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

Neşe Yıldız†

First version: August 16, 2004. This version: April 15, 2008.

Abstract

This note 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 note 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. Finally, the note studies the model where the information at hand consists of inequality constraints on non-parametric regression functions and shows the consistency of the plug-in estimator or M-estimators that agree with that estimator with probability approaching to one.

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 note 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. Finally, the note studies models in which the only available information to the researcher is in the form of non-parametric regression inequalities and shows that both the plug-in estimator and the proposed extremum estimator are consistent in the Hausdorff metric.

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.

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) and parts of Chernozhukov, Hong and Tamer (2007) 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 note 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 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 has two other advantages. First the data may be used to check whether this condition is satisfied. Second, this condition

[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 note their condition (4.6) holds.

can be extended to models of non-parametric regression inequalities.

The rest of this note 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 the certain models consisting of regression inequalities. Section 8 concludes. The main mathematical tools employed are described in the Appendix.

2 Description of the Parametric 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. Also for each set A, let $\overline{A}$ denote the closure of A. Suppose $\{X_i\}_{i=1}^n$ is an i.i.d. sequence of random variables defined on a complete and inner regular probability space $(\Omega, \mathcal{F}, P)$. 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. 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 $\varphi(\theta)$. For example, $\hat{g}_n(\theta)$ could be $\frac{1}{n}\sum_{j=1}^n \tilde{g}(X_j, \theta)$. 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,

$$\sup_{\hat{\theta} \in \hat{\Theta}_n} \inf_{\theta \in \Theta_0} ||\hat{\theta} - \theta|| \xrightarrow{P} 0, \tag{1}$$

and

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

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

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 $E\|\tilde{\varphi}(X,\theta)\| < \infty$ $\forall \theta \in \Theta$.*

Let $\eta > 0$. Since Θ is compact and $\tilde{g}$ is continuous, there exists $\alpha > 0$ such that $\|\tilde{g}(x,\theta') - \tilde{g}(x,\theta)\| < \eta$ whenever $\|\theta' - \theta\| < \alpha$. Since Θ is compact there exists $\{\theta_1, ..., \theta_K\} \subseteq \Theta$ such that every $\theta \in \Theta$ is less than α distant from some θ_k. This means that $\tilde{g}(x,\theta_k) - \frac{\eta}{2} < \tilde{g}(x,\theta) < \tilde{g}(x,\theta_k) + \frac{\eta}{2}$ for each fixed x. Let $\tilde{g}_{\eta,U}(x) := \tilde{g}(x,\theta_k) + \frac{\eta}{2}$ and $\tilde{g}_{\eta,L}(x) := \tilde{g}(x,\theta_k) - \frac{\eta}{2}$. Then $E(\tilde{g}_{\eta,U} - \tilde{g}_{\eta,L}) < \eta$. Then by Theorem 2 on p. 8 of Pollard (1984) we have $\sup_{\theta \in \Theta} ||\hat{g}_n(\theta) - g(\theta)|| \xrightarrow{a.s.} 0$. Similarly, we can show that $\sup_{\theta \in \Theta} ||\hat{\varphi}_n(\theta) - \varphi(\theta)|| \xrightarrow{a.s.} 0$.

3 Inequality Constraints Only:

I will first consider the case where $S = 0$, i.e. the case where we only have M inequality constraints.[2]

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,

[2]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.

so that

$$\sup_{\hat{\theta}\in\hat{\Theta}_n}\inf_{\theta_0\in\Theta_0}||\hat{\theta}-\theta_0|| \leq \sup_{\hat{\theta}\in\overline{\Theta}^{\epsilon}}\inf_{\theta_0\in\Theta_0}||\hat{\theta}-\theta_0||, \tag{3}$$

and

$$\sup_{\theta_0\in\Theta_0}\inf_{\hat{\theta}\in\hat{\Theta}_n}||\hat{\theta}-\theta_0|| \leq \sup_{\theta_0\in\Theta_0}\inf_{\hat{\theta}\in\underline{\Theta}^{\epsilon}}||\hat{\theta}-\theta_0||, \tag{4}$$

with probability approaching to 1. Thus, showing that (3) and (4) 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 (3) approaches 0 as ϵ decreases to 0:

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

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

Proof 3.1 *Consider first the case where $M = 1$. If $g(\theta) \geq 0$ for each $\theta \in \Theta$, then there is nothing to prove because in this case both Θ_0 and $\overline{\Theta}^{\epsilon}$ equal Θ for all $\epsilon > 0$. So suppose there is some $\underline{\theta} \in \Theta$ such that $g(\underline{\theta}) < 0$. Then for $\epsilon \in (0, -g(\underline{\theta}))$, $\overline{\Theta}^{\epsilon} \neq \Theta$. Next, let $\delta > 0$. We need to show that there exists $\overline{\epsilon}$ such that*

$$\sup_{\theta\in\overline{\Theta}^{\epsilon}}\inf_{\theta\in\Theta_0}||\theta-\theta_0|| \leq \delta \tag{5}$$

whenever $\epsilon \leq \overline{\epsilon}$. Consider $S := \{g(\theta) : \theta \in \Theta \setminus [\cup_{\theta_0\in\Theta_0} B_\delta(\theta_0)]\}$. If $S = \emptyset$, then $\overline{\Theta}^{\epsilon} \subseteq \cup_{\theta_0\in\Theta_0} B_\delta(\theta_0)$, and statement (5) is true. Otherwise, it must be that $s := \sup S > -\infty$. In addition, note that S is compact because closed subsets of a compact sets and images of compact sets under continuous functions are compact. Since g is continuous the supremum of S must be attained by an element of S, which implies that $s < 0$. We claim that statement (5) holds for $\overline{\epsilon} = -s/2$. To see this note that for any $\epsilon \leq \overline{\epsilon}$ there are two possibilities: either there exists θ' with $g(\theta') \in [-\epsilon, 0)$, or there is no such θ. In the latter case, $\overline{\Theta}^{\epsilon} = \Theta_0$, and $\sup_{\theta\in\overline{\Theta}^{\epsilon}}\inf_{\theta\in\Theta_0}||\theta-\theta_0|| = 0$. On the other hand, if $g(\theta') \in [-\epsilon, 0)$ for some θ' then because

$-\epsilon > s$, $g(\theta')$ cannot belong to S. But this means that $\overline{\Theta}^{\epsilon} \setminus \Theta_0 \subseteq \cup_{\theta_0 \in \Theta_0} B_\delta(\theta_0)$. In other words, $\sup_{\theta \in \overline{\Theta}^{\eta}} \inf_{\theta \in \Theta_0} ||\theta - \theta_0|| \leq \delta$. Since δ could be chosen arbitrarily, this proves the proposition for $M = 1$.

This result easily extends to the case where $g : \mathbb{R}^L \to \mathbb{R}^M$, with $1 < M < \infty$. To see this, let $h(\theta) := \min\{g_1(\theta), ..., g_M(\theta)\}$. Note that since each g_i is continuous, so is h. Now repeat the same arguments above with h in place of g. ■

Remark 3.1 *The first part of the proof is essentially the same as the proof of the corresponding direction in the consistency proof(s) of Andrews, Berry and Jia (2004), and Manski and Tamer (2002).*

To guarantee consistency of $\hat{\Theta}_n$ we also need to show that (4) 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 *Consider an open subset, $\mathcal{O}$ of $\mathbb{R}^I$, containing Θ.[3] Suppose I and M are finite integers. Assume that the function $\tilde{g} : \mathcal{X} \times \mathcal{O} \to \mathbb{R}^M$ is continuously differentiable in θ for almost every x, and that $E\left[\left|\frac{\partial \tilde{g}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty$ $\forall \theta \in \Theta$, $\forall m$, $\forall j$. 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

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

[3] Defining the function g on $\mathcal{O}$ as opposed to Θ is without loss of generality because of Tietze's extension theorem.

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 $\Theta^* \subseteq \mathrm{int}(\Theta)$. Then under assumptions (2.1), (2.2) and (3.1), we have*

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

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 $\Theta^* \subseteq \mathrm{int}(\Theta)$ and Θ^* is compact there exists $\eta > 0$, with η not depending on θ^*, such that $B_\eta(\theta^*) \subseteq \Theta$. 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*

$$\epsilon_1^* := \frac{1}{2} \inf\{h(\theta) : h(\theta) \geq 0, \theta \in \Theta \setminus \cup_{\theta^* \in \Theta^*} B_\rho(\theta^*)\}. \tag{6}$$

[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 $\mathbb{R}^M$.

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* $g(\theta) = t$ *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* θ_0^* *and*

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

Finally, since $h(\theta_0) \leq g_m(\theta_0)$, $\forall m$, $h(\theta') \geq \epsilon$, *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.* ■

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

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

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

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:

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

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 *Consider an open subset, $\mathcal{O}$ of $\mathbb{R}^I$, containing Θ. Suppose I and S are finite integers. Assume that the function $\tilde{\varphi} : \mathcal{X} \times \mathcal{O} \to \mathbb{R}^S$ is continuously differentiable in θ for almost every x, and that $E\left[\left|\frac{\partial \tilde{\varphi}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty$ $\forall \theta \in \Theta$, $\forall m$, $\forall j$. 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 ${}_pJ_q(\theta)$ denote the determinant of the submatrix consisting of the first p rows and q columns of $D\varphi(\theta)D\varphi(\theta)^T$. Let ${}_p\hat{J}_q(\theta)$ be defined in an analogous way with $\hat{\varphi}_n(\theta)$ replacing $\varphi(\theta)$. By assumption (4.1) ${}_sJ_s(\theta_0) > 0$ $\forall s$ and $\forall \theta_0 \in \Theta_0$. By assumption (4.1) and compactness of Θ_0 $\underline{\underline{\lambda}}^E := \min\{\inf\{{}_sJ_s(\theta) : \theta \in \Theta_0\} : s = 1, ..., S\} > 0$.*

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 a set $A_1 \subseteq \mathcal{X}^\infty$ with $P^\infty(A_1) > 1 - \frac{\delta}{2}$ and an integer N_1 such that $\forall x \in A_1$, $\forall n \geq N_1$ and $\forall s$, we have*

$$\sup_{\theta\in\Theta} |_s\hat{J}_s(x,\theta) -_s J_s(x,\theta)| < \frac{\underline{\underline{\lambda}}^E}{4}. \tag{7}$$

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 $K < \infty$ 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 $\nu \in \big(0, \min\{r, \frac{\eta}{K}, \frac{\epsilon}{K}\}\big)$. Using Assumption (2.2) and Egoroff's Theorem once more we can argue that there is a set $A_2 \subseteq \mathcal{X}^\infty$ with $P^\infty(A_2) > 1 - \frac{\delta}{2}$ and an integer N_2 such that $\forall x \in A_2$ and $\forall n \geq N_2$, we have

$$\sup_{\theta\in\Theta} \|\hat{\varphi}_n(x,\theta) - \varphi(\theta)\| < \nu. \tag{8}$$

Note that $P^\infty(A_1 \cap A_2) > 1 - \delta$, and that $\forall x \in A_1 \cap A_2$ and $\forall n \geq \max\{N_1, N_2\}$ both (7) and (8) hold. Next, consider any $\theta_0 \in \Theta_0$. Let $x \in A_1 \cap A_2$ and $n \geq \max\{N_1, N_2\}$. Then $\|\hat{\varphi}_n(x,\theta_0)\| < r$ (by (8)) 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 $\forall\epsilon > 0$

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

[9]For a statement of this theorem, please refer to p. 73 of Royden (1988).

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 $\hat{\Theta}_n \neq \emptyset$,

$$\hat{\Theta}_n = \hat{\Theta}_n^a := \{\theta' \in \Theta : \theta' \text{ minimizes } \hat{\varphi}_n(\theta)^T W \hat{\varphi}_n(\theta)\},$$

where W is a positive definite symmetric matrix. 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 *Consider an open subset, $\mathcal{O}$ of $\mathbb{R}^I$, containing Θ. Suppose I, M and S are finite integers. Assume that the functions $\tilde{\varphi} : \mathcal{X} \times \mathcal{O} \to \mathbb{R}^S$ and $\tilde{g} : \mathcal{X} \times \mathcal{O} \to \mathbb{R}^M$ are continuously differentiable in θ for almost every x, and that $E\left[\left|\frac{\partial \tilde{\varphi}_s(X,\theta)}{\partial \theta_j}\right|\right] < \infty$ and $E\left[\left|\frac{\partial \tilde{g}_m(X,\theta)}{\partial \theta_j}\right|\right] < \infty \forall \theta \in \Theta$, $\forall s, m\ \forall j$. 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 $\Theta^{**} \subseteq \mathrm{int}(\Theta)$. 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 $\Theta_0 \subseteq \mathrm{int}(\Theta)$ 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 ${}_pJ_q(\theta)$ *and* ${}_p\hat{J}_q(\theta)$ *be defined as in the proof of Proposition (4.1). Let* ${}_pJ_q^f(\theta)$ *and* ${}_p\hat{J}_q^f(\theta)$ *analogously with* f *and* $\hat{f}_n$ *replacing* φ *and* $\hat{\varphi}_n$*, respectively. Also let* $\underline{\underline{\lambda}} := \min\{\inf\{{}_sJ_s(\theta) : \theta \in \Theta_0\}, \inf\{{}_lJ_l^f(\theta) : \theta \in F\} : s = 1, ..., S, l = 1, ..., M+S\}$.

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* $A_1 \subseteq \mathcal{X}^\infty$ *with* $P^\infty(A_1) > 1 - \frac{\delta}{2}$ *such that* $\forall\omega \in A_1$, $\forall s = 1, ..., S$, $\forall l = 1, ..., M+S$ *and* $\forall n \geq N_1$ *we have*

$$\begin{aligned} \sup_{\theta\in\Theta}\left|{}_s\hat{J}_s(\theta) -_s J_s(\theta)\right| &< \frac{\underline{\underline{\lambda}}}{4}, \text{ and} \\ \sup_{\theta\in\Theta}\left|{}_l\hat{J}_l^f(\theta) -_l J_l^f(\theta)\right| &< \frac{\underline{\underline{\lambda}}}{4}. \end{aligned} \tag{9}$$

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*

$$\sup_{\theta\in\Theta}\left\|\begin{pmatrix} \hat{h}_n(\omega,\theta) \\ \hat{g}_n(\omega,\theta) \\ \hat{\varphi}_n(\omega,\theta) \end{pmatrix} - \begin{pmatrix} h(\theta) \\ g(\theta) \\ \varphi(\theta) \end{pmatrix}\right\| < \nu. \tag{10}$$

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 (10), $\hat{h}_n(\hat{\theta}_n) > 0$ 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 $\hat{\Theta}_n \neq \emptyset$

$$\hat{\Theta}_n = \hat{\Theta}_n^a := \{\theta \in \Theta : \theta \text{ minimizes } |\hat{h}_n(\theta')|1\{\hat{h}_n(\theta') \leq 0\} + \hat{\varphi}_n(\theta')^T W \hat{\varphi}_n(\theta')\},$$

where W is a positive definite matrix. 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

$$\sup_{\hat{\theta} \in \hat{\Theta}_n^1 \cap \hat{\Theta}_n^2} \inf_{\theta_0 \in \Theta_0^1 \cap \Theta_0^2} ||\hat{\theta} - \theta_0|| \leq \sup_{\hat{\theta} \in \overline{\Theta}^{\epsilon,1} \cap \overline{\Theta}^{\epsilon,2}} \inf_{\theta_0 \in \Theta_0^1 \cap \Theta_0^2} ||\hat{\theta} - \theta_0||, \tag{11}$$

$$\sup_{\theta_0 \in \Theta_0^1 \cap \Theta_0^2} \inf_{\hat{\theta} \in \hat{\Theta}_n^1 \cap \hat{\Theta}_n^2} ||\hat{\theta} - \theta_0|| \leq \sup_{\theta_0 \in \Theta_0^1 \cap \Theta_0^2} \inf_{\hat{\theta} \in \underline{\Theta}^{\epsilon,1} \cap \underline{\Theta}^{\epsilon,2}} ||\hat{\theta} - \theta_0||, \tag{12}$$

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,

$$\sup_{\theta_0 \in \Theta_0^l} \inf_{\hat{\theta} \in \underline{\Theta}^{\epsilon,l}} ||\hat{\theta} - \theta_0|| \to 0 \text{ for } l = 1, 2, \nRightarrow \sup_{\theta_0 \in \Theta_0^1 \cap \Theta_0^2} \inf_{\hat{\theta} \in \underline{\Theta}^{\epsilon,1} \cap \underline{\Theta}^{\epsilon,2}} ||\hat{\theta} - \theta_0|| \to 0.$$

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 (12) 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 Nonparametric Regression Inequalities:

Consider the model of the form

$$Y \leq g_0(X, Z_1) + \varepsilon, \tag{13}$$

with $E(\varepsilon|Z) = 0$, $Z = (Z_1^T, Z_2^T)^T$. Initially, we assume that Y is $M \times 1$ random vector, with $M < \infty$. g_0 denotes the unknown structural function of interest, X is a $d_x \times 1$ vector of explanatory variables, Z_1 and Z_2 are $d_1 \times 1$ and $d_2 \times 1$ vectors of instrumental variables, and ε is unobserved. Taking conditional expectation of equation (13) yields the integral equation

$$\pi(Z) = E(Y|Z) \leq E[g_0(X, Z_1)|Z] = \int g_0(x, z_1) \mathrm{d}F_{X|Z}(x). \tag{14}$$

Since $S := (Y, X^T, Z^T)^T$ are observed, π and F, the joint distribution of S, and hence $F_{X|Z}$, the conditional distribution of X given Z are identified. Our main goal is to be able make inferences about g_0 using the available information.

Let $\mathcal{L}_F^2$ denote the set of square integrable (with respect to F) and measurable functions of S, and let $W := (X^T, Z_1^T)^T$. In addition, let $\mathcal{L}_F^2(Y)$, $\mathcal{L}_F^2(W)$ and $\mathcal{L}_F^2(Z)$ denote subspaces of $\mathcal{L}_F^2$ consisting of real valued functions depending on Y, W or Z only. Note that we assume that $Y \in \mathcal{L}_F^2(Y)$ and $\pi(Z) \in \mathcal{L}_F^2(Z)$. Also define

$$T_F : \mathcal{L}_F^2(W) \to \mathcal{L}_F^2(Z),\ g \to T_F(g) := E[g(X, Z_1)|Z_1, Z_2] - \pi(Z_1, Z_2).$$

Note that if the Fréchet differential, $\delta T_F(g; h) \in \mathcal{L}_F^2(Z)$, of T_F at $g \in \mathcal{L}_F^2(W)$ with increment $h \in \mathcal{L}_F^2(W)$ exists, it is defined as a linear and continuous mapping with respect to h that is defined by

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

Since T_F is linear, $T_F(g+h) - T_F(g) = T_F(h)$. Thus, $\delta T_F(g; h)$ exists for each $g \in \mathcal{L}_F^2(W)$

and each $h \in \mathcal{L}^2_F(W)$; it is given by

$$\delta T_F(g;h) = E[h(W)|Z].$$

Since $\delta T_F(g;h)$ is continuous and linear in h by definition, we can write $\delta T_F(g;h) = T'_F(g)h$, where $T'_F(g)$ defines a transformation from $\mathcal{L}^2_F(W)$ into the normed linear space $B\big(\mathcal{L}^2_F(W), \mathcal{L}^2_F(Z)\big)$[10]. This transformation is called the Fréchet derivative T'_F of T_F. Note in our case T'_F does not depend on g. Thus, in our case, T'_F is trivially continuous in g and the mapping T_F is continuously Fréchet differentiable.

Define g a regular point of T_F if $T'(g)$ maps $\mathcal{L}^2_F(W)$ onto $\mathcal{L}^2_F(Z)$. In our case, $T'(g) = T'_F$ for each $g \in \mathcal{L}^2_F(W)$, so each $g \in \mathcal{L}^2_F(W)$ is a regular point if for each $r \in \mathcal{L}^2_F(Z)$ there exists an $h \in \mathcal{L}^2_F(W)$ such that $T'_F h = E[h(W)|Z] = r(Z)$. When $\sigma(W) \subseteq \sigma(Z)$. Note that in this case X is exogenous because

$$E(\varepsilon|X) = E[E(\varepsilon|W)|X] = E\{E[E(\varepsilon|Z)|W]|X\} = 0.$$

In addition, in this special case, every $g \in \mathcal{L}^2_F(W)$ is a regular point of T_F.

Suppose $\mathcal{G}$ is a non-empty and compact (in the $\mathcal{L}^2$ norm) subset of $\mathcal{L}^2_F(W)$. Define

$$\begin{aligned}
\mathcal{G}_0 &:= \{g \in \mathcal{G} : \int g(x, z_1)\mathrm{d}F_{X|Z}(x|z) - \pi(Z) \geq 0\}, \\
\mathcal{G}^* &:= \{g \in \mathcal{G} : \int g(x, z_1)\mathrm{d}F_{X|Z}(x|z) - \pi(Z) = 0\}, \\
\hat{\mathcal{G}} &:= \{g \in \mathcal{G} : \int g(x, z_1)\mathrm{d}\hat{F}_{X|Z}(x|z) - \hat{\pi}(Z) \geq 0\}, \\
\overline{\mathcal{G}}^\epsilon &:= \{g \in \mathcal{G} : \int g(x, z_1)\mathrm{d}F_{X|Z}(x|z) - \pi(Z) \geq -\epsilon \cdot \mathbf{1}\}, \\
\underline{\mathcal{G}}^\epsilon &:= \{g \in \mathcal{G} : \int g(x, z_1)\mathrm{d}F_{X|Z}(x|z) - \pi(Z) \geq \epsilon \cdot \mathbf{1}\},
\end{aligned}$$

[10]This is the space of all bounded linear operators from $\mathcal{L}^2_F(W)$ into $\mathcal{L}^2_F(Z)$. Note that since $\mathcal{L}^2_F(Z)$ is complete, this space is complete as well.

where $\mathbf{1}$ is the M dimensional vector of ones.

Proposition 7.1 *Suppose $\mathcal{G}^* \subseteq \text{int}(\mathcal{G})$. Then as $\epsilon \to 0$,*

$$\sup_{g \in \overline{\mathcal{G}}^\epsilon} \inf_{g_0 \in \mathcal{G}_0} \|g - g_0\|_{\mathcal{L}^2_F} \to 0, \tag{15}$$

$$\sup_{g_0 \in \mathcal{G}_0} \inf_{g \in \underline{\mathcal{G}}^\epsilon} \|g - g_0\|_{\mathcal{L}^2_F} \to 0. \tag{16}$$

Proof 7.1 *Let $h(g; Z) := \min\{E[g_1(X, Z_1)|Z] - \pi_1(Z), ..., E[g_M(X, Z_1)|Z] - \pi_M(Z)\}$. Then the proof of Proposition (3.1) goes through in $\mathcal{L}^2$ norm without any modification. Thus, (16) is true. To show (16), first observe that there exist $\eta > 0$ such that $B_\eta^{\mathcal{L}^2}(\mathcal{G}^*) \subseteq \mathcal{G}$ because by the conditional version of Jensen's inequality h is a continuous function of g. Next, let $\delta > 0$. By the Generalized Inverse Function Theorem for each $g_0 \in \mathcal{G}_0$ there exist $r > 0$ and $K < \infty$, which do not depend on g_0, such that for each $t \in B_r^{\mathcal{L}^2(Z)}(0)$ the equation $T_F(g) = t$ has a solution, and the solution satisfies $\|g - g_0\|_{\mathcal{L}^2(W)} \leq K\|T_F(g) - T_F(g_0)\|_{\mathcal{L}^2(Z)}$. Then let $g_0 \in \mathcal{G}_0$ and consider $\epsilon \in \left(0, \min\left\{\frac{r}{\sqrt{M}}, \frac{\delta}{K\sqrt{M}}, \frac{\eta}{2\sqrt{M}}\right\}\right)$. Define $t_m(Z) := E[g_{0m}(X, Z_1)|Z] - \pi(Z) + \epsilon - h(g_0; Z)$ for $m = 1, ..., M$. The rest of the proof is the same as the last part of the proof of Proposition (3.2).* ■

To prove consistency of the plug-in estimator we also need to show that $\underline{\mathcal{G}}^\epsilon \subseteq \hat{\mathcal{G}} \subseteq \overline{\mathcal{G}}^\epsilon$ with probability approaching to one as the sample size increases. For this purpose suppose for the moment that the conditional distribution of X given Z is absolutely continuous with respect to Lebesgue measure. Then,

$$\begin{aligned}\hat{T}_F(g) - T_F(g) &= \int g(x, z_1)\hat{f}_{X|Z}(x|z)\mathrm{d}x - \hat{\pi}(z) - \left(\int g(x, z_1)f_{X|Z}(x|z)\mathrm{d}x - \pi(z)\right) \\ &= \int g(x, z_1)[\hat{f}_{X|Z}(x|z) - f_{X|Z}(x|z)]\mathrm{d}x - [\hat{\pi}(z) - \pi(z)].\end{aligned}$$

These equalities demonstrate that it is possible to find conditions on the joint distrbution of X and Z which will guarantee that $\underline{\mathcal{G}}^\epsilon \subseteq \hat{\mathcal{G}} \subseteq \overline{\mathcal{G}}^\epsilon$ with probability approaching to one as the sample size increases.

8 Conclusion

This note 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 note 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 note has focused on estimation only. The results of this note, however, can be combined with the methods developed in Chernozhukov, Hong and Tamer (2007) or Andrews, Berry and Jia (2004).

For parametric models most of the conditions proposed in this note 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 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.

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

[4] Beresteanu, A., and F. Molinari (2006): "Asymptotic Properties for a Class of Partially Identified Models," Working Paper, Duke University and Cornell University.

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

[6] Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," forthcoming in *Econometrica.*

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

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

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

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

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

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

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

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

[15] 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.

[16] 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.

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

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

9 Mathematical Tools:

Corollary 9.1 *(to the Generalized Inverse Function Theorem) Suppose every element, x_0, of a compact set F is a regular point of a continuously (Fréchet) differentiable transformation T mapping the Banach space $X = \mathbb{R}^I$ into the Banach space $Y = \mathbb{R}^M$. Then there is $s > 0$*

and $K < \infty$ such that for each $x_0 \in F$ and $y_0 = T(x_0)$, the equation $T(x) = y$ has a solution for every $y \in B_s(y_0)$, and the solution satisfies $||x - x_0|| \leq K||y - y_0||$.

Definition 9.1 *Let T be a continuously Fréchet differentiable transformation from an open set S in a Banach space X into a Banach space Y, and let $DT(\cdot)$ denote its Fréchet derivative. If $x_0 \in S$ is such that $DT(x_0)$ maps X onto Y, the point x_0 is said to be a* regular point *of the transformation T.*[11]

Before stating the proof of this corollary let me describe the notation, and provide some small results that are used in the proof:

Remark 9.1 $||DT(x_0)|| := \sup_{||z|| \leq 1} ||DT(x_0)z||$*, and it is equal to the absolute value of the largest eigenvalue of $[DT(x_0)^T DT(x_0)]$. In addition, if $x_0 \in F$ where F is a compact set, then by the Theorem of the Maximum, $||DT(x_0)||$ is uniformly continuous on F (as a function of x_0).*

Remark 9.2 *"Let $L_0 := \{x : DT(x_0)x = 0\}$. Since L_0 is a closed subspace, X/L_0 is a Banach space. Define the operator A on this space by $A_{x_0}[x] = DT(x_0)x$, where $[x]$ denotes the class of elements equivalent to x modulo L_0. The operator is well defined, since $x_1 \in [x] \Leftrightarrow DT(x_0)(x - x_1) = 0 \Leftrightarrow DT(x_0)x = DT(x_0)x_1$, so that equivalent elements x yield identical elements $y \in Y$. Furthermore, this operator is linear, continuous, one-to-one, and onto; hence, it has a linear inverse. Moreover, by the Banach inverse theorem its inverse is continuous." (Luenberger, page 240, beginning of the proof the Generalized Inverse Theorem.)*

Definition 9.2 *Let L be a subspace of a vector space X.*

1. *Two elements $x_1, x_2 \in X$ are said to be equivalent modulo L if $x_1 - x_2 \in L$. In this case, we write $x_1 \equiv x_2$ and let $[x]$ denote the equivalence class of x.*

[11]This definition is from Luenberger p. 240.

2. *The quotient space X/L consists of all equivalence classes modulo L with addition and scalar multiplication defined by $[x_1] + [x_2] = [x_1 + x_2]$, $\alpha[x] = [\alpha x]$. (Luenberger, page 41.)*

Proposition 9.1 *Let X be a Banach space, L a closed subspace of X, and X/L the quotient space with the quotient norm defined as $||[x]|| = \inf\{||x + l|| : l \in L\}$. Then X/L is also a Banach space. (Luenberger, page 42)*

This proposition is needed for the application of the Banach Inverse Theorem, which tells us that the mapping A_{x_0} is invertible.

Remark 9.3 *Note that $||A_{x_0}^{-1}|| := \sup_{||y||\leq 1} ||A_{x_0}^{-1}(y)||$. Since for each $y \in Y$, $A_{x_0}^{-1}$ is an equivalence class which consists of points that are all mapped to y by $DT(x_0)$*

$$\sup_{||y||\leq 1} ||A_{x_0}^{-1}(y)|| = \sup_{||y||\leq 1} \inf\{||x + v|| : v \in L_0\},$$

where x is such that $DT(x_0)x = y$.

In this note, we are concerned with the case where $X = \mathbb{R}^I$ and $Y = \mathbb{R}^M$, each equipped with the corresponding Euclidean norm. For this special case, let us examine $||A_{x_0}^{-1}||$ more closely. Note that for each fixed $x \in X$, minimizing $||x + s||$ is the same as minimizing $||x + s||^2$. On the other hand, if the minimum exists the necessary condition for $\tilde{v}$ to be a minimizer is

$$2(x + \tilde{v}) + DT(x_0)^T\lambda = 0 \Leftrightarrow \tilde{v} = -x - \frac{1}{2}DT(x_0)^T\lambda,$$

for some $\lambda \in \mathbb{R}^M$. In addition, $\tilde{v} \in L_0 \Leftrightarrow \lambda = -2(DT(x_0)DT(x_0)^T)^{-1}DT(x_0)x$, so that $\tilde{v} = -\big(I - DT(x_0)^T[DT(x_0)DT(x_0)^T]^{-1}DT(x_0)\big)x$.

$$\sqrt{(x + \tilde{v})^T(x + \tilde{v})} = \sqrt{x^T DT(x_0)^T[DT(x_0)DT(x_0)^T]^{-1}DT(x_0)x}.$$

Using $DT(x_0)x = y$, we see that this last expression equals $\sqrt{y^T[DT(x_0)DT(x_0)^T]^{-1}y}$. Thus, $\sup_{||y||\leq 1} ||A_{x_0}^{-1}(y)||$ *equals the square root of the largest eigenvalue of* $[DT(x_0)DT(x_0)^T]^{-1}$.

Remark 9.4 ***Definition 9.3*** *If X and Y are normed spaces, and A is a bounded linear operator from X to Y, then the adjoint operator $A^* : Y^* \rightarrow X^*$ is defined by the equation $< x, A^*y^* >=< Ax, y^* >$, where $< \cdot, \cdot >$ denotes the inner product operator (Luenberger (1969), page 150).*

Theorem 9.1 *The adjoint operator A^* of the bounded linear operator $A : X \rightarrow Y$ is linear and bounded with $||A^*|| = ||A||$ (Luenberger (1969), page 151).*

If $X = \mathbb{R}^I$ and $Y = \mathbb{R}^M$, each equipped with the corresponding Euclidean norm, then a bounded linear function from X to Y could be represented by an $I \times M$ real matrix. Moreover, $A^* = A^T$. thus, the operator norms of A and A^T are equal.

Remark 9.5 *If $DT(x_0)$ has rank M then the $M \times M$, symmetric, positive-definite matrix $DT(x_0)DT(x_0)^T$ has M real and positive eigenvalues. Moreover, if $\lambda_1, ..., \lambda_M$ denote the eigenvalues of $DT(x_0)DT(x_0)^T$ ordered from largest to smallest according to their size then $1/\lambda_M, ..., 1/\lambda_1$ represent the eigenvalues of $[DT(x_0)DT(x_0)^T]^{-1}$ in the same order.*

Proof of the Corollary: Note that $||DT(\cdot)||$ is uniformly continuous on F. Therefore, for each $\epsilon > 0$, we can choose $r > 0$, which is independent of x_0, such that $||x - x_0|| < r \Rightarrow ||DT(x) - DT(x_0)|| < \epsilon$. On the other hand, since F is compact and $||A_{x_0}^{-1}||$ is continuous in x_0, $\sup\{||A_{x_0}^{-1}|| : x_0 \in F\} := c < \infty$. Now we take our ϵ to be in $(0, \frac{1}{4}c)$, then take r_c to be the r value corresponding to this value of ϵ (i.e. for all $x_0 \in F$, $||x - x_0|| < r_c \Rightarrow ||DT(x) - DT(x_0)|| < \epsilon$). Then take $s \in (0, \frac{1}{4c}r_c]$, and $K = 4c$. ■