

Assumption ID2 implies that for a fixed $\theta_1 \in \Theta_1^0$ the sub-vector θ_2 is point-identified. In particular, it imposes that there exists a unique $\theta_2^0 := \theta_2(\theta_1^0)$, where $(\theta_1^0, \theta_2^0) = \theta^0$. In models where there are only inequality moment restrictions $(d = 0)$ and the function g^0 is continuous, Assumption ID2 is typically not valid. This is so because in general, for any fixed value $\theta_1 \in \Theta_1^0$, there will be a sequence $\{\theta_{2j} : j \geq 1\}$ for which $g_k^0(\theta_1, \theta_{2j}) \leq 0$ for every $k \in \{1, ..., s\}$. An exception to this rule is the case where three or more inequalities $k \in \{1, ..., s\}$ specify a unique $\theta_2(\theta_1) \in \Theta_2$ for every fixed value $\theta_1 \in \Theta_1^0$. However, in this case, we may combine these three or more inequalities to construct an equality moment restriction.

Example 1 - Imbens and Manski (2004) *In this example the function g^0 is such that for each fixed value θ_1 there is a unique θ_2 such that*

$$g^0(\theta_1, \theta_2) = \theta_1 - E^0[Y_i W_i] - \theta_2 E^0[(1 - W_i)] = 0,$$

as long as $E^0[(1 - W_i)] > 0$.

Example 3 - A Binary Simultaneous Game with Complete Information *For any value of the vector $\delta := (\delta_1, ..., \delta_N)$, under the regularity conditions of the Implicit Function Theorem, there is a unique value of $\beta := (\beta_1, ..., \beta_N)$ that satisfies the equation:*

$$\mathbb{P}^0\{a_i = (0, ..., 0)\} - E^0[F_u(-Z_i'\beta_1, ..., -Z_i'\beta_N)] = 0.$$

Additionally, under similar regularity conditions, there is a unique value of the vector β that satisfies the equation:

$$\mathbb{P}^0\left\{a_i = (1, ..., 1)\right\} - E^0\left[F_u\left(\delta_1 (N-1) + Z_i'\beta_1, ..., \delta_N (N-1) + Z_i'\beta_N\right)\right] = 0$$

for a given value of the vector $\delta := (\delta_1, ..., \delta_N)$. Hence, by setting $\theta_1 := \delta$ and $\theta_2 := \beta$, Assumption ID2 is verified under the the regularity conditions of the Implicit Function Theorem.

As Example 3 in Section 2.1 illustrates, the use of simulation to compute $\hat{g}_n(\cdot)$ is essential to applications of discrete choice decision. For that reason we avoid assuming differentiability of $\hat{g}_n(\cdot)$ by imposing only differentiability in the limit through Assumption G(i) (below). Assumption G(i) is analogous to the assumption of asymptotic differentiability of the sample moments found in Pakes and Pollard (1989), and the reader is referred to their paper for a more comprehensive discussion. Assumption G(ii) (below) is sufficient to guarantee the criterion function has a unique minimum. For any $\varepsilon > 0$, let $B_\varepsilon(\theta)$ denote an open ball around $\theta \in \Theta \subset \mathbb{R}^{p+q}$ and assume:

Assumption G: (i) For any $\theta \in \Theta$ there exists a nonrandom $(d+s) \times (p+q)$ matrix $G^0(\theta)$ with full column rank such that $g^0\left(\bar{\theta}\right) = g^0(\theta) + G^0(\theta)\left(\bar{\theta} - \theta\right) + o\left(\left\|\bar{\theta} - \theta\right\|\right)$ as $\left\|\bar{\theta} - \theta\right\| \to 0$ for $\theta \in \Theta \cap B_\varepsilon(\theta)$ for some $\varepsilon > 0$. (ii) For any $\theta \in \Theta$, the $(d+s) \times q$-submatrix $G_2^0(\theta)$ composed of the columns of $G^0(\theta)$ corresponding to the parameters in the subvector θ_2 is such that the matrix composed of its first d rows has full column rank q and $d \geq q$.

For any $\theta_1 \in \Theta_1^0$, define the random function $\tilde{M}(\theta_1, \cdot)$ on $\mathbb{R}^q$ as

$$\tilde{M}(\theta_1, t) := \left\|c^0(\theta_1, \theta_2(\theta_1)) \cdot I\left(\Psi(\theta_1, \theta_2(\theta_1)) + G_2^0(\theta_1, \theta_2(\theta_1))\, t\right)\right\|^2_{W^0(\theta_1, \theta_2(\theta_1))}$$

and define a random $\mathbb{R}^q$-vector $\tilde{t}$ such that

$$\tilde{M}\left(\theta_1, \tilde{t}\right) := \inf_{t \in \mathbb{R}^q} \tilde{M}(\theta_1, t).$$

In general, $\tilde{t}$ will not be uniquely defined, but it is uniquely defined under Assumption G(ii). Given $\theta_1 \in \Theta_1$, Assumption G(ii) makes the function $\tilde{M}(\theta_1, t)$ strictly convex in t for every realization of $\Psi(\theta_1, \theta_2(\theta_1))$ and that guarantees the uniqueness of $\tilde{t}$ for every realization of $\Psi(\theta_1, \theta_2(\theta_1))$.

We now state the main result of this section:

Theorem 3 *Under Assumptions ID, ID2, M, W and G, for any* $\theta_1 \in \Theta_{1I}^0$ *and* $(\theta_1, \theta_2(\theta_1)) \in int(\Theta)$*, the sequence of estimators* $\hat{\theta}_{2n}(\theta_1) \in \Theta_2$ *defined as*

$$M_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \leq \inf_{\theta_2 \in \Theta_2} M_n(\theta_1, \theta_2) + o_p(1)$$

is such that

(a) $M_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \to_d \inf_{t \in \mathbb{R}^q} \tilde{M}(\theta_1, t)$ *and*

(b) $n^{1/2}\left(\hat{\theta}_{2n}(\theta_1) - \theta_2(\theta_1)\right) \to_d \tilde{t}$.

Remark: Assumption G(ii) is typically only valid in models with equality constraints and therefore the result does not extend to the case where the model is specified only in terms of inequalities ($d = 0$) in a trivial way.

Remark: We can always artificially augment the parameter space by some extra parameter to cover the case where the identified set is a singleton, i.e., $\Theta_I^0 := \left\{\theta^0\right\}$.

The matrix $G_2^0(\theta)$ can be interpreted as the derivative of $g^0(\cdot)$ with respect to θ_2 evaluated at $(\theta_1, \theta_2(\theta_1))$. This interpretation suggests an estimator $\hat{G}_{2n}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right)$ composed of q columns $\hat{G}_{j,2n}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right)$ for $j = 1, ..., q$:

$$\hat{G}_{j,2n}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) := \varepsilon_n^{-1}\left(\hat{g}_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1) + \varepsilon_n u_j\right) - \hat{g}_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right)\right), \tag{1}$$

where ε_n is a sequence that converges in probability to one and u_j is a q-dimensional unit vector whose j-th coordinate is equal to one and all other coordinates are equal to zero. This is in fact the estimator proposed by Pakes and Pollard (1989), which is commonly used to compute standard errors for estimates of simulation based estimators. Under the conditions of Theorem 3 we have $\hat{\theta}_{2n}(\theta_1) \to_p \theta_2(\theta_1)$, and using Assumptions M(ii) and G(i) together with the CMT we have $\hat{G}_{2n}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \to_p G_2^0(\theta_1, \theta_2(\theta_1))$ for every fixed $\theta_1 \in \Theta_1^0$.

4 Inference

In this section, we discuss how to use the asymptotic results derived in the previous section to perform inference on the true parameter $\theta^0 \in \Theta_I^0$ (or on $\theta_1^0 \in \Theta_{1I}^0$ in the case where θ_2 is identified).

4.1 A Test of Overidentifying Restrictions

We may be interested in testing the null hypothesis $H_0 : \mathbb{P}^0 \in \mathcal{P}^*$ against the alternative $H_1 : \mathbb{P}^0 \notin \mathcal{P}^*$, where $\mathcal{P}^* := \left\{\mathbb{P}^{(\theta_1^*,\theta_2,\gamma)} : (\theta_2, \gamma) \in \Theta_2 \times \Gamma\right\}$ for some known null value $\theta_1^* \in \Theta_1$. The result in Theorem 3(a) can then be used to derive a test of overidentifying restrictions. Under the assumptions of Theorem 3, we have

$$M_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) \to_d \inf_{t\in\mathbb{R}^q} \tilde{M}(\theta_1^*, t)$$

under the null.

Under Assumptions M(ii) and W we have $\hat{g}_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) = g^0\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) + o_p(1)$ and $W_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) = W^0\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) + o_p(1)$. The CMT then gives us

$$\begin{aligned}
\inf_{\theta_2\in\Theta_2} \left\|I\left(g^0(\theta_1^*, \theta_2)\right)\right\|^2_{W^0(\theta_1^*,\theta_2)} &\leq \left\|I\left(g^0\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right)\right)\right\|^2_{W^0\left(\theta_1^*,\hat{\theta}_{2n}(\theta_1^*)\right)} \\
&= n^{-1}M_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) + o_p(1).
\end{aligned}$$

Therefore, under the alternative,

$$\inf_{\theta_2\in\Theta_2} \left\|I\left(g^0(\theta_1^*, \theta_2)\right)\right\|^2_{W^0(\theta_1^*,\theta_2)} > 0$$

by Assumptions ID and W, which imples that $M_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) \to_p +\infty$, under the alternative.

The discussion above suggests that a test $H_0 : \mathbb{P}^0 \in \mathcal{P}^*$ against the alternative $H_1 : \mathbb{P}^0 \notin \mathcal{P}^*$ may be constructed if we can find a sequence of critical values $\{\nu^\alpha_{1n}(\theta_1^*) : n \geq 1\}$ (depending on θ_1^*) which consistently estimates the $(1-\alpha)$ quantile of the random variable $\inf_{t\in\mathbb{R}^q} \tilde{M}(\theta_1^*, t)$. Given this sequence, we have:

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{P}^0\left\{M_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) \leq \nu^\alpha_{1n}(\theta_1^*)\right\} &\geq 1-\alpha \text{ for } \mathbb{P}^0 \in \mathcal{P}^* \\
\limsup_{n\to\infty} \mathbb{P}^0\left\{M_n\left(\theta_1^*, \hat{\theta}_{2n}(\theta_1^*)\right) \leq \nu^\alpha_{1n}(\theta_1^*)\right\} &= 0 \text{ otherwise.}
\end{aligned}$$

We address the issue of finding this sequence $\{\nu^\alpha_{1n}(\theta_1^*) : n \geq 1\}$ in Section 4.3 below.

4.2 Confidence Sets

We want to find a sequence of sets $\{CS_n : n \geq 1\}$ with $CS_n \subset \Theta$ such that $\liminf_{n\to\infty} \mathbb{P}^0 \left\{\theta^0 \in CS_n\right\} \geq 1-\alpha$. The basic idea is to exploit the duality between confidence sets and hypothesis testing to construct confidence sets CS_n that have the correct coverage rate asymptotically.[8]

If for each $\theta \in \Theta$ we can find a sequence of critical values $\{\nu_n^\alpha(\theta) : n \geq 1\}$ (depending on θ) such that

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}^0 \{M_n(\theta) \leq \nu_n^\alpha(\theta)\} &\geq 1-\alpha \text{ for } \theta \in \Theta_I^0 \text{ and} \\ \limsup_{n\to\infty} \mathbb{P}^0 \{M_n(\theta) \leq \nu_n^\alpha(\theta)\} &= 0 \text{ otherwise,} \end{aligned}$$

we can then define the confidence set $CS_n := \{\theta \in \Theta : M_n(\theta) \leq \nu_n^\alpha(\theta)\}$. Because for any $\theta \in \Theta$, $\theta \in CS_n$ if and only if $M_n(\theta) \leq \nu_n^\alpha(\theta)$, the sequence of sets $\{CS_n : n \geq 1\}$ will then be such that

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}^0 \{\theta \in CS_n\} &\geq 1-\alpha \text{ for } \theta \in \Theta_I^0 \text{ and} \\ \limsup_{n\to\infty} \mathbb{P}^0 \{\theta \in CS_n\} &= 0 \text{ otherwise.} \end{aligned}$$

In particular, $\liminf_{n\to\infty} \mathbb{P}^0 \left\{\theta^0 \in CS_n\right\} \geq 1-\alpha$.

Analogously, if the parameter θ_2^0 is identified, as in Assumption ID2, we may substitute $M_n(\theta)$ above for $\inf_{\theta_2\in\Theta_2} M_n(\theta_1,\theta_2)$ and define $CS_n^1 := \{\theta_1 \in \Theta_1 : \inf_{\theta_2\in\Theta_2} M_n(\theta_1,\theta_2) \leq \nu_{1n}^\alpha(\theta_1)\}$. The sequence of confidence sets $\{CS_n^1 : n \geq 1\}$ will have the correct coverage rate asymptotically for the parameter θ_1^0 if we can find a sequence of critical values $\{\nu_{1n}^\alpha(\theta_1) : n \geq 1\}$ (depending on θ_1) such that

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}^0 \left\{\inf_{\theta_2\in\Theta_2} M_n(\theta_1,\theta_2) \leq \nu_{1n}^\alpha(\theta_1)\right\} &\geq 1-\alpha \text{ for } \theta_1 \in \Theta_{1I}^0 \text{ and} \\ \limsup_{n\to\infty} \mathbb{P}^0 \left\{\inf_{\theta_2\in\Theta_2} M_n(\theta_1,\theta_2) \leq \nu_{1n}^\alpha(\theta_1)\right\} &= 0 \text{ otherwise.} \end{aligned}$$

In order to construct these confidence sets CS_n or CS_n^1 we need to find sequences of critical value functions $\{\nu_n^\alpha(\theta) : n \geq 1\}$ and $\{\nu_{1n}^\alpha(\theta_1) : n \geq 1\}$ with such characteristics.

[8] See Theorem 4 in Lehman (1986, Chap. 3).

4.3 Critical Values

The main dificulty of finding critical values for $M_n(\theta)$, with $\theta \in \Theta_I^0$, lies in the fact that their asymptotic distributions are not pivotal, they depend on the parameters $c^0(\theta)$, $\Delta^0(\theta)$ and $W^0(\theta)$. The same applies to the case where θ_2 is identified, but the asymptotic distribution of $M_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right)$ depends additionally on the matrix $G_2(\theta)$.

Under Assumptions M(ii) and W, the matrices $\Delta^0(\theta)$ and $W^0(\theta)$ can be consistently estimated by $\hat{\Delta}_n(\theta) := \hat{\Omega}_n(\theta) - \hat{g}_n(\theta)\hat{g}_n(\theta)'$ and $W_n(\theta)$ and do not pose a problem for computing the critical values. As in (1), the same applies to the matrix $G_2(\theta_1)$ in the case where the parameter θ_2 is identified.

Hence, provided that we can find $c_n(\theta) \to_p c^0(\theta)$, the $(1-\alpha)$-quantiles of the distribution of $\|c^0(\theta) \cdot \Psi(\theta)\|^2_{W^0(\theta)}$ may be consistently estimated by simulating a standard normal random $\mathbb{R}^{d+s}$-vector Z^* many times and computing the $(1-\alpha)$-quantiles of the distribution of the random variable $\left\|c_n(\theta) \cdot \hat{\Delta}_n(\theta)^{1/2} Z^*\right\|^2_{W_n(\theta)}$. Provided that we can find $c_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \to_p c^0(\theta_1, \theta_2(\theta_1))$, the case where the parameter θ_2 is identified can be treated analogously because $\hat{\Delta}_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \to_p \Delta^0(\theta_1, \theta_2(\theta_1))$ and $W_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \to_p W^0(\theta_1, \theta_2(\theta_1))$ under Assumptions M(ii) and W.

In many applications resampling methods have been used to avoid the estimation of nuisance parameters when computing critical values for tests and confidence sets. However, for the models discussed here the bootstrap will typically not consistently estimate the asymptotic distribution of $M_n(\theta)$ (or $\inf_{\theta_2 \in \Theta_2} M_n(\theta_1, \theta_2)$ in the case where the parameter θ_2 is identified). The incosistency of the bootstrap for estimating the asymptotic distribution of $M_n(\theta)$ follows from the lack of regularity in the asymptotic distribution of $M_n(\theta)$. Typically, for any $u \in \mathbb{R}^{p+q}$ such that $\|u\| < \infty$ the asymptotic distribution of $M_n\left(\theta + n^{-1/2}u\right)$ will depend on u. Using Theorem 2.3 in Beran (1997), we may show that the bootstrap distribution of $M_n(\theta)$ will not be a consistent. Bickel, Gotze, and van Zwet (1997, pg. 7) also address a similar case in their Example 6.

An alternative resampling method for obtainnig critical values for the asymptotic distributions of $M_n(\theta)$ (or $\inf_{\theta_2 \in \Theta_2} M_n(\theta_1, \theta_2)$ in the case where the parameter θ_2 is identified) is subsampling. Under the conditions discussed above, the limiting distribution of these test statistics is in fact continuous

(See Lemma ??). Hence, at least in the iid case, the consistency of subsampling critical values follows from Theorem 2.2.1 in Politis and Romano (1999). However, this does not hold in general when there are no equality moment restrictions ($d = 0$) and $g_k^0(\theta) < 0$ for all $k \in \{1, ..., s\}$. Shaikh (2005) shows the consistency in levels of tests based on subsampling critical values for functions similar to $M_n(\theta)$ when there are no equality moment constraints ($d = 0$). The methods developed here should be viewed as an alternative to subsampling methods such as the one discussed in Shaikh (2005). Andrews and Guggenberger (2005) also discuss the use of subsampling in this case.

In order to avoid the estimation of the selection vector $c^0(\theta)$, conservative procedures have been proposed in the literature. For the case of $M_n(\theta)$ for instance, a conservative procedure is proposed by PPHI (2005) for their test of overidentifying restrictions. Using our notation, their procedure is equivalent to taking the critical values to be of the $(1-\alpha)$-quantiles of the distribution of $\|I(\Psi(\theta))\|^2_{W^0(\theta)}$, which may be consistently estimated by simulating a standard normal random $\mathbb{R}^{d+s}$-vector Z^* many times and computing $(1-\alpha)$-quantiles of the distribution of $\left\| I\left(\hat{\Delta}_n(\theta)^{1/2} Z^*\right)\right\|^2_{W_n(\theta)}$.

In many applications, conservative procedures for computing the critical values should work fairly well. In particular, they should work well in the case where there are few inequality moment restrictions, i.e., the difference between $|c^0(\theta)|$ and $d+s$ (or $\sup_{\theta \in \Theta_I^0} |c^0(\theta)|$ in some cases) is small. However, as Example 3 in Section 2.1 illustrates, there are many applications where these procedures may not generate the desired results. In games with complete information involving many players, $|c^0(\theta)|$ may vary significantly across different $\theta \in \Theta_I^0$. Constructing a procedure that effectively tries to account for that may be desirable in these applications.

Here we introduce two procedures for estimating $c^0(\theta)$. One of the procedures is a modification of a moment selection procedure introduced by Andrews (1999a). The second procedure is based on a method suggested by Andrews (1999b, sec 6.5). In general, the second procedure will be less computationally intensive than the first one. Provided that we can consistently estimate the selection vector $c^0(\theta)$, the quantiles of the random variables $\tilde{M}(\theta)$ (or $Arg\inf_{t \in \mathbb{R}^q} \tilde{M}(\theta_1, t)$ in the case where θ_2 is identified) may be consistently estimated by simulation.

4.3.1 Modified Moment Selection Criterion

In order to estimate $c^0(\theta)$, we need to select the binding moment constraints among all constraints in Assumption ID. This can be viewed as a moment selection problem, and we now discuss how to select these binding moment constraints.

Let C be the set of possible selection vectors, which in general (but not always) will be defined as $C := \left\{c \in \{0,1\}^{d+s} : c_k = 1 \text{ for } k = 1, ..., d\right\}$.[9] Define the Modified Moment Selection Criterion (MMSC) function to be

$$MMSC(\theta, c) := \left\|n^{1/2} c \cdot I(-\hat{g}_n(\theta))\right\|^2_{W_n(\theta)} - h(|c|)\kappa_n,$$

where the function $h(\cdot)$ and the sequence $\{\kappa_n : n \geq 1\}$ are specified by the researcher and assumed to satisfy:

Assumption MMSC: (i) $h(\cdot)$ is strictly increasing and (ii) $\kappa_n \to \infty$ as $n \to \infty$ and $\kappa_n = o(n)$.

Among the selection vectors belonging to the set C, the MMSC will penalize the selection vectors with non-binding contraints. On the other hand, it introduces a "bonus term" $h(|c|)\kappa_n$ that rewards the selection vectors that utilize more moment restrictions. Typical examples of $h(|c|)\kappa_n$ used in the moment selection literature are: $|c|\log n$ (BIC), $2|c|$ (AIC) and $\varsigma |c| \log\log n$ for some $\varsigma > 2$ (HQIC).

Define the estimated selection vector $\{\hat{c}_n : \Theta \to C : n \geq 1\}$ to be

$$\hat{c}_n(\theta) := Arg\min_{c \in C} MMSC(\theta, c).$$

We now state a consistency result for $\hat{c}_n(\theta)$:

Theorem 4 *For any $u \in \mathbb{R}^{p+q}$ such that $\|u\| < \infty$ let $\{u_n : n \geq 1\}$ be an arbitrary sequence of $\mathbb{R}^{p+q}$-vectors converging to u. Under Assumptions ID, M, W, G(i), and MMSC, for any fixed $\theta \in \Theta_I^0$ such that $\theta \in int(\Theta)$, we have $\hat{c}_n\left(\theta + n^{-1/2}u_n\right) \to_p c^0(\theta)$.*

[9] As in Imbens and Manski (2004) and Rosen (2005), the researcher may choose to exploit the structure of the model to rule out certain selection vectors $c \in C$.

Remark: The assumption of $\theta \in int(\Theta)$ is only used to guarantee that the functions $\hat{g}_n(\theta)$ and $W_n(\theta)$ are defined on $(\theta + n^{-1/2}u_n)$ for n finite but sufficiently large. Assumption G(i) is only needed to guarantee that $\Psi_n(\theta + n^{-1/2}u_n) = O_p(1)$ for any fixed $\theta \in \Theta_I^0$. At a fixed point $\theta \in \Theta_I^0$, we have $\hat{c}_n(\theta) \rightarrow_p c^0(\theta)$ under Assumptions ID, M, W and MMSC.

The theorem establishes that $\hat{c}_n(\theta)$ is a consistent estimator of $c^0(\theta)$. That also holds true for $\hat{c}_n(\hat{\theta}_n)$, where $\hat{\theta}_n$ is some $n^{1/2}$-consistent estimator of θ such as $\hat{\theta}_n := (\theta_1, \hat{\theta}_{2n}(\theta_1))$ in Theorem 3. However, even in simple cases, ex-post model selection inference can have very poor performance in finite samples.[10] Because of its discreteness, $c^0(\theta)$ will be harder to estimate correctly when $g^0(\theta)$ has some coordinates that are close to zero but not exactly zero. We may avoid the inaccuracy of the asymptotic approximation due to the estimation of $c^0(\theta)$ by conservatively selecting more moments.

The fact that MMSC only penalizes the non-binding contraints implies that $\hat{c}_n(\theta)$ will typically be such that $c^0(\theta) \leq \hat{c}_n(\theta)$. This feature of MMSC implies that $\hat{c}_n(\theta)$ will tend to select more contraints than the constraints that are actually binding in the limit, generating conservative confidence intervals.

Define $J_n^*(\theta, v)$ to be the distribution function of the random variable

$$\left\| \hat{c}_n(\theta) \cdot \hat{\Delta}_n(\theta)^{1/2} Z^* \right\|_{W_n(\theta)}^2,$$

where Z^* is a multivariate standard normal random $\mathbb{R}^{d+s}$-vector and define

$$\nu_n^\alpha(\theta) := \inf\{v \in \mathbb{R}_+ : J_n^*(\theta, v) \geq 1 - \alpha\}.$$

Analogously, define $J_{1n}^*(\cdot)$ to be the distribution function of the random variable

$$\inf_{t \in \mathbb{R}^q} \left\| c_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) \cdot I\left(\hat{\Delta}_n^{1/2}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) Z^* + G_{2n}\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right) t\right) \right\|_{W_n\left(\theta_1, \hat{\theta}_{2n}(\theta_1)\right)}^2$$

and define

$$\nu_{1n}^\alpha(\theta) := \inf\{v \in \mathbb{R}_+ : J_{1n}^*(v) \geq 1 - \alpha\}.$$

Note that the distributions of $\tilde{M}(\theta)$ and $\inf_{t \in \mathbb{R}^q} \tilde{M}(\theta_1, t)$ are continuous at all points in $\mathbb{R}_+$. Hence, under Assumptions M(ii) and W (and G(i) in

[10] See Leeb and Potscher (2003) for instance.

the case where θ_2 is identified), the CMT gives us $\nu_n^\alpha(\theta) \to_p \nu_\infty^\alpha(\theta)$ and $\nu_{1n}^\alpha(\theta) \to_p \nu_{1\infty}^\alpha(\theta)$, where $\nu_\infty^\alpha(\theta)$ and $\nu_{1\infty}^\alpha(\theta)$ are the $(1-\alpha)$-quantiles of the random variables $\tilde{M}(\theta)$ and $\inf_{t\in\mathbb{R}^q} \tilde{M}(\theta_1, t)$, respectively.[11]

4.3.2 Selection Based on Wald Tests

A procedure based on Wald tests is discussed in Andrews (1999b, sec 6.5) and it is also related to issues discussed in Andrews (2000). The procedure is based on a sequence of one-sided Wald tests for $g_k^0(\theta) \geq 0$ against $g_k^0(\theta) < 0$ for each $k \in \{d+1, ..., d+s\}$. The estimator $\tilde{c}_n(\theta)$ introduced below is constructed by excluding the constraints k for which $g_k^0(\theta) \geq 0$ is rejected. Hence, for $n \in \mathbb{N}$ sufficiently large, we will typically select $\tilde{c}_n(\theta)$ with more constraints than implied by the vector $c^0(\theta)$, generating more conservative critical values.

Under suitable conditions such as the ones stated in Chow and Teicher (1978, pg. 343 and 344) for instance, the sequence $\{g_i(\theta) : 1 \leq i\}$ will typically satisy the Law of the Iterated Logarithm. We assume it holds in Assumption LIL below:

Assumption LIL: Let $u \in \mathbb{R}^{p+q}$ such that $\|u\| < \infty$ and $\{u_n : n \geq 1\}$ be an arbitrary sequence of $\mathbb{R}^{p+q}$-vectors converging to u. For any fixed $\theta \in \Theta_I^0$ such that $\theta \in int(\Theta)$ and $j \in \{1, ..., s\}$, we have:

$$\begin{aligned}
\liminf_{n\to\infty} (2n \log\log n)^{-1/2} \sum_{i=1}^{n} \frac{g_{d+j,n}\left(X_i, \theta + n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta + n^{-1/2}u_n\right)^{1/2}} &= -1 \text{ if } g_{d+j}^0(\theta) = 0 \text{ and} \\
\lim_{n\to\infty} \frac{n^{1/2}\hat{g}_{d+j,n}\left(\theta + n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta + n^{-1/2}u_n\right)^{1/2}} &= -\infty \text{ if } g_{d+j}^0(\theta) < 0
\end{aligned}$$

w.p.a.1, where $\hat{\omega}_{d+j,n}\left(\theta + n^{-1/2}u_n\right)$ is the sample mean of $\left\{g_{d+j}\left(X_i, \theta + n^{-1/2}u_n\right)^2 : 1 \leq i \leq n\right\}$.

The assumption of $\theta \in int(\Theta)$ is only used to guarantee that the functions $\hat{g}_n(\theta)$ and $W_n(\theta)$ are defined on $\left(\theta + n^{-1/2}u_n\right)$ for n finite but sufficiently large. We demonstrate how to verify Assumption LIL above with Example 2 of Section 2.1 for a fixed value $\theta \in \Theta_I^0$. The extention to the case $\left(\theta + n^{-1/2}u_n\right)$ is straightforward under a differentiability assumption such as Assumption G.

[11] In the case where there are no equality moment constraints ($d = 0$) but $g_k^0(\theta) = 0$ for some $k \in \{1, ..., s\}$, the same argument applies as long as $\alpha < 1/2$.

Example 2 *Let $\theta \in \Theta_I^0$ be such $g_2^0(\theta) := E^0[(Z_i'\theta)1\{Y_i \le 0\}] = 0$.*

$$\liminf_{n\to\infty} (2n\log\log n)^{-1/2} \sum_{i=1}^{n} \frac{g_2(X_i,\theta)}{\hat{\omega}_{2,n}(X_i,\theta)^{1/2}} =$$

$$\liminf_{n\to\infty} \left[\left(\frac{E^0\left[(Z_1'\theta)^2 1\{Y_1 \le 0\}\right]}{n^{-1}\sum_{i=1}^n (Z_i'\theta)^2 1\{Y_i \le 0\}} \right)^{1/2} \times \right.$$

$$\left. (2n\log\log n)^{-1/2} \sum_{i=1}^{n} \frac{(Z_i'\theta)1\{Y_i \le 0\}}{E^0\left[(Z_1'\theta)^2 1\{Y_1 \le 0\}\right]^{1/2}} \right] =$$

$$\liminf_{n\to\infty} \left(\frac{E^0\left[(Z_i'\theta)^2 1\{Y_i \le 0\}\right]}{n^{-1}\sum_{i=1}^n (Z_i'\theta)^2 1\{Y_i \le 0\}} \right)^{1/2} (-1) = -1$$

where the second equality holds almost surely by the Law of the Iterated Logarithm such as in Chow and Teicher (1978, pg. 343 and 344) since

$$\left\{ \frac{(Z_i'\theta)1\{Y_i \le 0\}}{E^0\left[(Z_1'\theta)^2 1\{Y_1 \le 0\}\right]^{1/2}} : 1 \le i \right\}$$

are independent random variables with mean zero and variance one. The last equality also holds almost surely by the Strong Law of Large Numbers, such as in Chow and Teicher (1978, pg. 121), since $\{(Z_i, Y_i) : i \ge 1\}$ are bounded random vectors.

On the other hand, if $\theta \in \Theta_I^0$ is such that $g_2^0(\theta) := E^0[(Z_i'\theta)1\{Y_i \le 0\}] < 0$, we may write:

$$\frac{n^{1/2}\hat{g}_{2,n}(\theta)}{\hat{\omega}_{2,n}(\theta)^{1/2}} = \frac{V^0[(Z_i'\theta)1\{Y_i \le 0\}]^{1/2}}{\hat{\omega}_{2,n}(\theta)^{1/2}} \frac{n^{1/2}(\hat{g}_{2,n}(\theta) - g_2^0(\theta))}{V^0[(Z_i'\theta)1\{Y_i \le 0\}]^{1/2}} + \frac{n^{1/2} g_2^0(\theta)}{\hat{\omega}_{2,n}(\theta)^{1/2}}.$$

Since $\{(Z_i, Y_i) : i \ge 1\}$ are bounded random vectors, the first term on the right hand side of the expression above is $O_p(1)$ by the Strong Law of Large Numbers such as in Chow and Teicher (1978, pg. 121) and the Lindeberg Central Limit Theorem such as in Chow and Teicher (1978, pg. 291). Applying the Strong Law of Large Numbers to $\hat{\omega}_{2,n}(\theta)$ once again, the second term on the right hand side of the expression above diverges to $-\infty$ w.p.a.1.

Define the first d coordinates of the vector $\tilde{c}_n(\theta)$ to be equal to one and its last s coordinates to be

$$\tilde{c}_{d+j,n}(\theta) := 1\left\{\frac{n^{1/2}\hat{g}_{d+j,n}(X_i,\theta)}{\hat{\omega}_{2,n}(X_i,\theta)^{1/2}} > -\eta_n^j\right\},$$

for $j \in \{1,...,s\}$, where the sequence $\{\eta_n^j : n \geq 1 \text{ and } j = 1,...,s\}$ satisfies the assumption below:

Assumption η: The sequence of $\mathbb{R}^s$-vectors $\{\eta_n := (\eta_n^1,...,\eta_n^s) : n \geq 1\}$ satisfies $\lim_{n\to\infty}\eta_n = 0_{(s)}$ and $\liminf_{n\to\infty}\eta_n^j(2\log\log n)^{1/2} > 1$ for all $j \in \{1,...,s\}$.

Under Assumptions LIL and η above, for each coordinate $j = 1,...,s$, we have

$$\mathbb{P}^0\left\{\liminf_{n\to\infty}\left(\frac{n^{1/2}\hat{g}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)^{1/2}} + \eta_n^j\right) > 0\right\} = \left\{\begin{array}{l} 1 \text{ if } g_{d+j}^0(\theta) = 0 \text{ and} \\ 0 \text{ if } g_{d+j}^0(\theta) < 0. \end{array}\right.$$

Hence, if $g_{d+j}^0(\theta) = 0$, for $n \in \mathbb{N}$ large enough we have:

$$\begin{aligned}\mathbb{P}^0\left\{\tilde{c}_{d+j,n}(\theta) = 1\right\} &= \mathbb{P}^0\left\{\frac{n^{1/2}\hat{g}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)^{1/2}} + \eta_n^j > 0\right\} \\ &\geq \mathbb{P}^0\left\{\liminf_{n\to\infty}\left(\frac{n^{1/2}\hat{g}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)^{1/2}} + \eta_n^j\right) > 0\right\} = 1.\end{aligned}$$

On the other hand, if $g_{d+j}^0(\theta) < 0$,

$$\mathbb{P}^0\left\{\tilde{c}_{d+j,n}(\theta) = 0\right\} = \mathbb{P}^0\left\{\frac{n^{1/2}\hat{g}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)}{\hat{\omega}_{d+j,n}\left(\theta+n^{-1/2}u_n\right)^{1/2}} \leq -\eta_n^j\right\} \to 1.$$

Since the convergence is joint for all $j \in \{1,...,s\}$, we have $\tilde{c}_n\left(\theta+n^{-1/2}u_n\right) \to_p c^0(\theta)$ for any $\theta \in \Theta_I^0$ such that $\theta \in int(\Theta)$. Valid critical values $\tilde{\nu}_n^\alpha(\theta)$ and $\tilde{\nu}_{1n}^\alpha(\theta)$ can then be constructed using $\tilde{c}_n(\theta)$ and $\tilde{c}_n\left(\theta_1,\hat{\theta}_{2n}(\theta_1)\right)$, as in the previous section.

4.4 Uniform Consistency in Levels

In the previous sections, we derived confidence sets $\{CS_n : n \geq 1\}$ that were asymptotically consistent in levels for a fixed data generating process $\mathbb{P}^0 \in \mathcal{P}$. However, for a fixed value of $n \in \mathbb{N}$, the confidence sets $\{CS_n : n \geq 1\}$ could have very different finite sample coverage rates for different $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$. In fact, in many models we can find a $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$ such that $\mathbb{P}^{(\theta,\gamma)} \{\theta \in CS_n\} < 1 - \alpha$ for any value of $n \in \mathbb{N}$ finite.

In this section, we investigate conditions on the class of models $\mathcal{P}$ under which we can guarantee uniform consistency in levels across different data generating processes $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$. This is an important property for confidence sets to have, and it is often considered as being part of the definition of confidence sets.[12] On the other hand, it is a property that is typically hard to prove in general. For instance, even in the simple case of a sample mean, Bahadur and Savage (1956) show that this is an impossible task if we do not impose restrictions over the class of models $\mathcal{P}$. Hence, we need to impose restrictions on the parametric family $\mathcal{P}$ to guarantee uniform consistency in levels of $\{CS_n : n \geq 1\}$. In order to do that, we introduce some notation:

For any arbitrary probability law $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$ define the functions on Θ:

$$
\begin{aligned}
g^{(\theta,\gamma)}(\theta^*) \quad &: \quad = \lim_{n\to\infty} E^{(\theta,\gamma)} n^{-1} \sum_{i=1}^{n} g_i(\theta^*), \\
\Omega^{(\theta,\gamma)}(\theta^*) \quad &: \quad = \lim_{n\to\infty} E^{(\theta,\gamma)} n^{-1} \sum_{i=1}^{n} g_i(\theta^*) \, g_i(\theta^*)' \text{ and} \\
\Delta^{(\theta,\gamma)}(\theta^*) \quad &: \quad = \Omega^{(\theta,\gamma)}(\theta^*) - g^{(\theta,\gamma)}(\theta^*) \, g^{(\theta,\gamma)}(\theta^*)'
\end{aligned}
$$

for $\theta^* \in \Theta$.

Here we decompose the parameter $\gamma \in \Gamma$ into three components $\gamma := (\gamma_1, \gamma_2, \gamma_3) \in \Gamma_1 \times \Gamma_2 \times \Gamma_3$. We rewrite Assumption ID for a generic probability law $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$ by assuming that there exists a nonempty subset $\Theta_I^{\gamma} \subset \Theta$ containing the true parameter θ such that we have

$$
\begin{aligned}
g_k^{(\theta,\gamma)}(\theta) \quad &= \quad 0 \text{ for } k = 1, ..., d \text{ and} \\
\left(g_k^{(\theta,\gamma)}(\theta) - \gamma_1 \right) \quad &= \quad 0 \text{ for } k = d+1, ..., d+s,
\end{aligned}
$$

where $g_k^{(\theta,\gamma)}(\theta)$ denotes the k-th coordinate of the $\mathbb{R}^{d+s}$-vector $g^{(\theta,\gamma)}(\theta)$, and

[12] See van der Vaart (1998, Ch. 14) for instance.

$\gamma_1 \in \Gamma_1 := \mathbb{R}^s_-$. We may define $\gamma_2 := vec\left(\Omega^{(\theta,\gamma)}(\theta)\right) \in \Gamma_2$ and $\gamma_3 \in \Gamma_3$ to be some infinite dimensional parameter.

Assumption P: Let $\mathcal{P}$ be the parametric class

$$\mathcal{P} := \left\{\mathbb{P}^{(\theta,\gamma)} : \theta \in \Theta \text{ and } \gamma := (\gamma_1, \gamma_2, \gamma_3) \in \Gamma := \Gamma_1 \times \Gamma_2 \times \Gamma_3\right\}.$$

To any arbitrary sequence $\left\{\left(\theta_{n,h}, \gamma_{n,h}\right) \in \Theta \times \Gamma : n \geq 1\right\}$ for which

$$\left(\theta_{n,h}, n^{1/2}\gamma_{1,n,h}, \gamma_{2,n,h}, \gamma_{3,n,h}\right) \to \left(\theta^h, h_1, h_2, h_3\right)$$

for $\left(\theta^h, h_1, h_2, h_3\right) \in cl(\Theta) \times \left\{\mathbb{R}^s_- \cup \{-\infty\}^s\right\} \times cl(\Gamma_2) \times cl(\Gamma_3)$ corresponds a sequence of $\left\{\mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \in \mathcal{P} : n \geq 1\right\}$ satisfying the following conditions:

$$n^{1/2}\left(\hat{g}_n(\theta_{n,h}) - \left(0'_{(d)}, \gamma'_{1,n,h}\right)'\right) \to {}_d N\left(0, \Delta^h\right), \tag{2}$$

$$\hat{\Delta}_n(\theta_{n,h}) \to {}_p \Delta^h \text{ and} \tag{3}$$

$$W_n(\theta_{n,h}) \to {}_p W^h \tag{4}$$

under $\mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)}$, where Δ^h, Ω^h and W^h are positive definite matrices, with W^h being diagonal.

We now state a result about the uniform consistency in levels for $\{CS_n; n \geq 1\}$:

Theorem 5 *For $\mathcal{P}$ satisfying Assumption P we have:*

$$\liminf_{n\to\infty} \inf_{\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}:\theta \in \Theta_I^\gamma} \mathbb{P}^{(\theta,\gamma)}\{\theta \in CS_n\} \geq 1 - \alpha.$$

Theorem 5 above provides conditions under which the confidence sets $\{CS_n : n \geq 1\}$ discussed in the previous section have good uniform behavior. Intuitively, this asymptotic result implies that $\{CS_n : n \geq 1\}$ should perform well in large (but finite) samples across different $\mathbb{P}^{(\theta,\gamma)} \in \mathcal{P}$ whenever $\mathcal{P}$ satisfies Assumption P above. In the next section, we investigate further the finite sample performance of $\{CS_n : n \geq 1\}$ as well as alternative methods available in the literature using Monte Carlo simulations.

5 Monte Carlo

6 Conclusion (Under Construction)

We point out that constructing the sets CS_n (or CS_n^1) may be computationally intensive, depending on the dimension of Θ (or Θ_1 for the case where θ_2 is identified). The main problem is that the critical value $\nu_n^\alpha(\theta)$ depends on each point θ. This problem is common in certain statistics that are designed to be robust against failure of identification.[13] As described in Kleibergen (2005), one may construct CS_n by selecting a grid of K parameters $\{\theta_k \in \Theta : 1 \leq k \leq K\}$ and computing the criterion function and $\nu_n^\alpha(\theta)$ for the points in this grid.

7 Appendix

Proof of Theorem 3. For notational simplicity, for any fixed $\theta_1 \in \Theta_{1I}^0$, denote $\theta_2(\theta_1)$ by θ_2 and the pair $(\theta_1, \theta_2(\theta_1))$ simply by θ.

Part (a)

We start by showing consistency of $\hat{\theta}_{2n}(\theta_1)$. By the definition of $\hat{\theta}_{2n}(\theta_1)$ we have

$$M_n\left(\theta_1, \hat{\theta}_{2n}\right) \leq \inf_{\theta_2^* \in \Theta_2} M_n(\theta_1, \theta_2^*) + o_p(1) \leq M_n(\theta) + o_p(1) = O_p(1),$$

[13] See Moreira (2003) and Kleibergen (2005).

where the last convergence result follows by Theorem 2.

By Assumptions ID, ID2, M(ii) and W we have that for any $\varepsilon > 0$ there exists $K_\varepsilon > 0$ such that

$$\inf_{\theta_2^* \in \Theta_2 : \theta_2^* \notin N_{\theta_1}} \left\| I\left(g^0\left(\theta_1, \theta_2^*\right)\right)\right\|^2_{W^0(\theta_1,\theta_2^*)} = K_\varepsilon,$$

where $N_{\theta_1} := \left\{\bar{\theta}_2 \in \Theta_2 : \left\|\bar{\theta}_2 - \theta_2\right\| < \varepsilon\right\}$. Hence, by Assumption M(ii), for any $\bar{\theta}_2 \notin N_{\theta_1}$, we have $M_n\left(\theta_1, \bar{\theta}_2\right) \rightarrow_p +\infty$ and therefore $\hat{\theta}_{2n}\left(\theta_1\right) \in N_{\theta_1}$ w.p.a.1, implying that $\hat{\theta}_{2n}\left(\theta_1\right) \rightarrow_p \theta_2\left(\theta_1\right)$.

Define $\hat{t}_n := n^{1/2}\left(\hat{\theta}_{2n}\left(\theta_1\right) - \theta_2\left(\theta_1\right)\right)$. Consistency of $\hat{\theta}_{2n}\left(\theta_1\right)$ is equivalent to $t_n n^{-1/2} = o_p\left(1\right)$. We now show that $\hat{t}_n = O_p\left(1\right)$.

By the definition of $\hat{t}_n$, we have

$$M_n\left(\theta_1, \theta_2 + n^{-1/2}\hat{t}_n\right) \leq \inf_{\theta_2^* \in \Theta_2} M_n\left(\theta_1, \theta_2^*\right) + o_p\left(1\right) \leq M_n\left(\theta_1, \theta_2\right) = O_p\left(1\right),$$

where the last convergense result follows by Theorem 2.

From Assumption W and the result above, all the coordinates $k \in \{1, ..., d+s\}$ of the vector $n^{1/2} I\left(\hat{g}_n\left(\theta_1, \theta_2 + n^{-1/2}\hat{t}_n\right)\right)$ are $O_p\left(1\right)$, in particular, the coordinates $k \in \{1, ..., d\}$. Therefore, for $k \in \{1, ..., d\}$, using Assumption G(i) and the fact that $n^{-1/2}\hat{t}_n = o_p\left(1\right)$, we have

$$\begin{aligned} O_p\left(1\right) &= n^{1/2}\hat{g}_{k,n}\left(\theta_1, \theta_2 + n^{-1/2}\hat{t}_n\right) \\ &= n^{1/2}\hat{g}_{k,n}\left(\theta\right) + G^0_{k,2}\left(\theta\right)\hat{t}_n + \left\|\hat{t}_n\right\| o_p\left(1\right), \end{aligned}$$

where $G^0_{k,2}\left(\theta\right)$ is the k-th row of the matrix $G^0_2\left(\theta\right)$. Since $\theta \in \Theta^0_I$, by Assumptions ID, ID2 and M(iii), we have $n^{1/2}\hat{g}_{k,n}\left(\theta\right) = O_p\left(1\right)$ for all $k \in \{1, ..., d\}$. This and Assumption G(ii) imply that there exists a constant $K > 0$ such that $0 \leq \left\|\hat{t}_n\right\| \leq K$ with probability arbitrarily close to one.

Hence,

$$M_n\left(\theta_1, \theta_2 + \hat{t}_n n^{-1/2}\right) = \inf_{t \in J_n} M_n\left(\theta_1, \theta_2 + t n^{-1/2}\right) + o_p\left(1\right),$$

where $J_n := \left\{t \in \mathbb{R}^q : \|t\| < K \text{ and } \theta_2 + n^{-1/2}t \in \Theta_2\right\}$.

From Assumptions ID2, M(iv) and G(i), and from the triangle inequality we have

$$\sup_{t\in J_n}\left\|n^{1/2}\hat{g}_n\left(\theta_1,\theta_2+tn^{-1/2}\right)-n^{1/2}g^0\left(\theta\right)-G_2^0\left(\theta\right)t-\Psi_n\left(\theta\right)\right\|=$$

$$=\sup_{t\in J_n}\left\|n^{1/2}\hat{g}_n\left(\theta_1,\theta_2+tn^{-1/2}\right)\pm n^{1/2}g^0\left(\theta_1,\theta_2+tn^{-1/2}\right)-n^{1/2}g^0\left(\theta\right)-G_2^0\left(\theta\right)t-\Psi_n\left(\theta\right)\right\|$$

$$\leq\sup_{t\in\mathbb{R}^q:\|t\|<K}\left\|\Psi_n\left(\theta_1,\theta_2+tn^{-1/2}\right)-\Psi_n\left(\theta\right)\right\|+K\,o\left(1\right)=o_p\left(1\right),$$

and from Assumptions M(i) and W we have

$$\sup_{t\in J_n}\left\|W_n\left(\theta_1,\theta_2+tn^{-1/2}\right)-W^0\left(\theta\right)\right\|=o_p\left(1\right).$$

It follows that

$$\sup_{t\in J_n}\left|M_n\left(\theta_1,\theta_2+tn^{-1/2}\right)-\left\|I\left(\Psi_n\left(\theta\right)+n^{1/2}g^0\left(\theta\right)+G_2^0\left(\theta\right)t\right)\right\|^2_{W^0(\theta)}\right|=o_p\left(1\right),$$

which implies

$$\inf_{t\in J_n}M_n\left(\theta_1,\theta_2+tn^{-1/2}\right)=\inf_{t\in J_n}\left\|I\left(\Psi_n\left(\theta\right)+n^{1/2}g^0\left(\theta\right)+G_2^0\left(\theta\right)t\right)\right\|^2_{W^0(\theta)}$$

w.p.a.1.

For any $\theta\in\Theta_I^0$, Assumption G(ii) implies that the function

$$\left\|I\left(\Psi_n\left(\theta\right)+n^{1/2}g^0\left(\theta\right)+G_2^0\left(\theta\right)t\right)\right\|^2_{W^0(\theta)}$$

is continuous and strictly convex in $t\in\mathbb{R}^q$ for every finite $n\in\mathbb{N}$ and every realization of $\Psi_n(\theta)+n^{1/2}g^0(\theta)$. Hence, it achieves a unique minimum inside $J:=\{t\in\mathbb{R}^q:\|t\|<K\}$. By Assumption M(i), for any $n\in\mathbb{N}$ finite but sufficiently large we have $J=J_n$ and therefore, by the CMT, we have

$$\begin{aligned}M_n\left(\theta_1,\theta_2+\hat{t}_nn^{-1/2}\right)&=\inf_{t\in J_n}\left\|I\left(\Psi_n\left(\theta\right)+n^{1/2}g^0\left(\theta\right)+G_2^0\left(\theta\right)t\right)\right\|^2_{W^0(\theta)}+o_p\left(1\right)\\&=\inf_{t\in J}\left\|I\left(\Psi_n\left(\theta\right)+n^{1/2}g^0\left(\theta\right)+G_2^0\left(\theta\right)t\right)\right\|^2_{W^0(\theta)}+o_p\left(1\right)\to_d\inf_{t\in J}\tilde{M}\left(\theta,t\right)\end{aligned}$$

Since $\tilde{t}$ is a tight random variable by Lemma 8, for a choice of K large enough, we have $\inf_{t\in J}\tilde{M}(\theta,t)=\inf_{t\in\mathbb{R}^q}\tilde{M}(\theta,t)$ with probability arbitrarily close to one and the result follows.

Part (b)

Let $F \subset \mathbb{R}^q$ be a closed set. Using arguments similiar to the ones used in Part (a), for every $\theta_1 \in \Theta_{1I}^0$ we have

$$\inf_{t \in F \cap J_n} M_n\left(\theta_1, \theta_2 + n^{-1/2}t\right) - M_n\left(\theta_1, \hat{\theta}_{2n}\right) \rightarrow_d \inf_{t \in F \cap J} \tilde{M}\left(\theta, t\right) - \inf_{t \in \mathbb{R}^q} \tilde{M}\left(\theta, t\right)$$

and therefore

$$\limsup_{n \to \infty} \mathbb{P}^0\left\{\inf_{t \in F \cap J_n} M_n\left(\theta_1, \theta_2 + n^{-1/2}t\right) \leq M_n\left(\theta_1, \hat{\theta}_{2n}\right)\right\} \leq \mathbb{P}^0\left\{\inf_{t \in F \cap J} \tilde{M}\left(\theta, t\right) \leq \inf_{t \in \mathbb{R}^q} \tilde{M}\left(\theta, t\right)\right\},$$

by the PT.

By Assumption M(i), for any $n \in \mathbb{N}$ finite but sufficiently large we have $J \subset J_n$. Hence,

$$\begin{aligned}\limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J\right\} &\leq \limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J_n\right\} \\ &\leq \limsup_{n \to \infty} \mathbb{P}^0\left\{\inf_{t \in F \cap J_n} M_n\left(\theta_1, \theta_2 + n^{-1/2}t\right) \leq M_n\left(\theta_1, \hat{\theta}_{2n}\right)\right\} \\ &\leq \mathbb{P}^0\left\{\inf_{t \in F \cap J} \tilde{M}\left(\theta, t\right) \leq \inf_{t \in \mathbb{R}^q} \tilde{M}\left(\theta, t\right)\right\} \leq \mathbb{P}^0\left\{\tilde{t} \in F \cap J\right\},\end{aligned}$$

where the last inequality follows by the strict convexity of $\tilde{M}\left(\theta, t\right)$ with respect to t.

For any closed set $F \subset \mathbb{R}^q$ we have

$$\limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F\right\} \leq \limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J\right\} + \limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J^c\right\},$$

where J^c denotes the complement of J. By the previous result, we have

$$\limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J\right\} \leq \mathbb{P}^0\left\{\tilde{t} \in F \cap J\right\}$$

and therefore

$$\limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F\right\} \leq \mathbb{P}^0\left\{\tilde{t} \in F\right\} + \limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \notin J\right\}.$$

Because $\hat{t}_n = O_p\left(1\right)$, $\limsup_{n \to \infty} \mathbb{P}^0\left\{\hat{t}_n \in F \cap J^c\right\}$ can be made arbitrarily small by the choice of K . The result will then follow by the PT. ■

Proof of Theorem 4. For any fixed $\theta \in \Theta_{1I}^0$, define $C^0(\theta) \subset C$ as the set of selection vectors that satisfy: $c_k = 0$ for all $k \in \{d+1, ...d+s\}$ such that $g_k^0(\theta) < 0$. The set $C^0(\theta)$ depends on the measure $\mathbb{P}^0$ and on $\theta \in \Theta$ through $g^0(\theta)$ and is generally unknown, but not random. We can also see that $c^0(\theta)$ uniquely maximizes $|c|$ in $C^0(\theta)$.

By assumption, $\theta \in int(\Theta)$ and $\|u\| < \infty$, and we have $\theta + n^{-1/2}u_n \in \Theta$ for $n \in \mathbb{N}$ finite but sufficiently large under Assumption M(i). Hence, $\hat{g}_n\left(\theta + n^{-1/2}u_n\right)$ and $W_n\left(\theta + n^{-1/2}u_n\right)$ are well-defined for $n \in \mathbb{N}$ finite but sufficiently large.

By Assumptions M(iv), G(i) ant M(ii), respectively, we have

$$\begin{aligned} \hat{g}_n^k\left(\theta + n^{-1/2}u_n\right) &= \hat{g}_n^k(\theta) + g_k^0\left(\theta + n^{-1/2}u_n\right) - g_k^0(\theta) + o_p\left(n^{-1/2}\right) \\ &= \hat{g}_n^k(\theta) + G^0(\theta)\, n^{-1/2}u_n + o_p\left(n^{-1/2}\right) \\ &= g_k^0(\theta) + o_p(1). \end{aligned} \quad (5)$$

By Assumption W, we have

$$\begin{aligned} \left\|W_n\left(\theta + n^{-1/2}u_n\right) - W^0\left(\theta + n^{-1/2}u_n\right)\right\| &\leq \\ \sup_{\theta \in \Theta}\left\|W_n(\theta) - W^0(\theta)\right\| + o_p(1) &= o_p(1). \end{aligned} \quad (6)$$

For any fixed selection vector $c \in C$, the function $(c \cdot I(-x))' A (c \cdot I(-x))$ is continuous on $\mathbb{R}^{d+s} \times \mathbb{R}^{(d+s)\times(d+s)}$. Hence, from (5), (6), and the CMT, we have

$$\left\|c \cdot I\left(-\hat{g}_n\left(\theta + n^{-1/2}u_n\right)\right)\right\|^2_{W_n\left(\theta + n^{-1/2}u_n\right)} = \left\|c \cdot I\left(-g^0(\theta)\right)\right\|^2_{W^0(\theta)} + o_p(1)$$

and therefore

$$\begin{aligned} & n^{-1}MMSC\left(\theta + n^{-1/2}u_n, c\right) = \\ &\quad = \left\|c \cdot I\left(-\hat{g}_n\left(\theta + n^{-1/2}u_n\right)\right)\right\|^2_{W_n\left(\theta + n^{-1/2}u_n\right)} - n^{-1}h(|c|)\,\kappa_n \\ &\quad = \left\|c \cdot I\left(-g^0(\theta)\right)\right\|^2_{W^0(\theta)} + o_p(1), \end{aligned}$$

by Assumption MMSC.

For any $\theta \in \Theta_I^0$, if $c \notin C^0(\theta)$, c is such that $c_k = 1$ for some $k \in \{d+1, ..., d+s\}$ for which $g_k^0(\theta) < 0$, then, by Assumptions ID and W, we have

$$\left\|c \cdot I\left(-g^0(\theta)\right)\right\|^2_{W^0(\theta)} > 0$$

and therefore $MMSC\left(\theta+n^{-1/2}h_n,c\right)\to_p+\infty$.

For any $\theta\in\Theta_I^0$, if $c\in C^0(\theta)$, we have

$$\begin{aligned} 0 &\leq n\times\left\|c\cdot I\left(-\hat{g}_n\left(\theta+n^{-1/2}u_n\right)\right)\right\|^2_{W_n\left(\theta+n^{-1/2}u_n\right)} \\ &= \left\|c\cdot I\left(-n^{1/2}\hat{g}_n^k(\theta)+O_p(1)\right)\right\|^2_{W_n\left(\theta+n^{-1/2}u_n\right)}=O_p(1), \end{aligned}$$

where the first equality holds by Assumptions M(iv) and G(i) and the convergence result holds by Assumption M(iii) and the CMT. In this case, for any $\theta\in\Theta_I^0$ and $c\in C^0(\theta)$ we have

$$n^{-1}MMSC\left(\theta+n^{-1/2}u_n,c\right)=O_p\left(n^{-1}\right)-h\left(|c|\right)o(1)=O_p(1)$$

by Assumption MMSC.

By Assumption MMSC, if c_1,c_2 belong to $C^0(\theta)$ and $|c_1|<|c_2|$ we have $\left(h(c_1)-h(c_2)\right)\kappa_n\to-\infty$ and therefore $MMSC(\theta+n^{-1/2}u_n,c_1)>MMSC(\theta+n^{-1/2}u_n,c_2)$ w.p.a.1 for any $\theta\in\Theta_I^0$ such that $\theta\in int(\Theta)$.

Hence, for any $\theta\in\Theta_I^0$ such that $\theta\in int(\Theta)$, $\hat{c}_n\left(\theta+n^{-1/2}u_n\right)$ maximizes $|c|$ over $c\in C^0(\theta)$ w.p.a.1, which implies the result since $c^0(\theta)$ uniquely maximizes $|c|$ in $C^0(\theta)$ and $C^0(\theta)$ is a finite set. ■

Proof of Theorem 5. There always exists a sequence $\left\{\left(\bar{\theta}_n,\bar{\gamma}_n\right):n\geq1\right\}\in\Theta\times\Gamma$, such that

$$\liminf_{n\to\infty}\inf_{\mathbb{P}^{(\theta,\gamma)}\in\mathcal{P}:\theta\in\Theta_I^\gamma}\mathbb{P}^{(\theta,\gamma)}\left\{\theta\in CS_n\right\}=\liminf_{n\to\infty}\mathbb{P}^{\left(\bar{\theta}_n,\bar{\gamma}_n\right)}\left\{M_n\left(\bar{\theta}_n\right)\leq v_n^\alpha\left(\bar{\theta}_n\right)\right\}.$$

Also, there always exists a subsequence $\left\{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right):k\geq1\right\}$ of $\left\{\left(\bar{\theta}_n,\bar{\gamma}_n\right):n\geq1\right\}$ for which $\lim_{k\to\infty}\mathbb{P}^{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right)}\left\{\bar{\theta}_{n_k}\in CS_{n_k}\right\}$ exists and

$$\begin{aligned} \liminf_{n\to\infty}\inf_{\mathbb{P}^{(\theta,\gamma)}\in\mathcal{P}:\theta\in\Theta_I^\gamma}\mathbb{P}^{(\theta,\gamma)}\left\{\theta\in CS_n\right\} &= \lim_{k\to\infty}\mathbb{P}^{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right)}\left\{\bar{\theta}_{n_k}\in CS_{n_k}\right\} \\ &= \lim_{k\to\infty}\mathbb{P}^{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right)}\left\{M_n\left(\bar{\theta}_{n_k}\right)\leq v_n^\alpha\left(\bar{\theta}_{n_k}\right)\right\}. \end{aligned}$$

To each point $\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right)\in\Theta\times G$ in $\left\{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right):k\geq1\right\}$ correponds a point $\left(\bar{\theta}_{n_k},n_k^{1/2}\bar{\gamma}_{1,n_k},\bar{\gamma}_{2,n_k},\bar{\gamma}_{3,n_k}\right)\in\Theta\times\Gamma$ because Γ_1 is a cone. Hence, there exists a subsequence $\left\{\left(\bar{\theta}_{n_{j,h}},\bar{\gamma}_{n_{j,h}}\right):j\geq1\right\}$ of $\left\{\left(\bar{\theta}_{n_k},\bar{\gamma}_{n_k}\right):k\geq1\right\}$ for which

$$\left(\bar{\theta}_{n_{j,h}},n_j^{1/2}\bar{\gamma}_{1,n_{j,h}},\bar{\gamma}_{2,n_{j,h}},\bar{\gamma}_{3,n_{j,h}}\right)\to\left(\theta^h,h_1,h_2,h_3\right)$$

for $\left(\theta^h, h_1, h_2, h_3\right) \in cl\left(\Theta\right) \times \left\{\mathbb{R}_-^s \cup \{-\infty\}^s\right\} \times cl\left(\Gamma_2\right) \times cl\left(\Gamma_3\right)$ as $j \to \infty$.

The result then follows by Lemma 7. ■

Lemma 6 *Take any arbitrary sequence* $\left\{\mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \in \mathcal{P} : n \geq 1\right\}$ *for which*

$$\left(\theta_{n,h}, n^{1/2}\gamma_{1,n,h}, \gamma_{2,n,h}, \gamma_{3,n,h}\right) \to \left(\theta^h, h_1, h_2, h_3\right)$$

for $\left(\theta^h, h_1, h_2, h_3\right) \in cl\left(\Theta\right) \times \left\{\mathbb{R}_-^s \cup \{-\infty\}^s\right\} \times cl\left(\Gamma_2\right) \times cl\left(\Gamma_3\right)$, *satisfying (2), (3) and (4). Define* $c^h \in C$ *to be the selection vector for which* $c^h_{d+k} := 1\left\{|h_{1,k}| < \infty\right\}$ *for all* $k \in \{1, ..., s\}$, *where* $h_{1,k}$ *is the* k*-th coordinate of the vector* h_1. *For this arbitrary sequence* $\left\{\mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \in \mathcal{P} : j \geq 1\right\}$ *we have* $\hat{c}_n\left(\theta_{n,h}\right) \to_p^{\left(\theta_{n,h},\gamma_{n,h}\right)} c^h$ *under Assumption MMSC.*

Proof. For any fixed selection vector $c \in C$, the function $\left(c \cdot I\left(-x\right)\right)' A\left(c \cdot I\left(-x\right)\right)$ is continuous on $\mathbb{R}^{d+s} \times \mathbb{R}^{(d+s)\times(d+s)}$. Hence, from (2), (3) and (4) and the CMT we have

$$n \times \left\|c \cdot I\left(-\hat{g}_n\left(\theta_{n,h}\right) \pm \left(0', \gamma'_{1,n,h}\right)'\right)\right\|^2_{W_n\left(\theta_{n,h}\right)} \rightsquigarrow^{\left(\theta_{n,h},\gamma_{n,h}\right)} \left\|c \cdot I\left(N\left(0, \Delta^h\right) - \left(0', h_1\right)'\right)\right\|^2_{W^h}.$$

Define $C^h \subset C$ to be the set of selection vectors that satisfy: $c_{d+k} = 0$ for all $k \in \{1, ...s\}$ such that $h_{1,k} = -\infty$.

If $c \notin C^h$, then $c_{d+k} = 1$ for some $k \in \{1, ..., s\}$ for which $h_{1,k} = -\infty$. In this case, by Assumption MMSC, we have $MMSC\left(\theta_{n,h}, c\right) \to_p^{\left(\theta_{n,h},\gamma_{n,h}\right)} +\infty$.

If $c \in C^h$, for any $\varepsilon > 0$ there exists some constant $K_\varepsilon > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \left\{MMSC\left(\theta_{n,h}, c\right) < K_\varepsilon\right\} < \varepsilon,$$

under Assumption MMSC.

By Assumption MMSC, if c_1, c_2 belong to C^h and $|c_1| < |c_2|$ we have $\left(h\left(c_1\right) - h\left(c_2\right)\right)\kappa_n \to -\infty$. Therefore,

$$\lim_{n\to\infty} \mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \left\{MMSC(\theta, c_1) > MMSC(\theta, c_2)\right\} \geq 1 - \varepsilon.$$

Hence, $\hat{c}_n\left(\theta_{n,h}\right)$ is such that

$$\lim_{n\to\infty} \mathbb{P}^{\left(\theta_{n,h},\gamma_{n,h}\right)} \left\{|c| \leq \left|\hat{c}_n\left(\theta_{n,h}\right)\right|\right\} \geq 1 - \varepsilon.$$

for all $c \in C^h$. Note that the selection vector c^h uniquely maximizes $|c|$ in C^h. Therefore, $\hat{c}_n(\theta_{n,h}) \rightarrow_p^{(\theta_{n,h},\gamma_{n,h})} c^h$. ■

Lemma 7 *Take any arbitrary sequence* $\left\{\mathbb{P}^{(\theta_{n,h},\gamma_{n,h})} \in \mathcal{P} : n \geq 1\right\}$ *for which*

$$\left(\theta_{n,h}, n^{1/2}\gamma_{1,n,h}, \gamma_{2,n,h}, \gamma_{3,n,h}\right) \rightarrow \left(\theta^h, h_1, h_2, h_3\right)$$

for $\left(\theta^h, h_1, h_2, h_3\right) \in cl(\Theta) \times \left\{\mathbb{R}_-^s \cup \{-\infty\}^s\right\} \times cl(\Gamma_2) \times cl(\Gamma_3)$*, satisfying (2), (3) and (4). Define* $c^h \in C$ *to be the selection vector for which* $c_{d+k}^h := 1\{|h_{1,k}| < \infty\}$ *for all* $k \in \{1, ..., s\}$*, where* $h_{1,k}$ *is the* k*-th coordinate of the vector* h_1*. Also, let* $\nu_n^\alpha(\theta_{n,h})$ *be the* $(1-\alpha)$*-quantile of the distribution of*

$$\left\|\hat{c}_n(\theta_{n,h}) \cdot I\left(\Delta_n(\theta_{n,h})^{1/2} Z^*\right)\right\|_{W_n(\theta_{n,h})}^2$$

with $\hat{c}_n(\theta_{n,h})$ *as in Lemma 6. For this arbitrary sequence* $\left\{\mathbb{P}^{(\theta_{n,h},\gamma_{n,h})} \in \mathcal{P} : j \geq 1\right\}$ *we have:*

(a) $\nu_n^\alpha(\theta_{n,h}) \rightarrow_p^{(\theta_{n,h},\gamma_{n,h})} \nu_\infty^{\alpha,h}$*, where* $\nu_\infty^{\alpha,h}$ *is the* $(1-\alpha)$*-quantile of the random variable* $\left\|c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$.

(b)

$$\liminf_{n\rightarrow\infty} \mathbb{P}^{(\theta_{n,h},\gamma_{n,h})}\left\{M_n(\theta_{n,h}) \leq \nu_n^\alpha(\theta_{n,h})\right\} \geq 1-\alpha.$$

Proof. (a)

From (2), (3) and (4), Lemma 6 and the CMT we have

$$\left\|\hat{c}_n(\theta_{n,h}) \cdot I\left(\Delta_n(\theta_{n,h})^{1/2} Z^*\right)\right\|_{W_n(\theta_{n,h})}^2 \rightarrow_p^{(\theta_{n,h},\gamma_{n,h})} \left\|c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2.$$

The distribution of $\left\|c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$ is continuous at all points in $\mathbb{R}^q$. Hence, $\nu_n^\alpha(\theta_{n,h}) \rightarrow_p^{(\theta_{n,h},\gamma_{n,h})} \nu_\infty^{\alpha,h}$, where $\nu_\infty^{\alpha,h}$ is the $(1-\alpha)$-quantile of the distribution $\left\|c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$ by the CMT.

(b)

Assumptions (2), (3) and (4) and the CMT give us

$$M_n(\theta_{n,h}) \rightsquigarrow^{(\theta_{n,h},\gamma_{n,h})} \left\|I\left(\left(\Delta^h\right)^{1/2} Z^* + h_1\right)\right\|_{W^h}^2.$$

From part (a) we have $\nu_n^\alpha(\theta_{n,h}) \to_p^{(\theta_{n,h},\gamma_{n,h})} \nu_\infty^{\alpha,h}$. Since the convergence of $M_n(\theta_{n,h})$ and $\nu_n^\alpha(\theta_{n,h})$ is joint, we have:

$$M_n(\theta_{n,h}) - \nu_n^\alpha(\theta_{n,h}) \rightsquigarrow^{(\theta_{n,h},\gamma_{n,h})} \left\| I\left(\left(\Delta^h\right)^{1/2} Z^* + h_1\right)\right\|_{W^h}^2 - \nu_\infty^{\alpha,h}. \tag{7}$$

Denote the distribution function of $\left\| I\left(\left(\Delta^h\right)^{1/2} Z^* + h_1\right)\right\|_{W^h}^2$ by $J_h(\cdot)$. For any h_1, we have

$$\left\| I\left(\left(\Delta^h\right)^{1/2} Z^* + h_1\right)\right\|_{W^h}^2 \leq \left\| c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$$

with probability one. The distribution $J_h(\cdot)$ is then stochastically dominated by the distribution function of $\left\| c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$ and therefore $J_h\left(\nu_\infty^{\alpha,h}\right) \geq 1-\alpha$.

Hence,

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}^{(\theta_{n,h},\gamma_{n,h})}\{\theta_{n,h} \in CS_n\} &= \liminf_{n\to\infty} \mathbb{P}^{(\theta_{n,h},\gamma_{n,h})}\{M_n(\theta_{n,h}) - \nu_n^\alpha(\theta_{n,h}) \leq 0\} \\ &\geq J_h\left(\nu_\infty^{\alpha,h}\right) \geq 1-\alpha,\end{aligned}$$

where the first inequality follows from (7) and the PT. The last inequality follows from the fact that $J_h(\cdot)$ is stochastically dominated by the distribution function of $\left\| c^h \cdot I\left(\left(\Delta^h\right)^{1/2} Z^*\right)\right\|_{W^h}^2$. ■

Lemma 8 *For any $\varepsilon > 0$ there exists some constant $K_\varepsilon > 0$ for which*

$$\mathbb{P}^0\left\{\left\|\tilde{t}\right\| < K_\varepsilon\right\} < \varepsilon.$$

Proof. For notational simplicity, for any fixed $\theta_1 \in \Theta_{1I}^0$, denote $\theta_2(\theta_1)$ by θ_2 and the pair $(\theta_1, \theta_2(\theta_1))$ simply by θ.

Because the function $\tilde{M}(\theta_1, t)$ is strictly convex in $t \in \mathbb{R}^q$, the random variable $\tilde{t} := Arg\inf_{t\in\mathbb{R}^q} \|c^0(\theta) \cdot I(\Psi(\theta) + G_2(\theta)t)\|_{W^0(\theta)}^2$ is defined uniquely for every realization of $\Psi(\theta)$. For every realization of $\tilde{t}$, the last s coordinates of the vector $c^0(\theta) \cdot I\left(\Psi(\theta) + G_2(\theta)\tilde{t}\right)$ belong to one and only one of the orthants of $\mathbb{R}^s$. Denote this orthant of $\mathbb{R}^s$ by $\pi\left(\tilde{t}\right)$. Because of the strict convexity of the function $\tilde{M}(\theta, t)$ in t, for every realization of $\Psi(\theta)$, we have

$\tilde{M}\left(\theta,\tilde{t}\right)$ equal to the infimum of $\tilde{M}\left(\theta,\tilde{t}\right)$ for $t\in\mathbb{R}^q$ subject to the last s coordinates of the vector $c^0\left(\theta\right)\cdot I\left(\Psi\left(\theta\right)+G_2\left(\theta\right)\tilde{t}\right)$ belonging to $\pi\left(\tilde{t}\right)$.[14]

Denote by $x^{\pi\left(\tilde{t}\right)}$ and $A^{\pi\left(\tilde{t}\right)}$ the remaining vector and matrix after deleting all the coordinates of the vector x and the matrix A for which the last s coordinates of the vector $c^0\left(\theta\right)\cdot I\left(\Psi\left(\theta\right)+G_2\left(\theta\right)\tilde{t}\right)$ are equal to zero. In this case, by applying standard projection arguments, we may write

$$\tilde{t}=-P_G^{\pi\left(\tilde{t}\right)}W^{0,\pi\left(\tilde{t}\right)}\left(\theta\right)^{-1/2}\Psi^{\pi\left(\tilde{t}\right)}\left(\theta\right)$$

where $P_G^{\pi\left(\tilde{t}\right)}:=\left(G_2^{\pi\left(\tilde{t}\right)}\left(\theta\right)'W^{0,\pi\left(\tilde{t}\right)}\left(\theta\right)^{-1}G_2^{\pi\left(\tilde{t}\right)}\left(\theta\right)\right)^{-1}G_2^{\pi\left(\tilde{t}\right)}\left(\theta\right)'W^{0,\pi\left(\tilde{t}\right)}\left(\theta\right)^{-1/2}$.

Hence, for some constant $0<K<\infty$ we have

$$\begin{aligned}\mathbb{P}^0\left\{\left\|\tilde{t}\right\|<K\right\} &= \sum_\pi \mathbb{P}^0\left\{\left\|\tilde{t}\right\|<K|\pi\left(\tilde{t}\right)=\pi\right\}\mathbb{P}^0\left\{\pi\left(\tilde{t}\right)=\pi\right\}\\ &\leq \max_\pi \mathbb{P}^0\left\{\left\|\tilde{t}\right\|<K|\pi\left(\tilde{t}\right)=\pi\right\}\\ &= \max_\pi \mathbb{P}^0\left\{\left\|-P_G^\pi W^{0,\pi}\left(\theta\right)^{-1/2}\Psi^\pi\left(\theta\right)\right\|<K\right\}.\end{aligned}$$

For any π the random variable $P_G^\pi W^{0,\pi}\left(\theta\right)^{-1/2}\Psi^\pi\left(\theta\right)$ is tight. Hence,

$$\max_\pi \mathbb{P}^0\left\{\left\|-P_G^\pi W^{0,\pi}\left(\theta\right)^{-1/2}\Psi^\pi\left(\theta\right)\right\|<K\right\}$$

can be made arbitrarily small by the choice of K. The result then follows. ■

References

[1] Anderson, T. and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *The Annals of Mathematical Statistics,* 20, 46-63.

[2] Andrews, D. (1994), "Empirical Processes Methods in Econometrics", *Handbook of Econometrics,* Vol. IV, eds. R. Engle and D. McFadden, North-Holland.

[14] See Mangasarain (1969, chap 7) for a proof of this result. Heuristically, the result establishes that if a function is strictly convex and the constraints are linear, its constrained infimum and its unconstrained infimum are the same if the constraint is not binding at the solution of the unconstrained problem.

[3] Andrews, D. (1999a), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation", *Econometrica,* 67, 543-564.

[4] Andrews, D. (1999b), "Estimation when a Parameter is on a Boundary", *Econometrica,* 67, 1341-1383.

[5] Andrews, D. (2000), "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space", *Econometrica,* 68, 399-405.

[6] Andrews, D. and Stock, J. (2005), "Inference with Weak Instruments", *2005 Econometric Society World Congress in London, Cowles Foundation Discussion Paper.*

[7] Andrews, D., Berry, S. and Jia, P. (2005), "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location", *Working Paper.*

[8] Andrews, D. and Guggenberger, P. (2006), "The Limit of Finite Sample Size and a Problem with Subsampling", *Working Paper.*

[9] Bahadur, R. and Savage (1956), "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems", *The Annals of Mathematical Statistics,* 27, 1115-1122.

[10] Bajari, P., Hong, H. and Ryan, S. (2004), "Identification and Estimation of Discrete Games of Complete Information", *Working Paper.*

[11] Beran, R. (1997), "Diagnosing Bootstrap Success", *Ann. Inst. Statisti. Math.*, 49, 1-24.

[12] Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, 60, 889-917.

[13] Bickel, P., Gotze, F. nad van Zwet, W. (1997), "Resampling Fewer than n Observations: Gains, Losses and Remedies for Losses", *Statistica Sinica*, 7, 1-31/

[14] Bresnahan, T. and Reiss, P. (1990), "Entry Monopoly Markets", *Review of Economic Studies*, 57, 57-81.

[15] Bresnahan, T. and Reiss, P. (1991), "Empirical Models of Discrete Games", *Journal of Econometrics*, 64, 355-373.

[16] Ciliberto, F. and Tamer, E. (2004), "Market Structure and Multiple Equilibria in Airline Markets", *Working Paper*.

[17] Chernozhukov, V., Hong, H. and Tamer, E., "Inference on Parameter Sets in Econometric Models", *Working Paper*.

[18] Chow, Y. and Teicher, H. (1978), "Probability Theory", Springer-Verlag.

[19] Guggenberger, P and Smith, R. (2006), "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification", *Econometric Theory,* 21, 667-709.

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

[21] Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression", *Econometrica*, 70, 1781-1804.

[22] Kleibergen, F. (2005), "Testing Parameters in GMM Without Assuming that They Are Identified", *Econometrica*, 73, 1103-1123.

[23] Leeb, H. and Potscher, B. (2005), "Model Selection Inference: Facts and Fiction", *Econometric Theory*, 21, 21-59.

[24] Lehman, E. (1986), "Testing Statistical Hypotheses", Springer-Verlag.

[25] Mangasarian, O. (1969), "Nonlinear Programming", McGraw-Hill Inc.

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

[27] McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", *Econometrica*, 57, 995-1026.

[28] Moreira, M. (2003), "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica*, 71, 1027-1048.

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

[30] Pakes, A., Porter, J., Ho, K. and Ishii, J. (2005), "Moment Inequalities and Their Application", *Working Paper.*

[31] Pollard, D. (1979), "General Chi-Square Goodness-of-Fit Tests with Data-Dependent Cells", *Zeitschrift fuer Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50, 317-331.

[32] Reiss, P. (1996), "Empirical Models of Discrete Strategic Choices", *American Economic Review,* 86, 421-426.

[33] Rosen, A. (2005), "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities", *Working Paper.*

[34] Shaikh, A. (2005), "Inference for a Class of Partially Identified Econometric Models", *Working Paper.*

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

[36] van der Vaart, A. (1998), "Asymptotic Statistics", Cambridge University Press.