1 Inference in Partially Identified Nonparametric Instrumental Variables Models (in collaboration with Yevgeniy Kovchegov) (** VERY PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CIRCULATE WITHOUT CHECKING WITH THE AUTHORS ***):

Answering many empirical questions in economics requires the knowledge of the structural relationship between a dependent variable and a variety of explanatory variables some of which might be endogenous. When the form of this structural relationship is parametrically specified instrumental variables methods are typically used to get consistent and asymptotically normal estimators for the finite dimensional vector of parameters, and thus, the structural function of interest. Unfortunately the parametric estimators are not robust to misspecification of the underlying structural relationship. Failure of the robustness of parametric methods raises the question whether it is possible to extend the instrumental variables estimation to non-parametric framework. This question was first studied in Newey and Powell (2003). Thus far, however, the theoretical analysis and empirical implementation of nonparametric instrumental variables methods have not advanced very much. This may have to do with the fact that identification is very hard to attain in these models. In the non-parametric framework we can estimate the conditional expectation of the structural function given the instruments as in the parametric framework, but unlike in the parametric case, knowledge of this conditional expectation is not sufficient for point identification of the structural function in the non-parametric case unless very strong assumptions are made on the joint distribution of the instruments and the explanatory variables. These assumptions restrict the flexibility of the model overall and significantly reduce the value of the non-parametric approach. This paper will devise methods for consistent estimation and asymptotically valid inference for both the set of structural functions that yield the same conditional expectation given the instruments and each element in this set without imposing assumptions that restrict this set to be almost surely singleton. To our knowledge, the only papers that try to do inference in the nonparametric setting studied by Newey & Powell (2003) and Ai & Chen (2003) without requiring point identification are Santos (2007) and Santos (2008). In the context of a non-parametric model with endogenous covariates, Santos (2007) proposes a test statistic for the null hypothesis that at least one element of the identified set satisfies a conjectured restriction. Santos (2008) proposes a consistent bootstrap procedure for the test statistic of Santos (2007). Neither of the Santos papers, however, tries to estimate the identified set, or incorporate the information that may be obtained about the identified set in the hypothesis testing procedure. To the best of our knowledge our paper is the first attempt at this.

In addition to the difficulty of attaining identification in the type of models we are proposing to study, we have to deal with the well-known ill posed inverse problem associated with the solutions of the integral equations that arise in these models. To recover the structural function of interest a certain linear integral operator, namely the conditional expectation operator, has to be inverted. Point identification of the structural function is equivalent to the invertibility of this operator. Even when assumptions are made to guarantee the invertibility of this mapping, these assumptions fail to guarantee that the inverse is continuous which means that small errors in approximating the image of this operator can translate to large errors in the pre-image. This problem is called an ill posed inverse problem. Although we are not going to seek assumptions that guarantee the invertibility of the conditional expectation operator we will still have to deal with the problem that pre-image of this operator may still vary discontinuously. To deal with this problem we can either impose restrictions on the stochastic relationship between the covariates and the instruments (i.e. impose conditions on their joint distribution) or we can try to restrict our search for structural functions that yield the same conditional expectation given the instruments within a fixed subset of all square integrable functions. Our preliminary investigations, the results of which are presented below, suggest that both of these venues may be fruitful. Although these preliminary results are promising we are planning to explore both of these venues further. We expect to be able to exploit recent advances in spectral theory in this respect.

Although the problem we are considering in this paper is nonparametric, for computational feasibility the procedures we propose will be implemented on a parametric sieve ($\{\mathcal{G}_J\}$) which will approximate the nonparametric family of functions we would like to make inference on. We expect that the Hilbert-Schmidt operator that we consider can be represented via a suitable tridiagonal (Jacobi) operator. The tridiagonal operator will in turn have a unique system of orthogonal polynomials with respect to a probability measure as the operator's eigenfunctions. This will produce a better basis for approximation of the underlying nonparametric space of functions than the Hermitian polynomials used by Newey & Powell (2003). Thus we expect to be able to work under weaker assumptions on $F_{X,Z}$ and the original Hilbert-Schmidt operator.

This paper will build on the recent econometrics literature that has developed methods for inference on finite dimensional parameter vectors in partially identified parametric models. More specifically, once we work with a parametric sieve then for each finite dimensional element of the sieve (i.e. for each fixed J) the inference problem becomes inference for finitely many parameters in a partially identified moment equality model, and the tools developed in the recent parametric partial identification literature become immediately applicable. These tools, however, need to be extended for the case where $J \to \infty$ and such extensions are not readily available.

2 Description of the Problem:

In this section notation and some standard assumptions for instrumental variables models will be introduced. Let Y denote the outcome of interest, such as quantity demanded. This outcome is a function of observed variables X plus some unobserved factors, ϵ:

$$Y = g_0(X) + \epsilon.$$

Some of the covariates, like the price of the good in question, are endogenous so that $E(\epsilon|X) \neq 0$. There are, however, instrumental variables, such as price of the same good in a different market, where

$$E(\epsilon|Z) = 0. \tag{1}$$

Our goal is to make inferences about the function g_0. Without any restrictions on this function and on ϵ such inference will not be possible. The minimal assumption we have to impose is that both g_0 and ϵ have finite variance. Under these assumptions $\pi_0(Z) := E[Y|Z] = E[g_0(X)|Z]$ is well defined. In addition, since Y, X and Z are observable, using non-parametric methods we can estimate $\pi_0(Z)$. To make inference about g_0 we are going to use the information contained in π. Towards this goal, we will define a mapping, T, from $\mathcal{L}^2(X)$ into $\mathcal{L}^2(Z)$, where $T(g) = E[g(X)|Z]$ for each $g \in \mathcal{L}^2(X)$. In particular, if we could invert this mapping and if can show that the inverse of this mapping is continuous we could consistently estimate g_0. As is well known, however, (cf. Newey & Powell (2003), Darolles, Florens & Renault (2002) and Florens & Carrasco (2007)) this mapping is not invertible unless strong assumptions are imposed on $F_{X,Z}$, the joint distribution of X and Z. Furthermore, even under the assumptions that guarantee invertibility of this mapping, continuity of the inverse is not guaranteed. This paper will focus on devising inference methods which do not require the mapping T to be invertible.

2.1 Inference when Support of endogenous covariates and the excluded instruments is finite:

Suppose that $X = (S, Z_1)$, where S and Z_1 denote the endogenous and exogenous covariates, respectively. If there are instruments not included in the outcome equation, Z_2, then $Z = (Z_1, Z_2)$. The simplest case to analyze is when S and W_2 are discrete with finite support conditional on Z_1. Conditional on $Z_1 = z_1$ let J denote the number of points in the support of S and K denote the number of points in the support of Z_2. Also let the conditional probabilities of each s_j given (z_1, z_{2k}) be denoted by $P(z_1)_{kj} := P(S = s_j|Z_2 = Z_{2k}, Z_1 = z_1)$, a $K \times J$ matrix. Then

$$\begin{pmatrix} \pi_0(z_1, z_{21}) \\ \vdots \\ \pi_0(z_1, z_{2K}) \end{pmatrix} = \begin{bmatrix} p_{11}(z_1) & \cdots & p_{1J}(z_1) \\ \vdots & \cdots & \vdots \\ p_{K1}(z_1) & \cdots & p_{KJ}(z_1) \end{bmatrix} \begin{pmatrix} g_0(s_1, z_1) \\ \vdots \\ g_0(s_J, z_1) \end{pmatrix}.$$

When $K \geq J$ and $rank(P(z_1)) = J$ for *a.e.* z_1, the $J \times 1$ vector $g_0(S, z_1) = (P^T(z_1)P(z_1))^{-1}\pi_0(z_1, Z_2)$ for *a.e.* z_1, and the inference problem is quite simple. The PI's note that the inference problem remains to be simple even when $K \leq J$, as long as $P(z_1)$ is still full rank for *a.e.* z_1, i.e. as long as $rank(P(z_1)) = K$ for *a.e.* z_1.[1] In this case even though the mapping $P : \mathbb{R}^J \to \mathbb{R}^K$ is not invertible, the mapping $A : \mathbb{R}^J/L_0 \to \mathbb{R}^K$ with $A[g] := Pg$ is where $\mathbb{R}^J/L_0$ denotes the quotient space of $\mathbb{R}^J$ modulo the subspace $L_0 := \{g \in \mathbb{R}^J : Pg = 0\}$. Note that here $[g]$ denotes the equivalence class of g which consists of all the points in $\mathbb{R}^J$ that mapped to the same point in $\mathbb{R}^K$ under the mapping P. The mapping A is one-to-one, onto, linear and invertible. Using these

[1] For simplicity z_1 in $P(z_1)$ will be dropped in the rest of this section, and all the statements in the rest of this section should be taken to hold for almost every z_1.

facts, we can directly show that there exists a neighborhood $\mathcal{N}(\pi_0)$ (this neighborhood is defined relative to the K dimensional Euclidean metric in this case) and a finite constant R (which does not depend on g_0 because of the linearity of the transformation T) such that the equation $Pg = \pi$ has a solution for each $\pi \in \mathcal{N}(\pi_0)$ which satisfies $\|g - g_0\| \leq R\|\pi - \pi_0\|$. Alternatively, we could note that for each $g \in \mathbb{R}^J$ the (Fréchet) derivative of the mapping Pg (with respect t.o g) is P. Consequently, when P is of full rank this derivative maps $\mathbb{R}^J$ onto $\mathbb{R}^K$ for every $g \in \mathbb{R}^J$. Thus, we get the same conclusion by applying the Generalized Inverse Function Theorem, which is stated at the end of this document for ease of reference. We will elaborate on how this conclusion helps our inference problem below.

2.2 Inference without excluded instruments:

This section discusses how the same type of conclusion as in the previous section can be reached when there are no excluded instruments. In this section X and Z could be discrete or continuous. Note that when there are no excluded instruments $\mathcal{L}^2(Z) \subseteq \mathcal{L}^2(X)$. In addition, define

$$\mathcal{B} := \{E[g(X)|Z] : g \in \mathcal{L}^2(X)\}.$$

This set is the image of the mapping T in $\mathcal{L}^2(Z)$, the set of square integrable functions that only depend on Z, which in this case is a subset of the set of square integrable functions that depend on X only. To get the desired conclusion we first argue that $\mathcal{B}$ is a Banach space when $\mathcal{L}^2(Z) \subseteq \mathcal{L}^2(X)$. The fact that it is a vector space follows from $0 \in \mathcal{L}^2(X)$ and linearity of the mapping T. To see that it is complete, let $\{\varphi_n\}_{n=1}^{\infty} \subseteq \mathcal{B}$ be a Cauchy sequence. Since $\mathcal{B} \subseteq \mathcal{L}^2(Z)$, it is also a Cauchy sequence in $\mathcal{L}^2(X)$. Thus, there exists $\varphi \in \mathcal{L}^2(Z)$ with $\varphi_n \xrightarrow{\mathcal{L}^2} \varphi$. To verify that $\mathcal{B}$ is complete we need to verify that $\varphi \in \mathcal{B}$. Note that $\varphi \in \mathcal{L}^2(X)$ because $\mathcal{L}^2(Z) \subseteq \mathcal{L}^2(X)$, and that $E[\varphi(Z)|Z] = \varphi(Z)$ so that $\varphi \in \mathcal{B}$. Thus, $\mathcal{B}$ is a Banach space. Also note that $\pi_0(Z) \in \mathcal{B}$.

Next, we argue that when T is viewed as a mapping from $\mathcal{L}^2(X)$ to $\mathcal{B}$ every point of $\mathcal{L}^2(X)$ is a regular point of T. The Fréchet differential of T at g with increment h, $\delta T(g; h)$, is defined by

$$\lim_{\|h\|_{\mathcal{L}^2} \to 0} \frac{\|T(g + h) - T(g) - \delta T(g; h)\|_{\mathcal{L}^2}}{\|h\|_{\mathcal{L}^2}}.$$

g is called a regular point of T if $\delta T(g; h)$ maps $\mathcal{L}^2(X)$ onto $\mathcal{L}^2(Z)$. In our case, $\delta T(g; h)$ does not depend on g, so each $g \in \mathcal{L}^2(X)$ is a regular point of T if for each $r \in \mathcal{L}^2(Z)$ there exists an $h \in \mathcal{L}^2(X)$ such that $\delta T(g; h) = E[h(X)|Z] = r(Z)$. Note that when T is viewed as a mapping from $\mathcal{L}^2(X)$ to $\mathcal{B}$ every point of $\mathcal{L}^2(X)$ is a regular point of T because for any $g \in \mathcal{L}^2(X)$ the Fréchet differential of T at g in the direction h is $E[h(X)|Z]$. So the Fréchet derivative maps $\mathcal{L}^2(X)$ onto $\mathcal{B}$ by construction. As a result, the Generalized Inverse Function Theorem tells us that there is a ball with radius η (here this is a ball in the $\mathcal{L}^2$-norm topology) centered at π_0 and a constant R (again, this constant will not even depend on g_0) such that for all π in that ball the equation $T(g) = \pi$ has a solution. Moreover, the solution will satisfy

$$\|h - g_0\| \leq R\|\pi - \pi_0\|. \tag{2}$$

To see how this result can be useful for inference suppose we have an estimator, π_n, for π_0 such that $\pi_n \in \mathcal{L}^2(Z)$ and $\pi_n \xrightarrow{\mathcal{L}^2} \pi$ as $n \to \infty$. Then define

$$\begin{aligned}\mathcal{G}_0 &:= \{g \in \mathcal{L}^2(X) : E[g(X)|Z] = \pi_0(Z)\ a.s.\},\\ \mathcal{G}_n &:= \{g \in \mathcal{L}^2(X) : E[g(X)|Z] = \pi_n(Z)\ a.s.\}.\end{aligned}$$

Here $\mathcal{G}_0$ is the set of functions that we would like to estimate, and this is the set of functions that are mapped to π_0 under the transformation T. We don't know π_0, so to make inferences about $\mathcal{G}_0$ we have to use $\mathcal{G}_n$ which is the set of functions that are mapped to π_n under T. Let $\varepsilon \in (0, \eta)$ and n be sufficiently large so that $\|\pi - \pi_n\|_{\mathcal{L}^2} < \varepsilon$. Consider $g_0 \in \mathcal{G}_0$. By (2) there exists $\tilde{g} \in \mathcal{G}_n$ such that $E[\tilde{g}(X)|Z] = \pi_n(Z)$ and $\|g_0 - \tilde{g}\|_{\mathcal{L}^2} \leq R\varepsilon$. Thus,

$$\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \mathcal{G}_n} \|g - g_0\|_{\mathcal{L}^2} \to 0.$$

Using almost identical steps, we can also show that

$$\sup_{g \in \mathcal{G}_n} \inf_{g_0 \in \mathcal{G}_0} \|g - g_0\|_{\mathcal{L}^2} \to 0.$$

These arguments tell us that

$$\max\{\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \mathcal{G}_n} \|g - g_0\|_{\mathcal{L}^2}, \sup_{g \in \mathcal{G}_n} \inf_{g_0 \in \mathcal{G}_0} \|g - g_0\|_{\mathcal{L}^2}\} \to 0.$$

In words, this means that the set $\mathcal{G}_n$ is consistent for $\mathcal{G}_0$ in the $\mathcal{L}^2$-norm based Hausdorff metric for sets.

In this and the previous section we considered estimating g_0 or the set of functions which satisfy the conditional mean restriction (1) by restricting the stochastic relationship between the covariates and the instruments. In particular, our initial efforts suggest that if we can impose conditions on the joint distribution of the covariates and the instruments that will ensure that the Hilbert-Schmidt operator of the mapping T is bounded below then we can show that the set $\mathcal{B}$ is a Banach space. Using that information we can apply the Generalized Inverse Function Theorem and make significant headway towards making inferences about g_0 or $\mathcal{G}_0$. We have identified two basic scenarios where we can do all this. We are currently investigating other possible conditions on $F_{X,Z}$ that will yield results similar in nature.

2.3 Inference without strong restrictions on $F_{X,Z}$:

In this section we will discuss how we can make inferences about the set of g's that satisfy the moment condition using information contained in π_0 without making assumptions on $F_{X,Z}$ beyond those that guarantee existence of some moments of certain functions of X and Z. As is well-known this approach runs into problems if we don't employ some sort of regularization method. To see where the problem lies suppose we try to proceed as in the previous subsections. In particular, let $\mathcal{B}$ be defined as above, and define

$$\begin{aligned} L_0 &:= \{g \in \mathcal{L}^2(X) : Tg = 0 \; a.s.\} \\ A[g] &:= Tg. \end{aligned}$$

As before $[g]$ denotes the equivalence class of functions that are all mapped to the same point (a.s.) under T. Note that each $[g] \in \mathcal{L}^2(X)/L_0$, the quotient space of $\mathcal{L}^2(X)$ modulo its subspace L_0. By Cauchy-Schwarz inequality L_0 is a closed subspace of $\mathcal{L}^2(X)$.

We can define the norm of $[g] \in \mathcal{L}^2(X)/L_0$ by

$$\|[g]\|_q = \inf\{\|g + l\|_{\mathcal{L}^2} : l \in L_0\}.$$

Thus, $\|[g]\|_q$ is the infimum of the norms of all elements in the equivalence class $[g]$. If $l \in L_0$ $\|[l]\|_q = 0$ and $\|[m]\|_q > 0$ if $[m] \neq [l]$ because L_0 is a closed subspace of $\mathcal{L}^2(X)$. If on the other hand, $\|[x - x_0]\|_q =: a$, and $\tilde{x}$ achieves a, $x_0 + \tilde{x} \in [x]$ and $\|\tilde{x} - x_0\| = a$.[2] Similarly, if x^* is the closest point in $[x]$ to x_0 then $x^* - x_0 \in [x - x_0]$ and $\|x^* - x_0\|_{\mathcal{L}^2} = \|[x - x_0]\|_q$.

Since T is is linear so is the mapping A, and thus, A is one to one. By definition, A is onto $\mathcal{B}$. Therefore it is invertible. Using the quotient space norm which was defined above, we could write

$$\|[g] - [g_0]\|_q = \|A^{-1}(A[g] - A[g_0])\|_q \leq \|A^{-1}\|_o \cdot \|A[g] - A[g_0]\|_{\mathcal{L}^2} = \|A^{-1}\|_o \cdot \|\pi - \pi_0\|_{\mathcal{L}^2}, \quad (3)$$

where $A[g] = \pi$ and $\|A^{-1}\|_o = \sup\{\|A^{-1}r\|_q : \|r\|_{\mathcal{L}^2} \leq 1\}$. The problem is that $\|A^{-1}\|_o$ does not have to be finite. Moreover, even if we try verifying

$$\|[g] - [g_0]\|_q \leq \sup\{\|A^{-1}r\|_q : r \in \mathcal{B}\} \cdot \|A[g] - A[g_0]\|_{\mathcal{L}^2}, \quad (4)$$

instead of (3), without further restrictions we cannot guarantee that $\sup\{\|A^{-1}r\|_q : r \in \mathcal{B}\} < \infty$. This is why we have to use a regularization method. The regularization method we are going to employ is to restrict our estimation to a compact subset of $\mathcal{L}^2(X)$. Thus, let $\mathcal{G} \subset \mathcal{L}^2(X)$ be a compact set. Since T is a continuous mapping $T(\mathcal{G}) \subseteq L^2(Z)$ is compact. If we show that $\mathcal{A} := \mathcal{G}/L_0 := \{[g] \in \mathbb{R}^J/L_0 : [g] \cap \mathcal{G} \neq \emptyset\}$ is compact in $\|\cdot\|_q$ topology, we can conclude that $A : \mathcal{A} \to T(\mathcal{G})$ is a one to one, onto and continuous function. Then by Theorem 5.6 on p. 167 of Munkres (1975), "Topology: A First Course", we can conclude that A is a homeomorphism. Then for π and π_0 both in $T(\mathcal{G})$, and $\rho := \pi - \pi_0$,

$$\|[g] - [g_0]\|_q \leq \|\rho\|_{\mathcal{L}^2} \sup_{\alpha \in (0,1)} \|A^{-1}(\pi_0 + \alpha\rho)\|_q \leq \sup\{\|A^{-1}r\|_q : r \in T(\mathcal{G})\} \cdot \|\pi - \pi_0\|_{\mathcal{L}^2}, \quad (5)$$

where we used the Mean Value Inequality (Proposition 2) on p. 176 of Luenberger (1969) and the fact that the Fréchet differential of A^{-1} at π_0 in the direction $\pi_0 + \alpha\rho$ equals $A^{-1}(\pi_0 + \alpha\rho)$ to

[2]Such $\tilde{x}$ exists because $\|\cdot\|_{\mathcal{L}^2}$ is a continuous operation, and the closed ball in $\mathcal{L}^2$ centered at 0 with radius $5a$ intersected with L_0 is a compact set.

write the first inequality. Note that by continuity of A^{-1} and compactness of $T(\mathcal{G})$ we can ensure that $R := \sup\{\|A^{-1}r\|_q : r \in T(\mathcal{G})\} < \infty$.

To see why $\mathcal{A}$ might be compact in $\|\cdot\|_q$ topology, consider, $\{O_\lambda : \lambda \in \Lambda\}$, an open cover for $\mathcal{A}$. Then each $[h] \in \mathcal{A}$ belongs to some O_{λ_h}. Since O_{λ_h} is open in $\|\cdot\|_q$ topology, there exists $\theta_h > 0$ such that the open ball centered at $[h]$ with radius θ_h is contained in O_{λ_h}. On the other hand, by the arguments given immediately after $\|\cdot\|_q$ was defined we know that $\|[h]-[g]\|_q < \theta_h \Rightarrow \exists g^* \in [g]$ such that $\|h - g^*\|_{\mathcal{L}^2} < \theta_h$. Note that $\{B^{\mathcal{L}^2}_{\theta_h}(h) : h \in \mathcal{G}\}$ is an open cover for $\mathcal{G}$, where $B^{\mathcal{L}^2}_{\theta_h}(h)$ denotes the open ball relative to the $\mathcal{L}^2$-norm with radius θ_h. Since $\mathcal{G}$ is compact, there exists $I < \infty$ such that $\mathcal{G} \subseteq \cup_{\{i=1,\ldots,I\}} B^{\mathcal{L}^2}_{\theta_{h_i}}(h_i)$. Then $\mathcal{A} \subseteq \cup_{\{i=1,\ldots,I\}} O_{\lambda_{h_i}}$.

Consider $g_0 \in \mathcal{G} \cap \mathcal{G}_0 \subseteq \text{int}(\mathcal{G})$, then there exists η, such that $B^{\mathcal{L}^2}_\eta(g_0) \subseteq \mathcal{G}$. Since $\|g - g_0\|_{\mathcal{L}^2} < \eta \Rightarrow \|[g] - [g_0]\|_q < \eta$, and $O := \{[g] \cap \mathcal{G} : [g] \in B^q_\eta([g_0])\}$ is an open subset of $\mathcal{A}$. By continuity of A^{-1} pre-image of O under the mapping A^{-1} is an open subset of $T(\mathcal{G})$. As a result, if $\mathcal{G} \cap \mathcal{G}_0 \subseteq \text{int}(\mathcal{G})$ then $T(\mathcal{G}) \cap \mathcal{G}_0 \subseteq \text{int}\big(T(\mathcal{G})\big)$. On the other hand, if π_0 is in the interior of $T(\mathcal{G})$ and $\pi_n \xrightarrow{\mathcal{L}^\in} \pi_0$, then $\pi_n \in T(\mathcal{G})$ for n sufficiently large. Then there exists $[g_n] \in \mathcal{A}$ such that $A[g_n] = \pi_n$. Moreover, by (5)

$$\|[g_n] - [g_0]\|_q \leq R\|\pi_n - \pi_0\|_{\mathcal{L}^2}.$$

Let $\epsilon > 0$, and n be sufficiently large so that $\|\pi_n - \pi_0\|_{\mathcal{L}^2} < \frac{R}{\epsilon}$. Consider $g_0 \in \mathcal{G} \cap \mathcal{G}_0$. We know that the equation $A[g_n] = \pi_n$ has a solution with $\|[g_n] - [g_0]\|_q < \epsilon$. In addition, we know that there exists $g_n^* \in [g_n]$ such that $\|g_n^* - g_0\|_{\mathcal{L}^2} = \|[g_n] - [g_0]\|_q < \epsilon$. Finally, when ϵ is sufficiently small, g_n^* must be an element of $\mathcal{G}$. This arguments tell us that when $\mathcal{G} \cap \mathcal{G}_0 \subseteq \text{int}(\mathcal{G})$,

$$\sup_{g_0 \in \mathcal{G} \cap \mathcal{G}_0} \inf_{g_n \in \mathcal{G} \cap \mathcal{G}_n} \|g_n - g_0\|_{\mathcal{L}^2} \to 0.$$

Replacing g_n with g_0 in the above arguments, we can also conclude that

$$\sup_{g_n \in \mathcal{G} \cap \mathcal{G}_n} \inf_{g_0 \in \mathcal{G} \cap \mathcal{G}_0} \|g_n - g_0\|_{\mathcal{L}^2} \to 0.$$

3 Estimation:

In his section we describe possible strategies of point and interval estimation of $\mathcal{G}_0$ as well as points in $\mathcal{G}_0$. In this section let $\mathcal{G}$ denote either $\mathcal{L}^2(X)$ or a compact subset of it to be able to talk about estimation under the settings considered in previous sections. In addition, let $\mathcal{G}_0$ denote $\mathcal{G}_0 \cap \mathcal{G}$ and Ω be a positive definite non-stochastic matrix. Then if $\mathcal{G}_0 \neq \emptyset$ we have

$$\begin{aligned} &\min_{g \in \mathcal{G}} E\big[\{Y - E[g(X)|Z]\}^T \Omega \{Y - E[g(X)|Z]\}\big] = 0 \\ &\text{argmin}_{g \in \mathcal{G}} E\big[\{Y - E[g(X)|Z]\}^T \Omega \{Y - E[g(X)|Z]\}\big] = \mathcal{G}_0. \end{aligned} \tag{6}$$

Moreover, if we have an estimator π_n for π_0 such that $\|\pi_n - \pi_0\|_{\mathcal{L}^2} \to 0$, we can estimate $\mathcal{G}_0$ by

$$\hat{\mathcal{G}}_n := \text{arginf}_{g \in \mathcal{G}} \sum_{i=1}^{n} \{E[g|Z_i] - \pi_n\}^T \Omega \{E[g|Z_i] - \pi_n\}. \tag{7}$$

Then the arguments in previous sections tell us that $\hat{\mathcal{G}}_n$ is consistent for $\mathcal{G}_0$ in the $\mathcal{L}^2$-norm based Hausdorff metric for sets.

The above optimization problem may be infeasible computationally. To get a computationally feasible estimator we might employ a parametric sieve $\{\mathcal{G}\}_J$ which approximates $\mathcal{G}$ as $J \to \infty$. For example, following Newey & Powell (2003) we can let $w = (x, z_1)$ and consider $g(w)$ of the form $b(w)^T\beta + g_1(w)$ where $b(w)$ and β are $r \times 1$ vectors of known functions and unknown parameters respectively. Let d denote the dimension of w. Supposing that the mean μ_w and variance Σ_w (non-singular) of w exist let $\tilde{w} := \Sigma_w^{-1/2}(w - \mu_w)$. For $\hat{\mu}_w$ and $\hat{\Sigma}_w$ denoting the sample mean and sample variance of w, let $\hat{w} := \hat{\Sigma}_w^{-1/2}(w - \hat{\mu}_w)$. Also let λ be a $d \times 1$ vector of non-negative integers with $|\lambda| := \sum_{l=1}^d \lambda_l$ and $w^\lambda := \Pi_{i=1}^d (w_l)^{\lambda_l}$. To make computation easier we can use a finite dimensional approximation to $\|g_1\|$. Consider a Hermite polynomial approximation to g_1 of the form

$$g_1(w) \approx \textstyle\sum_{j=1}^J \gamma_j p_j(\hat{w}), \quad \text{with} \quad p_j(w) = \exp(-w^T w) w^{\lambda(j)},$$

where $|\lambda(j)|$ is increasing in j. Whenever necessary we can impose the compactness restriction by bounding the norms of the coefficients γ and β. So let $\mathcal{G}_J$ denote the sets that are approximating the compact set $\mathcal{G}$. Let Θ_J denote the set of parameters of functions in $\mathcal{G}_J$. Then we can estimate $\mathcal{G}_J \cap \mathcal{G}_0$ by minimizing

$$Q_n^J(\theta) := \frac{1}{n}\sum_{i=1}^n \{Y_i - \hat{E}[b|Z_i]^T\beta - \sum_{j=1}^J \gamma_j \hat{E}[p_j|Z_i]\}^T \Omega \{Y_i - \hat{E}[b|Z_i]^T\beta - \sum_{j=1}^J \gamma_j \hat{E}[p_j|Z_i]\}$$

with respect to $\theta = (\beta^T, \gamma^T)^T$. If we let $\hat{\Theta}_{Jn}$ the set of θ values that minimize the sample objective function above, then by arguments given in previous sections we can show that $\hat{\mathcal{G}}_{Jn} := \{b^T\hat{\beta} + \sum_j \hat{\gamma}_j p_j : (\hat{\beta}^T, \hat{\gamma}^T)^T \in \hat{\Theta}_{Jn}\}$ is a consistent estimator for $\mathcal{G}_J \cap \mathcal{G}_0$. In addition, the same arguments suggest that by letting J go to infinity at an appropriate rate we can show that

$$\sup_{g_0 \in \hat{\mathcal{G}}_n} \inf_{g_n \in \mathcal{G}_{Jn}} \|g_n - g_0\|_{\mathcal{L}^2} \to 0,$$

and

$$\sup_{g_n \in \mathcal{G}_{Jn}} \inf_{g_0 \in \hat{\mathcal{G}}_n} \|g_n - g_0\|_{\mathcal{L}^2} \to 0.$$

The existing literature on partially identified parametric models can help us construct two types of confidence sets for each fixed J using this criterion function. The first type of confidence set will contain the $\mathcal{G}_J \cap \mathcal{G}_0$ with a pre-specified probability. In contrast the second type of confidence set will have the property that for each element of the set $\mathcal{G}_J \cap \mathcal{G}_0$ the probability that this element falls into the confidence set will be larger than or equal to a pre-specified level. Having an estimator for the identified set is important for both types of confidence sets. For the first type of confidence set the quantiles of the supremum of $a_n Q_n^J(\theta)$ over the identified set, denoted by $\mathcal{C}_n$ following Chernozhukov, Hong & Tamer (2007), is crucial as the the desired confidence set is given by $\{\theta \in \Theta_J : a_n Q_n^J(\theta) \leq \hat{c}\}$, where $\hat{c}$ is an estimator for the relevant quantile of the random variable that $\mathcal{C}_n$ converges in distribution to.[3] For confidence sets of the second type typically the first step is to show that $a_n Q_n^J(\theta)$ converges in distribution to a random variable $\mathcal{C}(\theta)$. Then the confidence set can be taken as $\{\theta \in \Theta_J : a_n Q_n^J(\theta) \leq \tilde{c}\}$, where $\tilde{c}(\theta)$ is a consistent estimator for the relevant quantile of $\mathcal{C}(\theta)$ for each θ. As noted by Chernozhukov, Hong & Tamer (2007) and Chernozhukov & Fernandez-Val (2005) truncating $\tilde{c}(\theta)$ by the supremum of $\tilde{c}(\theta)$ where the supremum is taken over an estimator for the identified set "leads to substantial finite-sample power improvements".

Although the recent literature on partially identified parametric models can help us create confidence sets for $\mathcal{G}_J \cap \mathcal{G}_0$, it provides little help in how to modify these confidence sets as $J \to \infty$ to get in the limit a confidence set for $\mathcal{G} \cap \mathcal{G}_0$. We expect that a significant portion of the work in this project will be devoted to figuring out a good way of doing that.

Finally, above we mentioned approximation of g_1 via Hermite polynomials, but we might be able to get better results by using a different approximation. In particular, since the Hilbert-Schmidt operator that we consider is bounded and self-adjoint we can represent it via a suitable tridiagonal (Jacobi) operator. The tridiagonal operator will in turn have a unique system of orthogonal polynomials with respect to a probability measure as the operator's eigenfunctions. This will produce a better basis for approximation of $g_1(w)$ than the Hermitian polynomials used by Newey & Powell (2003). Thus we will be able to sharpen their result to work under weaker assumptions on $F_{X,Z}$ and the original Hilbert-Schmidt operator. In general, the tridiagonal representation will allow us to use the most recent tools of spectral theory in extending the work of Newey & Powell and the others. See [5] and [9].

4 The Generalized Inverse Function Theorem:

In this section we state the *Generalized Inverse Function Theorem* given on page 240 of Luenberger (1969).

Definition 4.1 *Let T be a continuously Fréchet differentiable transformation from an open set D in a Banach space X into a Banach space Y. If $x_0 \in D$ is such that*[4] *$T'(x_0)$ maps X onto Y, the point x_0 is said to be a* regular point *of the transformation T.*

Theorem 4.1 (Generalized Inverse Function Theorem, Luenberger (1969)) *Let x_0 be a regular point of a transformation T mapping the Banach space X into the Banach space Y. Then there is a neighborhood $N(y_0)$ of the point $y_0 = T(x_0)$ (i.e., a sphere centered at y_0) and a constant K such that the equation $T(x) = y$ has a solution for every $y \in N(y_0)$, and the solution satisfies $||x - x_0|| \leq K||y - y_0||$.*

[3] If α is the desired coverage probability of the confidence set then the relevant quantile is the α-quantile.

[4] T' denotes the Fréchet derivative of T.

Consistency of Plug-In Estimators of Upper Contour and Level Sets*

Neşe Yıldız†

** Under revision. Please do not circulate without author's permission. **

First version: August 16, 2004. This version: February 28, 2009.

Abstract

This paper studies the problem of estimating the set of finite dimensional parameter values defined by a finite number of moment inequality or equality conditions and gives conditions under which the estimator defined by the set of parameter values that satisfy the estimated versions of these conditions is consistent in Hausdorff metric. This paper also suggests extremum estimators that with probability approaching to one agree with the set consisting of parameter values that satisfy the sample versions of the moment conditions.

KEYWORDS: partial identification, moment equalities, moment inequalities.

*I would like to thank Yevgeniy Kovchegov, Greg Brumfield, Paulo Barelli and William Hawkins for valuable discussions.

†Correspondence: University of Rochester, Department of Economics, 231 Harkness Hall, Rochester, NY 14627; Email: nyildiz@mail.rochester.edu; Phone: 585-275-5782; Fax: 585-256 23 09.

1 Introduction

The appeal of estimation methods based on moment conditions to economists is largely due to their intimate link to economic theory. The optimization problems of economic agents facing uncertainty yield moment conditions which can be exploited to make inferences about the parameters of the agents' utility, cost, or production functions. This paper studies the problem of estimating the set of finite dimensional parameter values defined by a finite number of moment inequality or equality conditions and gives conditions under which the estimator defined by the set of parameter values that satisfy the estimated versions of these conditions is consistent in Hausdorff metric. If the set of parameter values that satisfy the population moment conditions is not empty, then for large sample sizes the estimator sets based on the sample versions of the moment conditions will not be empty either. When the sample size is small, however, these estimated sets may be empty. To deal with this problem, I also propose alternative estimators which are non-empty even for small sample sizes and with probability approaching to one agree with the set of parameters satisfying the sample versions of the moment conditions. These alternative estimators are sets consisting of the minima of a certain criterion function.

Developing methods for making inferences in the context of partially identified econometric models, that is models in which restrictions imposed on the model do not uniquely determine the parameters of interest, but contain useful information about the values these parameters may take, is an active area of research. Recently, Horowitz and Manski (2000) devised confidence intervals for the identified set of univariate parameters. Imbens and Manski (2004) constructed confidence intervals for the univariate parameter itself, rather than for the entire identified set. In the context of interval data Manski and Tamer (2002) proposed several extremum estimators for multidimensional sets of identified regression parameters and provided conditions for (Hausdorff) consistency of these estimators. Chernozhukov, Hong and Tamer (2002, 2007) were the first to develop confidence intervals for a general class of partially identified models. Romano and Shaikh (2006 a, b) constructed confidence intervals for the identified set and for individual parameters in the identified set, respectively, by iterating the procedure presented in Chernozhukov, Hong and Tamer (2002 and 2007). Imbens and Manski (2004), Romano and Shaikh (2006 a,b) and Chernozhukov, Hong and Tamer (2007) also investigate the robustness of these confidence sets in the underlying probability measure. In the context of economic models of entry Andrews, Berry and Jia (2004) studies estimation and inference problems for profit function parameters that satisfy inequality constraints representing necessary conditions for Nash equilibrium. Rosen (2006) presents a different, computationally simple method of constructing confidence sets for parameters in models characterized by a finite number of inequalities. For moment inequality models, Pakes, Porter, Ho and Ishii (2006) suggests a specification test and construct confidence sets for parameter values on the boundary of the identified set in moment inequality models. Beresteanu and Molinari (2006) provides inference methods for models where the identified set can be written as the Aumann expectation of a set valued random variable. Using tools of optimal mass transportation theory Galichon and Henry (2006a) demonstrates how a one sided Kolmogorov-Smirnov test statistic could be used to make inferences about parameters of interest in certain partially identified models. Building on the results of this paper Galichon and Henry (2006b) develops a method of constructing confidence regions in general partially identified models. Their method is based on projecting a large deviation region for multivariate quantile function that generates the data into a large deviation region for the identified set. Canay (2007) studies the problem of inference on the parameters that compose the identified set for moment inequality models and shows that inference based on the empirical likelihood ratio statistic has certain optimality properties. Bugni (2007) proposes a new bootstrap based method for the inference problem for the identified set as well as for each parameter in the identified set. For moment inequality models Andrews and Soares (2007) construct confidence sets for each parameter in the identified set using a new method called generalized moment selection (GMS) which utilizes information about the slackness of the sample moment conditions to infer which population moment conditions are binding. They also investigate asymptotic power of several inference procedures and show that GMS tests dominate subsampling, m out of n bootstrap, and plug-in asymptotic tests in this sense.

When the identified set has multiple disconnected parts with strictly positive distance between the parts, and the identified set is estimated by the collection of minima of a sample criterion function, for finite sample sizes it is possible that the criterion function will attain its minimum in neighborhoods of only a strict subset of these disconnected parts, never picking up neighborhoods of all the parts at once. To deal with this problem Manski and Tamer (2002), Chernozhukov, Hong and Tamer (2002), parts of Chernozhukov, Hong and Tamer (2007) and Bugni (2008) introduce some extra slackness into the objective function or the constraints themselves; this extra slackness or "tolerance" goes to zero as the sample size grows. In certain special cases, constructing consistent estimators without relying on such a "tolerance" parameter is possible. The "degeneracy" condition in Chernozhukov, Hong and Tamer (2007), which is a high level condition in the sense that it is not a condition on the primitives of the model, describes such models.[1]

This paper imposes restrictions on the moment functions and the parameter set which allow the researcher to construct consistent estimators that do not require any "tolerance" parameter. The estimator proposed here for the model comprised of inequality constraints only is very closely linked with the one proposed in Andrews, Berry and Jia (2004). The distinction is in the assumptions imposed on the model. In particular, Andrews, Berry and Jia (2004) devises a consistent estimator for a set of parameter values, Θ_+, over which a population criterion function is minimized so that their estimator consists of parameter

[1]For moment inequality models condition (4.6) in Chernozhukov, Hong and Tamer (2007) is a sufficient condition for "degeneracy". Under the conditions imposed on the moment functions in this paper their condition (4.6) holds.

values that minimize the sample version of the criterion function. To show that this estimator is consistent they assume either that Θ_+ is singleton or that the closure of the interior of Θ_+ is the same as Θ_+. This last condition rules out equality constraints. In contrast, I show the consistency of almost the same estimator by imposing a rank condition on the derivative matrix of the underlying moment conditions. In addition to allowing me to consider inequality as well as equality constraints, this condition can be extended to models of non-parametric regression inequalities as in Yildiz (2008).

The rest of this paper is organized as follows. Section 2 describes the problem. Sections 3 through 5 discuss models where the number of moment conditions does not exceed the dimension of the parameter space. Section 3 gives an estimator and shows its consistency for models comprised of inequality constraints only. Section 4 does the same thing for models consisting of equality conditions only. Section 5 studies models where both types of constraints are available. Section 6 studies the model characterized by inequality constraints only for the case where the number of inequality constraints exceeds the dimension of the parameter space. Section 7 discusses examples. Section 8 concludes. The main mathematical tools employed are described in the Appendix.

2 Description of the Problem:

Let $\Theta \subseteq \mathbb{R}^I$ denote the parameter set. Let $\mathbf{0}$ and $\mathbf{1}$ denote a vector of zeros and ones, respectively, with the size of these vectors inferred from the context. In addition, $B_r(a)$ denotes the open ball with radius r centered at a. For each set A, let $\overline{A}$ denotes the closure of A. Also for sets A, B the Hausdorff distance between them is defined as

$$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\}. \tag{1}$$

Suppose $\{X_i\}_{i=1}^n$ is an i.i.d. sequence of random variables defined on a complete and inner regular probability space $(\Omega, \mathcal{F}, P)$.[2] Let $\mathcal{X}$ and F_X denote the common support and law of X_i. For each $\theta \in \Theta$, let $g(\theta) := E_X \tilde{g}(X, \theta)$ and $\varphi(\theta) := E_X \tilde{\varphi}(X, \theta)$, where $\tilde{g}, \tilde{\varphi}$ are known up to the parameter vector θ, and the images of $\tilde{g}, \tilde{\varphi}$ are a subsets of $\mathbb{R}^M, \mathbb{R}^S$, respectively. I, M and S are all finite. The object of interest will be $\Theta_0 := \{\theta \in \Theta : g(\theta) \geq \mathbf{0}, \varphi(\theta) = \mathbf{0}\}$. I will assume that we have n observations on X and for each value of θ estimators, $\hat{g}_n(\theta), \hat{\varphi}_n(\theta)$ are available for $g(\theta)$ and $\varphi(\theta)$. In this paper, $\hat{g}_n(\theta) = \frac{1}{n}\sum_{j=1}^n \tilde{g}(X_j, \theta)$ and $\hat{\varphi}_n(\theta) = \frac{1}{n}\sum_{j=1}^n \tilde{\varphi}(X_j, \theta)$, but obviously these could be replaced by other estimators that converge to the corresponding population moments uniformly almost surely. The proposed estimator for the set Θ_0 is then $\hat{\Theta}_n := \{\theta \in \Theta : \hat{g}_n(\theta) \geq \mathbf{0}, \hat{\varphi}_n(\theta) = \mathbf{0}\}$. We will show that under our assumptions,

$$\sup_{\theta \in \hat{\Theta}_n} \inf_{\theta_0 \in \Theta_0} ||\hat{\theta} - \theta|| \xrightarrow{P} 0, \tag{2}$$

and

$$\sup_{\theta_0 \in \Theta_0} \inf_{\theta \in \hat{\Theta}_n} ||\theta_0 - \theta|| \xrightarrow{P} 0. \tag{3}$$

The following assumptions will be used in various parts of this paper:

Assumption 2.1 *The parameter set $\Theta \subseteq \mathbb{R}^I$ is compact and convex. In addition, $\Theta_0 \neq \emptyset$.*

Assumption 2.2 *$\tilde{g}$ and $\tilde{\varphi}$ are continuous in θ for almost every x. Moreover, $E\|\tilde{g}(X, \theta)\| < \infty$ and $E\|\tilde{\varphi}(X, \theta)\| < \infty$ $\forall \theta \in \Theta$.*

Under Assumptions 2.1 and 2.2, we have

$$\sup_{\theta \in \Theta} ||\hat{g}_n(\theta) - g(\theta)|| \overset{a.s.}{\rightarrow} 0, \text{ and} \tag{4}$$

$$\sup_{\theta \in \Theta} ||\hat{\varphi}_n(\theta) - \varphi(\theta)|| \overset{a.s.}{\rightarrow} 0, \tag{5}$$

[2]That is, for each $A \in \mathcal{F}$, there exists $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$ with each A_n an open subset of A such that $P(A_n) \uparrow P(A)$.

by Pollard (1984, p. 8, Theorem 2), for example.

3 Inequality Constraints Only:

I will first consider the case where $S = 0$, i.e. the case where we only have M inequality constraints.[3]

In showing the consistency of $\hat{\Theta}_n$ the following sets will be very useful:

$$\begin{aligned} \overline{\Theta}^{\epsilon} &:= \{\theta \in \Theta : g(\theta) \geq -\epsilon \cdot \mathbf{1}\}, \\ \underline{\Theta}^{\epsilon} &:= \{\theta \in \Theta : g(\theta) \geq \epsilon \cdot \mathbf{1}\}, \end{aligned}$$

Note that for $\epsilon > 0$, $\underline{\Theta}^{\epsilon} \subseteq \Theta_0 \subseteq \overline{\Theta}^{\epsilon}$. On the other hand, Assumption (2.2) implies that for each ϵ there exists N_{ϵ} such that for all $n \geq N_{\epsilon}$, $\underline{\Theta}^{\epsilon} \subseteq \hat{\Theta}_n \subseteq \overline{\Theta}^{\epsilon}$ with probability close to 1, so that

$$\sup_{\hat{\theta} \in \hat{\Theta}_n} \inf_{\theta_0 \in \Theta_0} ||\hat{\theta} - \theta_0|| \leq \sup_{\hat{\theta} \in \overline{\Theta}^{\epsilon}} \inf_{\theta_0 \in \Theta_0} ||\hat{\theta} - \theta_0||, \tag{6}$$

and

$$\sup_{\theta_0 \in \Theta_0} \inf_{\hat{\theta} \in \hat{\Theta}_n} ||\hat{\theta} - \theta_0|| \leq \sup_{\theta_0 \in \Theta_0} \inf_{\hat{\theta} \in \underline{\Theta}^{\epsilon}} ||\hat{\theta} - \theta_0||, \tag{7}$$

with probability approaching to 1. Thus, showing that (6) and (7) both converge to 0 as ϵ converges to 0 implies that $\hat{\Theta}_n$ is consistent for Θ_0 in the Hausdorff metric. The following proposition shows that (6) approaches 0 as ϵ decreases to 0:

Proposition 3.1 *If Θ is compact, g is continuous and $\Theta_0 \neq \emptyset$, we have*

$$\sup_{\theta \in \overline{\Theta}^{\epsilon}} \inf_{\theta_0 \in \Theta_0} ||\theta - \theta_0|| \to 0 \ \text{ as } \ \epsilon \to 0.$$

[3]To be precise, moment inequalities can be transformed into moment equalities by specifying $\varphi_m(\theta) := |g_m(\theta)|\{g(\theta) \leq 0\}$. But this specification is not very useful for our purposes because our Jacobian condition will not be applicable to the φ obtained this way.

Proof 3.1 *See Andrews, Berry and Jia (2004), or Manski and Tamer (2002).*

To guarantee consistency of $\hat{\Theta}_n$ we also need to show that (7) converges to 0 as ϵ approaches 0. This direction, however, requires an additional assumption. To state the required assumption we need to define:

Definition 3.1 $h(\theta) := \min\{g_1(\theta), ..., g_M(\theta)\}$, $\Theta^* := \{\theta \in \Theta : h(\theta) = 0\}$.

Assumption 3.1 *The function* $\tilde{g} : \mathcal{X} \times \Theta \to \mathbb{R}^M$ *is continuously differentiable in* θ *for each* $\theta \in \text{int}(\Theta)$ *for almost every* x*, and that* $E\left[\left|\frac{\partial \tilde{g}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty$, $\forall \theta \in \text{int}(\Theta)$, $\forall m$, $\forall j$. *In addition, for each* $\theta^* \in \Theta$ *such that* $h(\theta^*) = 0$, $Dg(\theta^*)$*, the Jacobian of* g *evaluated at* θ^**, has rank* M*.*

Note that the continuous differentiability of $\tilde{g}$ in θ and the absolute integrability of this derivative combined with the Dominated Convergence Theorem imply that $g(\theta)$ is continuously differentiable and its derivative, $Dg(\theta)$, equals $E[D\tilde{g}(X, \theta)]$. In addition, the continuity of the derivative of $\tilde{g}$ with respect to θ for almost every x combined with the compactness of Θ and Theorem 2 on page 8 of Pollard (1984) imply that

$$\sup_{\theta \in \Theta} \|D\hat{g}_n(\theta) - Dg(\theta)\| \stackrel{\text{a.s.}}{\to} 0.$$

While this result is not used in the proof of the next Proposition, it will be useful in Section (5).

Proposition 3.2 *Suppose* $M \leq I$ *and that* $\Theta^* \subseteq \text{int}(\Theta)$*. Then under assumptions (2.1), (2.2) and (3.1), we have*

$$\sup_{\theta_0 \in \Theta_0} \inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \to 0 \text{ as } \epsilon \to 0.$$

Proof 3.2 *Note that if each g_m is continuous, so is $h(\cdot)$. Thus, Θ^* is a closed subset of Θ. Since Θ itself is compact, this means that Θ^* is compact as well. Moreover, since $\Theta^* \subseteq \text{int}(\Theta)$ and Θ^* is compact there exists $\eta > 0$, with η not depending on θ^*, such that $B_\eta(\theta^*) \subseteq \Theta$ for all $\theta^* \in \Theta^*$. On the other hand, the rank of $Dg(\theta^*)$ is the same as the $Dg(\theta^*)Dg(\theta^*)^T$, which is a symmetric positive definite matrix. Thus, rank $\left[Dg(\theta^*)Dg(\theta^*)^T\right] = M$ means $Dg(\theta^*)Dg(\theta^*)^T$ has M strictly positive eigenvalues.*[4] *Moreover, using the Courant Fischer min-max theorem,*[5] *we can write each eigenvalue of $Dg(\theta)Dg(\theta)^T$ as the value function of an optimization problem with a continuous objective function and a continuous constraint correspondence, so that the Theorem of Maximum would imply that these eigenvalues are all continuous functions of θ.*[6] *Since Θ^* is compact, this tells us that $\inf_{\theta^* \in \Theta^*} \lambda_M(\theta^*) =: \underline{\lambda} > 0$, where $\lambda_M(\theta)$ denotes the minimum eigenvalue of $Dg(\theta)Dg(\theta)^T$, and that there exists $\tilde{\rho} > 0$ such that $||\theta - \theta^*|| \leq \tilde{\rho} \Rightarrow \lambda_M(\theta) \geq \frac{1}{2}\underline{\lambda}$, for all $\theta^* \in \Theta^*$. Thus, every element of the compact set $\cup_{\theta^* \in \Theta^*} \overline{B_\rho(\theta^*)}$, where $\rho := \min\left\{\frac{\eta}{2}, \tilde{\rho}\right\}$, is a* regular *point of g.*[7] *In addition, $\cup_{\theta^* \in \Theta^*} \overline{B_\rho(\theta^*)} \subseteq \Theta$. Next, consider*

$$\epsilon_1^* := \frac{1}{2} \inf\{h(\theta) : h(\theta) \geq 0, \theta \in \Theta \setminus \cup_{\theta^* \in \Theta^*} B_\rho(\theta^*)\}. \tag{8}$$

Arguments similar to those given in the proof of proposition (3.1) imply that $\epsilon_1^* > 0$. The arguments up to this point show that whenever $h(\theta) \in [0, \epsilon_1^*]$ for a given θ, then that θ must belong to Θ and be within ρ distance of some $\theta^* \in \Theta^*$, and hence, $Dg(\theta)$ must have rank M. Therefore, by a corollary to the Generalized Inverse Function Theorem[8]*, we know that for each θ_0 satisfying $h(\theta_0) \in [0, \epsilon_1^*]$, there exist $r > 0$ and $K < \infty$, where r and K do not depend on θ_0, but may depend on ϵ_1^*, such that for each $t \in B_r(g(\theta_0))$, the equation $g(\theta) = t$*

[4] For these results, see for example Chapter 11 of Amemiya (1994).

[5] A statement and proof of this theorem is given on pages 115-117 of Bellman (1970).

[6] For a statement of the Theorem of the Maximum, refer to p. 963 of Mas-Colell, Whinston and Green (1995).

[7] θ is a *regular* point of g if $Dg(\theta)$ maps $\mathbb{R}^I$ onto $\mathbb{R}^M$.

[8] See p.240-242 of Luenberger (1969) for a statement and proof of the theorem, and the Appendix of this document for the proof of the corollary.

has a solution. Moreover, the solution satisfies $||\theta - \theta_0|| \leq K||g(\theta) - g(\theta_0)||$.

Let $\delta > 0$*, and consider* $0 < \epsilon < \min\{\frac{r}{\sqrt{M}}, \frac{\delta}{K\sqrt{M}}, \epsilon_1^*, \frac{\eta}{2K\sqrt{M}}\}$*. For any* θ_0 *with* $h(\theta_0) \geq \epsilon$*,* $\theta_0 \in \underline{\Theta}^\epsilon$*, we have* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| = 0$*. Thus, consider* θ_0 *such that* $h(\theta_0) \in [0, \epsilon)$*. Note that if there is no such* θ_0 *we have nothing to prove because in that case* $\Theta_0 = \underline{\Theta}^\epsilon$*. Next, let* $t \in \mathbb{R}^M$ *be defined by* $t_m := g_m(\theta_0) + \epsilon - h(\theta_0)$*. Then*

$$\|t - g(\theta_0)\| = \sqrt{\sum_m (g_m(\theta_0) + \epsilon - h(\theta_0) - g_m(\theta_0))^2} = \sqrt{M}(\epsilon - h(\theta_0)) \leq \sqrt{M}\epsilon < r.$$

Therefore, there is a θ' *such that* $g_m(\theta') = g_m(\theta_0) + \epsilon - h(\theta_0)$*, and*

$$\|\theta_0 - \theta'\| \leq K\sqrt{M}\epsilon < \delta.$$

To argue that $\theta' \in \Theta$*, note that since* $h(\theta_0) \in [0, \epsilon_1^*)$*,* θ_0 *must be within* ρ *distance to some* $\theta_0^* \in \Theta^*$ *and*

$$\|\theta' - \theta_0^*\| \leq \|\theta' - \theta_0\| + \|\theta_0 - \theta_0^*\| \leq K\sqrt{M}\epsilon + \frac{\eta}{2} < \eta.$$

Thus, $\theta_0^* \in B_\eta(\Theta^*) \subseteq \Theta$*. Finally, since* $h(\theta_0) \leq g_m(\theta_0)$*,* $\forall m$*,* $h(\theta') = g_m(\theta_0) + \epsilon - h(\theta_0) \geq \epsilon$*, i.e.* $\theta' \in \underline{\Theta}^\epsilon$*. Thus,* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \leq ||\theta' - \theta_0|| < \delta$*. Since the way* ϵ *was chosen did not depend on* θ_0*, we have* $\inf_{\theta \in \underline{\Theta}^\epsilon} ||\theta - \theta_0|| \leq \delta$ *for each* $\theta_0 \in \Theta_0$*. Since* δ *was chosen arbitrarily these arguments prove the proposition.* ■

Before concluding this section let us note that since $\Theta_0 \neq \emptyset$

$$\Theta_0 = \{\theta \in \Theta : \theta \text{ minimizes } |h(\theta)| 1\{h(\theta) \leq 0\}\}.$$

If $\Theta^* \subseteq \text{int}(\Theta)$, Assumptions (2.2) and (3.1) guarantee that for large n, $\hat{\Theta}_n$ will be not empty with probability approaching to 1. Nevertheless, for small sample sizes, $\hat{\Theta}_n$ could be empty. This problem can be easily fixed, however, by considering the following alternative estimator which equals $\hat{\Theta}_n$ whenever the latter is not empty:

$$\hat{\Theta}_n^a \;=\; \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}(\theta)| 1\{\hat{h}(\theta) \leq 0\}\}.$$

4 Equality Constraints Only:

This section studies the case where $M = 0$ and $S \leq I$, that is the identified set is defined by equality constraints only. Moreover, the number of equality constraints is less than or equal to the dimension of the parameter space. The set we would like to estimate is $\Theta_0 = \{\theta \in \Theta : \varphi(\theta) = \mathbf{0}\}$, and the proposed estimator is $\hat{\Theta}_n := \{\theta \in \Theta : \hat{\varphi}(\theta) = \mathbf{0}\}$. As before, our goal is to show that $d_H(\Theta_0, \hat{\Theta}_n) \xrightarrow{P} 0$. For this purpose define

$$\overline{\Theta}^{\epsilon} := \{\theta \in \Theta : -\epsilon \cdot \mathbf{1} \leq \varphi(\theta) \leq \epsilon \cdot \mathbf{1}\} = \{\theta \in \Theta : \varphi(\theta) \geq -\epsilon \cdot \mathbf{1}, -\varphi(\theta) \geq -\epsilon \cdot \mathbf{1}\}.$$

By Assumption (2.2) $\hat{\Theta}_n \subseteq \overline{\Theta}^{\epsilon}$ as $n \to \infty$ w.p. 1. In addition, if Θ is compact, Θ_0 is non-empty and φ is continuous, Proposition (3.1) implies that

$$\sup_{\theta \in \overline{\Theta}^{\epsilon}} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \to 0 \text{ as } \epsilon \to 0.$$

Thus,

$$\sup_{\theta \in \hat{\Theta}_n} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \xrightarrow{\text{P}} 0.$$

To show the other direction for Hausdorff consistency of the estimator we need to introduce an assumption analogous to Assumption (3.1):

Assumption 4.1 *The function $\tilde{\varphi} : \mathcal{X} \times \Theta \to \mathbb{R}^S$ is continuously differentiable in θ for each $\theta \in \text{int}(\Theta)$ and for almost every x, and that $E\left[\left|\frac{\partial \tilde{\varphi}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty$ $\forall \theta \in \text{int}(\Theta)$, $\forall m$, $\forall j$. In addition, for each $\theta^* \in \Theta$ such that $\varphi(\theta^*) = 0$, $D\varphi(\theta^*)$, the Jacobian of φ evaluated at θ^*,* *has rank S.*

Using arguments similar to those given immediately after Assumption (3.1) we can argue that for each $s = 1, ..., S$ and $i = 1, ..., I$

$$\sup_{\theta\in\Theta}\left|\frac{\partial\hat{\varphi}_{ns}(\theta)}{\partial\theta_i} - \frac{\partial\varphi_s(\theta)}{\partial\theta_i}\right| \xrightarrow{\text{a.s.}} 0.$$

Proposition 4.1 *Suppose $\Theta_0 \subseteq \text{int}(\Theta)$. In addition, suppose that Assumptions (2.1), (2.2) and (4.1) hold. Then $d_H(\Theta_0, \hat{\Theta}_n) \xrightarrow{\text{P}} 0$.*

Proof 4.1 *Since Θ_0 is compact and $\Theta_0 \subseteq \text{int}(\Theta)$ we could show that there exists $\eta > 0$ such that if $\theta' \in B_\eta(\theta_0)$ for some $\theta_0 \in \Theta_0$ then $\theta' \in \Theta$. On the other hand, recall that a real square matrix is positive if and only if determinants associated with all of its upper left submatrices are positive. Let $J_p(\theta)$ denote the determinant of the submatrix consisting of the first p rows and p columns of $D\varphi(\theta)D\varphi(\theta)^T$. Let $\hat{J}_p(\theta)$ be defined in an analogous way with $\hat{\varphi}_n(\theta)$ replacing $\varphi(\theta)$. By assumption (4.1) $J_s(\theta_0) > 0$ $\forall s$ and $\forall\theta_0 \in \Theta_0$. By assumption (4.1) and compactness of Θ_0 $\underline{\underline{\lambda}}^E := \min\{\inf\{J_s(\theta) : \theta \in \Theta_0\} : s = 1, ..., S\} > 0$.*

To show that

$$\sup_{\theta\in\Theta_0}\inf_{\theta'\in\hat{\Theta}_n} ||\theta' - \theta|| \xrightarrow{\text{P}} 0,$$

let $\delta > 0$, and $\epsilon > 0$. By the arguments given just before this proposition and by Egoroff's Theorem[9] *there exists an integer N_1 such that $\forall n \geq N_1$ there exists a set $A_1 \subseteq \mathcal{X}$ with $P(A_1) > 1 - \frac{\delta}{3}$ and an integer N_1 such that $\forall x \in A_1$, $\forall n \geq N_1$ and $\forall s$, we have*

$$\sup_{\theta\in\Theta}|\hat{J}_s(x,\theta) - J_s(x,\theta)| < \frac{\underline{\underline{\lambda}}^E}{4}. \quad (9)$$

This means that for each $x \in A_1$ and for each $n > N_1$ every $\theta_0 \in \Theta_0$ is a regular point of $\hat{\varphi}_n(x,\cdot)$. Therefore, by the Corollary of the Generalized Inverse Function Theorem there exist

[9] This theorem is stated in the Appendix for ease of reference.

$K < \infty$ and $r > 0$ such that for all $\theta_0 \in \Theta_0$ and all $y \in \mathbb{R}^S$ with $y \in B_r(\hat{\varphi}_n(\theta_0))$ the equation $y = \hat{\varphi}_n(\theta)$ has a solution and the solution, $\hat{\theta}_n$, satisfies $\|\hat{\theta}_n - \theta_0\| \leq K\|\hat{\varphi}_n(\hat{\theta}_n) - \hat{\varphi}_n(\theta_0)\|$. Let $\nu \in \left(0, \min\{r, \frac{\eta}{K}, \frac{\epsilon}{K}\}\right)$. Using Assumption (2.2) and Egoroff's Theorem once more we can argue that there is a set $A_2 \subseteq \mathcal{X}$ with $P(A_2) > 1 - \frac{\delta}{3}$ and an integer N_2 such that $\forall x \in A_2$ and $\forall n \geq N_2$, we have

$$\sup_{\theta \in \Theta} \|\hat{\varphi}_n(x, \theta) - \varphi(\theta)\| < \nu. \tag{10}$$

Note that $P(A_1 \cap A_2) > 1 - \frac{2}{3}\delta$. Since $(\Omega, \mathcal{F}, P)$ is inner regular there exists a compact set $A \subseteq A_1 \cap A_2$ with $P(A) > 1 - \delta$ such that $\forall x \in A$ and $\forall n \geq \max\{N_1, N_2\}$ both (9) and (10) hold. Next, consider any $\theta_0 \in \Theta_0$. Let $x \in A$ and $n \geq \max\{N_1, N_2\}$. Then $\|\hat{\varphi}_n(x, \theta_0)\| < r$ (by (10)) and θ_0 is a regular point of $\hat{\varphi}_n(x, \cdot)$. Thus, there exists $\hat{\theta}_n(x)$ with $\hat{\varphi}_n(x, \hat{\theta}_n(x)) = 0$. Moreover, $\|\hat{\theta}_n(x) - \theta_0\| \leq K\|\hat{\varphi}_n(x, \hat{\theta}_n(x)) - \hat{\varphi}_(x, \theta_0)\| < K\nu < \min\{\eta, \epsilon\}$ meaning that $\hat{\theta}_n(x) \in \Theta$ and is less than ϵ distant away from θ_0. Since $\theta_0 \in \Theta_0, \delta > 0$ and $\epsilon > 0$ were chosen arbitrarily and since r, K, N_1 and N_2 do not depend on θ_0 these arguments prove that $\forall \epsilon > 0$

$$P\left(\sup_{\theta_0 \in \Theta_0} \inf_{\hat{\theta}_n \in \hat{\Theta}_n} \|\hat{\theta}_n - \theta_0\| > \epsilon\right) \to 0, \text{ as } n \to \infty. \quad \blacksquare$$

The arguments given in this proof demonstrate that for all sufficiently large n the probability that $\hat{\Theta}_n \neq \emptyset$ will be close to 1. When $\hat{\Theta}_n \neq \emptyset$,

$$\hat{\Theta}_n = \hat{\Theta}_n^a := \{\theta' \in \Theta : \theta' \text{ minimizes } |\hat{h}_n(\theta)| 1\{\hat{h}_n(\theta) \leq 0\}\},$$

where $\hat{h}_n(\theta) = \min\{\hat{\varphi}_{1n}(\theta), ..., \hat{\varphi}_{Sn}(\theta), -\hat{\varphi}_{1n}(\theta), ..., -\hat{\varphi}_{Sn}(\theta)\}$. Since we have shown that $\hat{\Theta}_n$ is consistent for Θ_0 in Hausdorff metric, this means that $\hat{\Theta}_n^a$ is also consistent for Θ_0 in the same metric. On the other hand, since $\hat{\varphi}_n$ is continuous in θ $\hat{\Theta}_n^a$ will be non-empty. This suggests that $\hat{\Theta}_n^a$ may be preferable to $\hat{\Theta}_n$ for small sample sizes.

5 Inequality and Equality Constraints Together:

In this section we turn to the case where the set that needs to be estimated is $\Theta_0 = \{\theta \in \Theta : g(\theta) \geq \mathbf{0},\ \varphi(\theta) = \mathbf{0}\}$, and the proposed estimator is
$\hat{\Theta}_n := \{\theta \in \Theta : \hat{g}_n(\theta) \geq \mathbf{0},\ \hat{\varphi}_n(\theta) = \mathbf{0}\}$. As before, our goal is to show that $d_H(\Theta_0, \hat{\Theta}_n) \xrightarrow{P} 0$. This time we define

$$\overline{\Theta}^{\epsilon} := \{\theta \in \Theta : g(\theta) \geq -\epsilon \cdot \mathbf{1},\ -\epsilon \cdot \mathbf{1} \leq \varphi(\theta) \leq \epsilon \cdot \mathbf{1}\} = \{\theta \in \Theta : h^E(\theta) \geq -\epsilon\},$$

where $h^E(\theta) = \min\{\xi_j(\theta) : j = 1, ..., M+2S\}$, with $\xi_j(\theta) = g_j(\theta)$ for $j = 1, ..., M$, $\xi_{M+j}(\theta) = \varphi_j(\theta)$ and $\xi_{M+S+j}(\theta) = -\varphi_j(\theta)$ for $j = 1, ..., S$. With this definition, it is easy to see that if Θ is compact, $\Theta_0 \neq \emptyset$ and g and φ are continuous, by Proposition (3.1) we have

$$\sup_{\theta \in \overline{\Theta}^{\epsilon}} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \to 0 \text{ as } \epsilon \to 0.$$

Assumption (2.2) then implies that

$$\sup_{\theta \in \hat{\Theta}_n} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \xrightarrow{P} 0.$$

As in the previous section, to show the other direction for Hausdorff consistency of the estimator we need to strengthen Assumption (2.2) and modify Assumption (3.1). Recall that $h(\theta) = \min\{g_m(\theta) : m = 1, ..., M\}$. Let $\tilde{f}(\theta) := \begin{pmatrix} \tilde{g}(\theta) \\ \tilde{\varphi}(\theta) \end{pmatrix}$, $\hat{f}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \tilde{f}(X_i, \theta)$ and $f(\theta) = E_X \tilde{f}(X, \theta)$. In addition, let $\Theta^{**} := \{\theta \in \Theta : h(\theta) = 0, \varphi(\theta) = 0\}$.

Assumption 5.1 *The functions $\tilde{\varphi} : \mathcal{X} \times \Theta \to \mathbb{R}^S$ and $\tilde{g} : \mathcal{X} \times \Theta \to \mathbb{R}^M$ are continuously differentiable in θ, for each $\theta \in \operatorname{int}(\Theta)$ and for almost every x, and that $E\left[\left|\frac{\partial \tilde{\varphi}_s(X,\theta)}{\partial \theta_j}\right|\right] < \infty$ and $E\left[\left|\frac{\partial \tilde{g}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty$, $\forall \theta \in \operatorname{int}(\Theta)$, $\forall s, m\ \forall j$. In addition, for each $\theta^* \in \Theta_0$, $D\varphi(\theta^*)$ has rank S, and for each $\theta^* \in \Theta_0$ such that $h(\theta^*) = 0$, $Df(\theta^*)$ has rank $S + M$.*

Note that Assumption (2.2) implies that $\sup_{\theta\in\Theta}\|\hat{h}_n(\theta)-h(\theta)\| \xrightarrow{a.s} 0$.

Proposition 5.1 *Suppose $\Theta^{**}\subseteq \text{int}(\Theta)$. In addition, suppose that Assumptions (2.1), (2.2) and (5.1) hold. Then $d_H(\Theta_0,\hat{\Theta}_n)\xrightarrow{P} 0$.*

Proof 5.1 *Using Assumption (5.1) and arguments as in the proof of Proposition (3.2) we can show that there is $\rho>0$ such that every element of $\cup_{\theta^*\in\Theta^{**}}\overline{B_\rho(\theta^*)}$ is a regular point of f and $\cup_{\theta^*\in\Theta^*}\overline{B_\rho(\theta^*)}\subseteq\Theta$. Let $\epsilon_1^*:=\frac{1}{2}\inf\{h(\theta):\theta\in\Theta_0\setminus\cup_{\theta^*\in\Theta^*}B_\rho(\theta^*)\}$. Again we can show that if $h(\theta)\in[0,\epsilon_1^*]$ and $\varphi(\theta)=0$ for a given θ then that θ must be a regular point of f belonging to Θ.*

Define $E:=\{\theta\in\Theta:h(\theta)\geq\epsilon_1^*,\varphi(\theta)=\mathbf{0}\}$, $F:=\{\theta\in\Theta:h(\theta)\in[0,\epsilon_1^*],\varphi(\theta)=\mathbf{0}\}$. Note that E and F are compact and $\Theta_0=E\cup F$. Also, using continuity of h, compactness of Θ and the fact that $\Theta_0\subseteq\text{int}(\Theta)$ we can show that there exists $\rho_2>0$ such that $\forall e\in E$ $\|\theta-e\|<\rho_2\Rightarrow h(\theta)\geq\frac{\epsilon_1^*}{2}$ and $\theta\in\Theta$.

Let $J_p(\theta)$ and $\hat{J}_p(\theta)$ be defined as in the proof of Proposition (4.1). Let $J_p^f(\theta)$ and $\hat{J}_p^f(\theta)$ analogously with f and $\hat{f}_n$ replacing φ and $\hat{\varphi}_n$, respectively. Also let $\underline{\underline{\lambda}}:=\min\{\inf\{J_s(\theta):\theta\in\Theta_0\},\inf\{J_l^f(\theta):\theta\in F\}:s=1,...,S,l=1,...,M+S\}$.

Let $0<\delta<1$ and $\epsilon>0$. By Assumption (2.2) and Egoroff's Theorem, there exists a positive integer N_1 and a set $A_1\subseteq\mathcal{X}$ with $P(A_1)>1-\frac{\delta}{2}$ such that $\forall\omega\in A_1$, $\forall s=1,...,S$, $\forall l=1,...,M+S$ and $\forall n\geq N_1$ we have

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\hat{J}_s(\theta)-J_s(\theta)\right| &< \frac{\underline{\underline{\lambda}}}{4},\ \text{and} \\ \sup_{\theta\in\Theta}\left|\hat{J}_l^f(\theta)-J_l^f(\theta)\right| &< \frac{\underline{\underline{\lambda}}}{4}.\end{aligned} \tag{11}$$

This means that for each $\omega\in A_1$ and each $n>N_1$, every $\theta_0\in\Theta_0$ is a regular point of $\hat{\varphi}_n(\omega,\cdot)$ and every $\theta\in F$ is a regular point of $\hat{f}_n(\omega,\cdot)$. Then our Corollary to the Generalized Inverse Function Theorem implies that there exist $r>0$ and $K<\infty$ such that $\forall\omega\in A_1$ and $\forall n>N_1$ the following two conditions hold:

(i) $\forall\theta_0 \in \Theta_0$ *the equation* $\hat{\varphi}_n(\omega,\theta) = y$ *has a solution* $\forall y \in B_r(\hat{\varphi}_n(\omega,\theta_0))$*, and the solution satisfies* $\|\theta - \theta_0\| \leq K\|y\|$*;*

(ii) $\forall\theta_0 \in F$ *the equation* $\hat{f}_n(\omega,\theta) = y$ *has a solution* $\forall y \in B_r(z)$ *where* $z = \hat{f}_n(\omega,\theta_0)$*. Moreover, the solution satisfies* $\|\theta - \theta_0\| \leq K\|y - z\|$*.*

Next, let $\nu \in \left(0, \min\left\{\frac{r}{\sqrt{M+1}}, \frac{\rho}{K\sqrt{M+1}}, \frac{\rho_2}{K}, \frac{\epsilon}{K\sqrt{M+1}}, \frac{\epsilon_1^*}{2}\right\}\right)$*. Using Assumption (2.2) and Egoroff's Theorem once more, we can argue that there exists a positive integer* N_2 *and a set* $A_2 \subseteq \mathcal{X}^\infty$ *with* $P^\infty(A_2) > 1 - \frac{\delta}{2}$ *such that* $\forall\omega \in A_2$*, and* $\forall n \geq N_2$ *we have*

$$\sup_{\theta\in\Theta}\left\|\begin{pmatrix}\hat{h}_n(\omega,\theta)\\ \hat{g}_n(\omega,\theta)\\ \hat{\varphi}_n(\omega,\theta)\end{pmatrix} - \begin{pmatrix}h(\theta)\\ g(\theta)\\ \varphi(\theta)\end{pmatrix}\right\| < \nu. \tag{12}$$

Consider $\theta_0 \in \Theta_0$*. Let* $\omega \in A_1 \cap A_2$ *and* $n > \max\{N_1, N_2\}$*. Suppose* $\theta_0 \in E$*. As in the proof of Proposition (4.1) we can show that there exists* $\hat{\theta}_n(\omega) \in \Theta$ *such that* $\hat{\varphi}_n(\omega, \hat{\theta}_n(\omega)) = 0$*,* $\|\hat{\theta}_n(\omega) - \theta_0\| < \epsilon$ *and* $\|\hat{\theta}_n(\omega) - \theta_0\| < \rho_2$*. This last expression guarantees that* $h(\hat{\theta}_n) \geq \frac{\epsilon_1^*}{2}$*. Moreover, by expression (12),* $\hat{h}_n(\hat{\theta}_n) > 0$ *and hence,* $\hat{\theta}_n \in \hat{\Theta}_n$ *for such* ω *and* n*. Next, suppose that* $\theta_0 \in F$*. Let* $t_m := \hat{g}_{nm}(\theta_0) + h(\theta_0) - \hat{h}_n(\theta_0)$ *for* $m = 1, ..., M$ *and* $t_m = 0$ *for* $m = M+1, ..., M+S$*. Then* $\|t - \hat{f}_n(\omega,\theta_0)\| = \sqrt{M(h(\theta_0) - \hat{h}_n(\theta_0))^2 + \|\hat{\varphi}_n(\theta_0) - \varphi(\theta_0)\|^2} < \sqrt{M+1}\nu < r$*. Note that* $t \geq \mathbf{0}$ *since* $h(\theta_0) \geq 0$ *and since* $\hat{h}_n(\theta_0) \leq \hat{g}_{nm}(\theta_0)$ $\forall m$*. Since* θ_0 *is a regular point of* $\hat{f}_n(\omega,\cdot)$ *there exists* $\hat{\theta}'_n(\omega)$ *satisfying* $\hat{f}_n(\omega, \hat{\theta}'_n(\omega)) = t$ *and* $\|\hat{\theta}'_n(\omega) - \theta_0\| \leq K\|t - \hat{f}_n(\omega,\theta_0)\| < K\sqrt{M+1}\nu < \epsilon$*. Finally,* $\|\hat{\theta}'_n(\omega) - \theta_0\| < \rho \Rightarrow \hat{\theta}'_n(\omega) \in \Theta$*. Since* $\theta_0 \in \Theta_0$ *was chosen arbitrarily, and all these arguments hold regardless of which* $\theta_0 \in \Theta_0$ *is chosen the results indicate that* $\forall n > \max N_1, N_2$*,*

$$P(\sup_{\theta\in\Theta_0}\inf_{\theta'\in\hat{\Theta}_n}\|\theta_0 - \theta'\| < \epsilon) > 1 - \delta. \quad \blacksquare$$

Note that when $\hat{\Theta}_n \neq \emptyset$

$$\hat{\Theta}_n = \hat{\Theta}_n^a := \{\theta \in \Theta : \theta \text{ minimizes } |\hat{H}_n(\theta')|1\{\hat{H}_n(\theta') \leq 0\}\},$$

where $\hat{H}_n(\theta) = \min\{\hat{h}_n(\theta), \hat{\varphi}_{1n}(\theta), ..., \hat{\varphi}_{Sn}(\theta), -\hat{\varphi}_{1n}(\theta), ..., -\hat{\varphi}_{Sn}(\theta)\}$. Since Θ is compact and the objective function is continuous this alternative estimator will be non-empty for each sample size. Since $\hat{\Theta}_n$ will be non-empty with probability approaching to 1 and since the two estimators agree whenever $\hat{\Theta}_n \neq \emptyset$ $\hat{\Theta}_n^a$ will be consistent as well.

6 More Moment Inequalities than Parameters:

Suppose we have $M \geq I$ and $S = 0$. In this case we cannot use the Generalized Inverse Function Theorem because the derivative map will not be onto. Nevertheless, we can try to solve this problem by breaking the set of moment inequalities into subsets where the number of elements in each subset is at most I. Suppose this gives us L subsets, with $M_l \leq I$, denoting the cardinality of the l^{th} subset. The analysis of the previous subsection can be applied to each subset of inequalities provided that the assumptions we made there hold for each subset. In particular, for $l = 1, ..., L$ define $\Theta_0^l := \{\theta \in \Theta : g^l(\theta) \geq 0\}$ and $\hat{\Theta}_n^l := \{\theta \in \Theta : \hat{g}^l(\theta) \geq 0\}$. The analysis in section (3) shows that for each $l = 1, ..., L$, $d_H(\Theta_0^l, \hat{\Theta}_n^l) \xrightarrow{P} 0$. Using this fact along with $\Theta_0 = \cap_{l=1}^L \Theta_0^l$ and $\hat{\Theta}_n = \cap_{l=1}^L \hat{\Theta}_n^l$, one might try to argue that $d_H(\Theta_0, \hat{\Theta}_n) \xrightarrow{P} 0$. We cannot, however, immediately come to this conclusion based on the analysis in section (3); we need to make extra assumptions to guarantee this result. To illustrate this point, consider the case where $L = 2$. Define for $l = 1, 2$, $\overline{\Theta}^{\epsilon,l}$, $\underline{\Theta}^{\epsilon,l}$, and h_l as in section (3) with g^l replacing g in each case. Then with probability approaching to 1, we have

$$\sup_{\hat\theta\in\hat\Theta_n^1\cap\hat\Theta_n^2}\inf_{\theta_0\in\Theta_0^1\cap\Theta_0^2}||\hat\theta-\theta_0|| \leq \sup_{\hat\theta\in\overline{\Theta}^{\epsilon,1}\cap\overline{\Theta}^{\epsilon,2}}\inf_{\theta_0\in\Theta_0^1\cap\Theta_0^2}||\hat\theta-\theta_0||, \quad (13)$$

$$\sup_{\theta_0\in\Theta_0^1\cap\Theta_0^2}\inf_{\hat\theta\in\hat\Theta_n^1\cap\hat\Theta_n^2}||\hat\theta-\theta_0|| \leq \sup_{\theta_0\in\Theta_0^1\cap\Theta_0^2}\inf_{\hat\theta\in\underline{\Theta}^{\epsilon,1}\cap\underline{\Theta}^{\epsilon,2}}||\hat\theta-\theta_0||, \quad (14)$$

for sufficiently large n. Since the proof of Proposition (3.1) did not make any assumptions about the size of M relative to that of I, that proposition is valid for $M \geq I$, and we have

$$\sup_{\hat\theta\in\overline{\Theta}^{\epsilon,1}\cap\overline{\Theta}^{\epsilon,2}}\inf_{\theta_0\in\Theta_0^1\cap\Theta_0^2}||\hat\theta-\theta_0|| \to 0.$$

Unfortunately,

$$\sup_{\theta_0\in\Theta_0^l}\inf_{\hat\theta\in\underline{\Theta}^{\epsilon,l}}||\hat\theta-\theta_0|| \to 0 \text{ for } l=1,2, \not\Rightarrow \sup_{\theta_0\in\Theta_0^1\cap\Theta_0^2}\inf_{\hat\theta\in\underline{\Theta}^{\epsilon,1}\cap\underline{\Theta}^{\epsilon,2}}||\hat\theta-\theta_0|| \to 0.$$

This results from the fact that even if $\Theta_0^1\cap\Theta_0^2$, $\underline{\Theta}^{\epsilon,1}$ and $\underline{\Theta}^{\epsilon,2}$ are all non-empty the intersection of $\underline{\Theta}^{\epsilon,1}$ and $\underline{\Theta}^{\epsilon,2}$ could be empty, which would imply that the expression on the right hand side of (14) is infinite. We can deal with this problem in more specialized models. The following assumption describes the additional condition these specialized models require:

Assumption 6.1 *Let* $h(\theta) := \min\{h^l(\theta) : l = 1,...,L\}$ *and* $\Theta^* = \{\theta \in \Theta : h(\theta) = 0\}$. *Suppose either that* Θ_0 *is singleton or for each* $\theta^* \in \Theta^*$ *there is* j_0*, with* j_0 *possibly dependent on* θ^**, such that* h *is strictly increasing or strictly decreasing in* θ_{j_0} *at* $\theta^* \in \Theta^*$.

Proposition 6.1 *Suppose Assumptions (2.1), (2.2) and (6.1) hold, and that* $\Theta^* \subseteq \text{int}(\Theta)$. *In addition, suppose that the moment functions could be broken into* L *groups as described above such that each group of functions satisfies Assumption (2.1) for* Θ^* *as defined here. Then* $d_H(\hat\Theta_n^a, \Theta_0) \xrightarrow{P} 0$ *where* $\hat\Theta_n^a = \{\theta \in \Theta : \theta \text{ minimizes } |\hat h(\theta)|1\{\hat h(\theta) \leq 0\}\}$.

Proof 6.1 *Note that all of the assumptions except Assumption 2 of Theorem 1 of Andrews,* *Berry and Jia (2004) immediately follow from the assumptions we imposed and our previous analysis. Here we will verify that (i)* $\forall\theta_0 \in \text{int}(\Theta_0)$ $h(\theta_0) > 0$*, and that (ii)* $\Theta_0 = \overline{\text{int}(\Theta_0)}$*.*

To see that (i) holds, suppose towards contradiction there is $\theta_0 \in \text{int}(\Theta_0)$ *with* $h(\theta_0) = 0$*. Then* $h^{l_0}(\theta_0)$ *must be 0 for some* $l_0 \in \{1, ..., L\}$*. Then by Assumption (3.1) and our Inverse Function Theorem, for all sufficiently small* $\epsilon > 0$ *there is* $\underline{\theta}_\epsilon$ *with* $h(\underline{\theta}_\epsilon) \leq h^{l_0}(\underline{\theta}_\epsilon) \leq -\epsilon$ *and* $\|\underline{\theta}_\epsilon - \theta_0\| \leq K\epsilon$ *for some finite* K*. This shows that* θ_0 *cannot be in the interior of* Θ_0*.*

To see why (ii) holds, recall that Θ_0 *is a closed set, so that* $\Theta_0 = \overline{\Theta_0} \supseteq \overline{\text{int}(\Theta_0)}$ *by definition of the closure of a set. Next, let* $\theta_0 \in \Theta_0$*. If* $\theta_0 \in \text{int}(\Theta_0)$ *then* $\theta_0 \in \overline{\text{int}(\Theta_0)}$ *as well. If* $\theta_0 \notin \text{int}(\Theta_0)$ *then that means for all* $\delta > 0$ *there exists* $\theta'(\delta) \in B_\delta(\theta_0)$ *with* $h(\theta'(\delta)) < 0$*. Since* h *is continuous, this implies that* $h(\theta_0) = 0$*. On the other hand, by Assumption (6.1) there is* j_0 *such that* h *is either increasing or decreasing in* θ_{j_0} *at* θ_0*. Let* $\theta_{tj} = \theta_{0j}$ *for* $j \neq j_0$*,* $\theta_{tj_0} = \theta_{0j_0} + \frac{1}{t}$ *if* h *is increasing in* θ_{j0} *and* $\theta_{tj_0} = \theta_{0j_0} - \frac{1}{t}$ *if* h *is decreasing in* θ_{j0} *at* θ_0*. Since* $\theta_0 \in \Theta^* \subseteq \text{int}(\Theta)$ *for some sufficiently large* n_1 $\{\theta_t\}_{t=n_1}^{\infty} \subseteq \text{int}(\Theta_0)$*. Moreover* $\theta_t \to \theta_0$ *as* $t \to \infty$*. Thus,* $\theta_0 \in \overline{\text{int}(\Theta_0)}$*, and* $\Theta_0 \subseteq \overline{\text{int}(\Theta_0)}$*, and the Proposition follows from Theorem 1 of Andrews, Berry and Jia (2004).* ■

7 Examples:

In this section we review two economic models where the parameters of the model are partially identified. We study the meaning of the Jacobian condition of this paper in the context of these models. These models are also discussed in Chernozhukov, Hong and Tamer (2007).

7.1 Bracketed Data:

Suppose random variable Y is unobserved, but is known to be bounded above and below by observed random variables Y_2 and Y_1 respectively. Let $\mu_j := E(Y_j)$ for $j = 1, 2$, and

$\theta := E(Y)$. Suppose that $\mu_2 > \mu_1$.[10] In this example,

$$g(\theta) = (\mu_2 - \theta, \theta - \mu_1)^T,$$

and $\Theta_0 = [\mu_1, \mu_2]$. The number of moment conditions is larger than the dimension of θ. Note, however, that $h(\theta) = \begin{cases} \theta - \mu_1 & \text{if } \theta \leq \frac{\mu_1+\mu_2}{2}, \\ \mu_2 - \theta & \text{if } \theta \geq \frac{\mu_1+\mu_2}{2}. \end{cases}$. Moreover, $\Theta^* = \{\mu_1, \mu_2\}$, and h is differentiable on Θ^*, with $h'(\mu_1) = 1$ and $h'(\mu_2) = -1$. Therefore, we can estimate Θ_0 consistently by minimizing $\hat{Q}^a(\theta) = |\hat{h}(\theta)|1\{\hat{h}(\theta) \leq 0\}$, where $\hat{h}(\theta) = \begin{cases} \theta - \overline{Y}_1 & \text{if } \theta \leq \frac{\overline{Y}_1+\overline{Y}_2}{2}, \\ \overline{Y}_2 - \theta & \text{if } \theta \geq \frac{\overline{Y}_1+\overline{Y}_2}{2}. \end{cases}$ Finally, note also that $\text{cl}(\text{int}([\mu_1, \mu_2])) = [\mu_1, \mu_2]$.Thus, the Andrews, Berry and Jia (2004) condition is easily verified.

7.2 Returns to Schooling:

Suppose potential income, Y, is related to educational attainment, $Educ$ by the equation

$$Y = \alpha_0 + \alpha_1 Educ + \alpha_2 Educ^2 + U.$$

Suppose also that quarter of birth, B, is orthogonal to U. Then letting $\theta := (\alpha_0, \alpha_1, \alpha_2)$ and $\mathcal{A}$ denoting a compact non-empty subset of $\mathbb{R}^3$ we have

$$\varphi(\theta) = \begin{pmatrix} E(Y - \alpha_0 - \alpha_1 Educ - \alpha_2 Educ^2) \\ E[(Y - \alpha_0 - \alpha_1 Educ - \alpha_2 Educ^2)B] \end{pmatrix},$$

and

$$\Theta_0 = \{(\alpha_0, \alpha_1, \alpha_2) \in \mathcal{A} : \varphi(\theta) = 0\}.$$

[10]Note that if $\mu_2 = \mu_1$ then the identification problem disappears.

Note that the Andrews, Berry and Jia (2004) approach is not applicable in this case. The Jacobian condition imposed in this paper on the other hand is that the rank of the following matrix is 2:

$$D\varphi(\theta) = \begin{bmatrix} -1 & -E(Educ) & -E(Educ^2) \\ -E(B) & -E(Educ \cdot B) & -E(Educ^2 \cdot B) \end{bmatrix}.$$

This condition will be satisfied as long as it is not the case that $E(Educ{\cdot}B) = E(Educ){\cdot}E(B)$ and $E(Educ^2 \cdot B) = E(Educ^2) \cdot E(B)$.

Instead of estimating conditional expectation of potential income given schooling one could also try to estimate a certain quantile function of returns to schooling. For example, Chernozhukov and Hansen (2006) consider a model in which the τ (with $\tau \in (01,))$ structural quantile function for returns to schooling is specified to be a linear function, so that the moment conditions at hand are

$$E[(\tau - 1\{Y \leq X^T\theta_0\})Z] = 0. \tag{15}$$

Then assuming that the potential outcome is a continuous random variable the Jacobian condition of this paper is satisfied if the $d_z \times d_x$ matrix

$$D\varphi(\theta) = -E\begin{bmatrix} ZX_1 f_Y(X^T\theta|X,Z) & ZX_2 f_Y(X^T\theta|X,Z) & \cdots & ZX_{d_x} f_Y(X^T\theta|X,Z) \end{bmatrix}$$

has rank d_z for all θ values satisfying equation (15).

8 Conclusion

This paper has proposed conditions under which the most intuitive estimator in parametric partially identified moment equality and inequality models as well as non-parametric regression inequality models is consistent for the identified set. The paper has also proposed alternative M-estimators which agree with the set of parameters satisfying the sample versions of the moment conditions that characterize the model. This paper has focused on estimation only. The results of this paper, however, can be combined with the methods developed in Chernozhukov, Hong and Tamer (2007) or in other papers.

For parametric models most of the conditions proposed in this paper are for models in which the number of moment conditions is less than or equal to the dimension of the parameter space. When the number of moment conditions is larger than the number of parameters one could use the conditions proposed here by selecting as many of the conditions as the number of parameters to consistently estimate the set of parameters that satisfy the selected subset of moment conditions. This set of course will always contain the identified set. Nevertheless, one could iteratively use the subsampling procedure of Chernozhukov, Hong and Tamer (2007) by taking the estimator for this set, denoted by $\overline{\Theta}_n$, as the starting point for the iterations. Iterating on this procedure using the whole parameter set as the initial point was suggested by Romano and Shaikh (2006 a,b). I expect that using $\overline{\Theta}_n$ as the initial point as opposed to the whole parameter set would significantly decrease the computational burden of this method.

9 Appendix:

This section states the Generalized Inverse Function Theorem and its Corollary which are used throughout the paper and proves the corollary. We also state Egoroff's Theorem in this section. The statement of the Generalized Inverse Function Theorem is taken from Luenberger(1969), p. 240. Note that the transformation T is assumed to be continuously Féchet differentiable.

Theorem 9.1 *(Generalized Inverse Function Theorem) Let x_0 be a regular point of a transformation T mapping the Banach space X into the Banach space Y. Then there is a neigh-*

borhood $N(y_0)$ of the point $y_0 = T(x_0)$ (i.e., a sphere centered at y_0) and a constant K such that the equation $T(x) = y$ has a solution for every $y \in N(y_0)$ and the solution satisfies $\|x - x_0\| \leq K\|y - y_0\|$."

Corollary 9.1 *Suppose every point in the compact set $A \subseteq X$ is a regular point of T. Then there exist $r > 0$ and $K < \infty$ such that the following statement holds for each $x_0 \in A$ and $y_0 = T(x_0)$: Whenever $y \in B_r(y_0)$, with $y_0 = T(x_0)$ the equation $y = T(x)$ has a solution and the solution satisfies $\|x - x_0\| \leq K\|y - y_0\|$.*

Proof 9.1 *Since T is Fréchet differentiable it maps compact subsets of X into compact sets, so that $T(A)$ is compact. Since each $x_0 \in A$ is regular the theorem tells us that there exists r_{y_0}, K_{y_0} as in the statement of the theorem. Now $\cup_{y_0 \in T(A)} B_{\frac{r_{y_0}}{2}}(y_0)$ is an open cover for $T(A)$. Thus, there exist $y_{01}, ..., y_{0J}$ such that $T(A) \subseteq \cup_{j=1}^{J} B_{\frac{r_{y_{0j}}}{2}}(y_{0j})$. Let $r := \frac{1}{2}\min\{r_{y_{01}}, ..., r_{y_{0J}}\}$. Next observe that $\cup_{y_0 \in T(A)} B_r(y_0)$ is also an open cover of $T(A)$ and it has a finite subcover $y_{01}, ..., y_{0S}$. Let $K := \max\{K_{y_{01}}, ..., K_{y_{0S}}\}$. Now consider an arbitrary $x_0 \in A$ with $y_0 = T(y_0)$. Suppose $y \in B_r(y_0)$. Then $y \in B_{2r}(y_{0s}) \subseteq B_{r_{y_{0s}}}(y_{0s})$ for some $s = 1, ..., S$. Morover,$y_{0s} = T(x_{0s}$ for some $x_{0s} \in A$, and the equation $y = T(x)$ has a solution and the solution satisfies $\|x - x_{0s}\| \leq K_{y_{0s}}\|y - y_{0s}\| \leq K\|y - y_{0s}\|$.* ■

Theorem 9.2 *(Egoroff's Theorem from p. 73-74 of Royden (1988).) "If $< f_n >$ is a sequence of measurable functions that converge to a real valued function f a.e. on a measurable set E of finite measure, then given $\eta > 0$, there is a subset $A \subset E$ with $mA < \eta$ such that f_n converges uniformly on $E \sim A$."*

References

[1] Amemiya, T. (1994): *Introduction to Statistics and Econometrics,* Cambridge: Harvard University Press.

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

[3] Andrews, D. K., and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection.," Cowles Foundation Working Paper CFDP 1631.

[4] Bellman, R. (1970): *Introduction to Matrix Analysis,* 2nd ed. New York: McGraw-Hill.

[5] Beresteanu, A., and F. Molinari (2006): "Asymptotic Properties for a Class of Partially Identified Models," forthcoming in *Econometrica.*

[6] Bugni, F. (2007): "Bootstrap Inference in Partially Identified Models," Unpublished Manuscript, Department of Economics, Northwestern University.

[7] Canay, I. A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Unpublished Manuscript, Department of Economics, University of Wisconsin.

[8] Chernozhukov, V., and C. Hansen (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491-525.

[9] Chernozhukov, V., H. Hong, and E. Tamer (2002): "Parameter Set Inference in a Class of Econometric Models," Working Paper, MIT.

[10] Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243-84.

[11] Galichon, A., and M. Henry (2006a): "Inference in Incomplete Models'" Columbia University Discussion Paper 0506-28.

[12] Galichon, A., and M. Henry (2006b): "Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models," Unpublished Manuscript, Harvard University and Columbia University.

[13] Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845-72.

[14] Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, Paperback ed. New York: John Wiley & Sons.

[15] Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519-46.

[16] Mas-Colell, A., M. D. Whinston, and J.R. Green (1995): *Microeconomic Theory,* New York: Oxford University Press.

[17] Pakes, A. , J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Applications," Working Paper, Harvard University.

[18] Pollard, D. (1984): *Convergence of Stochastic Processes,* New York: Springer-Verlag.

[19] Romano, J., and A. M. Shaikh (2006a): "Inference for the Identifed Set in Partially Identifed Econometric Models," Technical Report 2006-10, Department of Statistics, Stanford University.

[20] Romano, J., and A. M. Shaikh (2006b): "Inference for Identifable Parameters in Partially Identifed Econometric Models," Technical Report 2006-9, Department of Statistics, Stanford University.

[21] Rosen, A. (2006): "Confidence Sets for Partially Identi.ed Parameters that Satisfy a Finite Number of Moment Inequalities," Working Paper, University College London.

[22] Royden, H., L. (1988): *Real Analysis,* 3rd ed. New York: Macmillan Publishing Company and London: Collier Macmillan Publishing.