DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

SIMPLE TWO-STAGE INFERENCE FOR A CLASS OF PARTIALLY IDENTIFIED MODELS

Xiaoxia Shi
University of Wisconsin at Madison

Matthew Shum
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1376

May 2013

Simple Two-Stage Inference for A Class of Partially Identified Models*

Xiaoxia Shi†
University of Wisconsin at Madison

Matthew Shum‡
California Institute of Technology

April 16, 2012

Abstract

This note proposes a new two-stage estimation and inference procedure for a class of partially identified models. The procedure can be considered an extension of classical minimum distance estimation procedures to accommodate inequality constraints and partial identification. It involves no tuning parameter, is nonconservative and is conceptually and computationally simple. The class of models includes models of interest to applied researchers, including the static entry game, a voting game with communication and a discrete mixture model.

Keywords: Implicit Function Theorem, Hausdorff Consistency, Minimum Distance, Partial Identification, Two-stage Inference.

The recent literature on partially identified models has focused on general econometric formulations requiring complicated procedures. Examples of the general formulations include the moment inequality models and the models defined by intersection bounds.[1] In these general formulations, several difficulties for estimation and inference are recognized: (1) available set estimators that are consistent in Hausdorff distance take the form of a level set of a criterion function, where the level is arbitrary (see Chernozhukov et al. (2007)); such arbitrariness arguably constitutes the reason that consistent estimation of the identified set has been overshadowed by confidence set construction in this literature; (2) valid inference

*We thank Yanqin Fan, Patrik Guggenberger, Bruce Hansen and Jack Porter for useful comments and suggestions.

†xshi@ssc.wisc.edu

‡mshum@caltech.edu

[1]e.g. Chernozhukov, Hong and Tamer (2007), Andrews and Soares (2010),Bugni (2010),Canay (2010), Romano and Shaikh ((2008),(2010)) and Chernozhukov, Lee and Rosen (2008).

procedures often require simulation of either the test statistic or the critical values, which may rely on tuning parameters that are hard to choose.[2] Also, in the general models, there is a nearly theological debate on whether we should focus on confidence sets that cover the whole identified set, or those that cover each point in the identified set, with a fixed probability.[3]

In this note, we show that, for a special yet meaningful class of partially identified models, the difficulties above do not arise. These models are of a two-stage nature and we propose new two-stage procedures for the consistent estimation of the identified set and for constructing the confidence set. We show that (1) a sample analogue estimator for the identified set is consistent in Hausdorff distance and the estimator does not rely on an arbitrarily chosen level; (2) asymptotically valid confidence sets can be constructed by inverting simple squared-error type tests with χ^2 critical values, so that no tuning parameter is needed; moreover, the test underlying the confidence set is nonconservative and similar; finally, (3) confidence sets covering the identified set and those covering each point in the identified set with a given probability coincide in a large subclass. The class of models considered here include entry games, voting games, and discrete mixture models, all of which have been of interest to applied researchers.

The main contributions of this note are to provide a new consistent set estimator and a simple confidence set for this class of models. Our procedure can be considered an extension of classical minimum distance estimation procedures to accommodate inequality constraints and partial identification. Besides those, a technical contribution of this note is a new proof of consistency for set estimators. The new proof utilizes an Implicit Correspondence Lemma (ICL) which we prove by generalizing the Implicit Function Theorem. Both the new consistency proof technique and the ICL may be useful in more general models.

There are a small number of papers that address the consistent estimation problem under partial identification. These are Andrews, Berry and Jia (2004), Chernozhukov et al. (2007), and Yildiz (2012). The class of models treated in our paper is different from those treated in those papers. Thus, the assumptions made are not exactly comparable. Nevertheless, we will compare these conditions briefly below, after stating our main consistency result. Moreover, our proof technique is different from that of all the papers mentioned above.

The literature on constructing confidence sets for partially identified model is much larger. For a current survey, see the introduction of Andrews and Shi (2009).

[2]see e.g. Stoye (2010), Andrews and Soares (2010) and Chernozhukov et al. (2008).

[3]The distinction was first pointed out by Imbens and Manski (2004). Subsequently authors in this literature either advocate for one, or propose separate procedures for both. Andrews and his coauthors are representative of the former approach, while Romano and Shaikh (2008),(2010) have taken the latter approach.

In the next section, we describe our model framework, and provide several examples. Section 2 establishes the Hausdorff consistency of our estimated set; Section 3 presents results on confidence set and give conditions under which CS's covering the whole identified set and each point in the set coincide. Assumptions required for the results in Sections 2 and 3 are minimal and we illustrate the verification of them using the entry game example. Technical proofs for the theorems are given in the appendix.

1 The Two-Stage Model

The model considered consists of two stages. In the first stage, a parameter $\beta \in \mathcal{B} \subset R^{d_\beta}$ is point identified and has a consistent and asymptotically normal (CAN) estimator $\hat{\beta}_n$. In the second stage, the model relates the true value β_0 of β to a structural parameter θ (with true value θ_0), through some inequality/equality restrictions:

$$\begin{aligned} g^e(\theta_0, \beta_0) &= 0 \\ g^{ie}(\theta_0) &\geq 0, \end{aligned} \tag{1.1}$$

where $g^{ie} : \Theta \to R^{d_1}$ defines the the inequality restrictions, $g^e : \Theta \times \mathcal{A} \to R^{d_2}$ defines the equality restrictions and $\theta \in \Theta \subset R^{d_\theta}$. The parameter θ is potentially partially-identified. The identified set of θ is

$$\Theta_0 = \{\theta \in \Theta : g^e(\theta, \beta_0) = 0 \text{ and } g^{ie}(\theta) \geq 0\}. \tag{1.2}$$

The two-stage model is closely related to the classical minimum distance problem, but differs from the latter in the partial (vs. point) identification of θ and in the presence of the inequality constraints.

In the model (1.1), the inequality constraints do not depend on β_0. This is not particularly restrictive because one can always convert an inequality constraint into an equality constraint by introducing a slackness parameter, say γ, and adding an inequality constraint: $\gamma \geq 0$. This trick is used in Example 1.1 below.

The two-stage model includes several useful examples which have been studied in the empirical literature on partially identified models. We describe one example – the first one – in detail to illustrate the applicability our framework. The other two are described only briefly to save space. We note that our two-stage model in general is not a special case of the moment inequality models even though the three examples given below are.

Example 1.1. (Entry Game) Following Andrews et al. (2004) and Ciliberto and Tamer

(2009), consider the complete information game and allow only pure strategy equilibria. Take the 2 player game without covariates (homogeneous markets) as a starting example. Player j, $j = 1, 2$ enters the market if the profit of entering exceeds 0: $y_j = \{\pi_j \geq 0\}$. The profit $\pi_j = a_j + \delta_j y_{-j} + \varepsilon_j$, where a_j is the expected monopoly profit, δ_j is the competition effect which is assumed to be negative and $(\varepsilon_1, \varepsilon_2)$ follows a distribution known up to a parameter σ: $F(\cdot, \cdot; \sigma)$. Then the model predicts the probabilities of $(0, 0)$ and $(1, 1)$: $g_{00}(a, \delta, \sigma)$ and $g_{11}(a, \delta, \sigma)$ and the upper bounds for the probabilities of $(0, 1)$ and $(1, 0)$: $g_{01}(a, \delta, \sigma)$ and $g_{10}(a, \delta, \sigma)$, where $a = (a_1, a_2)'$ and $\delta = (\delta_1, \delta_2)'$. The outcome probabilities $p_{00}, p_{11}, p_{01}, p_{10}$ are the first stage point identified parameters. In the second stage, the structural parameters (a, δ, σ) are identified by the equalities/inequalities:

$$\begin{aligned} g_{00}(a, \delta, \sigma) - p_{00} &= 0 \\ g_{11}(a, \delta, \sigma) - p_{11} &= 0 \\ g_{01}(a, \delta, \sigma) - p_{01} &\geq 0 \\ g_{10}(a, \delta, \sigma) - p_{10} &\geq 0. \end{aligned} \tag{1.3}$$

The equalities/inequalities in (1.3) do not fall immediately into our general framework because the inequalities involve the first-stage parameters. However, we can introduce a nuisance second stage parameter γ, add the restriction $\gamma = p_{01}$ and rewrite the inequalities to only involve $(a, \delta, \sigma, \gamma)$. Specifically, let $\beta = (p_{00}, p_{11}, p_{01}, p_{10})$, $\theta = (a, \delta, \sigma, \gamma)$ for a nuisance parameter $\gamma \in [0, 1]$,

$$\begin{aligned} g^e(\theta, \beta) &= \begin{pmatrix} g_{00}(a, \delta, \sigma) - p_{00} \\ g_{11}(a, \delta, \sigma) - p_{11} \\ \gamma - p_{01} \end{pmatrix}, \text{ and} \\ g^{ie}(\theta) &= \begin{pmatrix} g_{01}(a, \delta, \sigma) - \gamma \\ g_{10}(a, \delta, \sigma) - (1 - g_{00}(a, \delta, \sigma) - g_{11}(a, \delta, \sigma) - \gamma) \end{pmatrix}. \end{aligned} \tag{1.4}$$

Then the entry game model is written in the form of (1.1).

Allowing covariates is easy. We can simply estimate $p_{00}(x) \equiv \Pr(0, 0|x), ..., p_{10}(x) \equiv \Pr(1, 0|x)$ in the first stage either fully nonparametrically, or use some flexible parametric form. Then in the second stage, use $(p_{00}(x), p_{11}(x), p_{01}(x), p_{10}(x))$ in place of $(p_{00}, p_{11}, p_{01}, p_{10})$. The estimated $a(x)$ and $\delta(x)$ will be the monopoly profit and the competition effects conditional on x. Generalizing the example to a game with more than 2 players can be done following Ciliberto and Tamer (2009).

Example 1.2. (Deliberative voting model) Iaryczower, Shi and Shum (2012) estimate a committee voting model in which judges have the opportunity to communicate their private information before submitting their votes. In this model, the vector of probabilities of the different vote profiles $\vec{p}_v$ is identified from the first stage. In the second stage, given $\vec{p}_v$, the structural parameters, θ, describing the judges' preferences, information qualities and prior are identified through a finite number of incentive compatibility (IC) constraints of the judges, corresponding to g^{ie}, and the equilibrium conditions (EC) – corresponding to the equality constraints g^e – which match the equilibrium voting outcomes predicted by the model with the $\vec{p}_v$ estimated in the first-stage.

Example 1.3. (Discrete mixture model) Consider a structural model with discrete unobserved heterogeneity, where a (discrete) outcome variable y is drawn according to a known parametric mixture distribution $f(y|\theta, \eta)$ characterized by structural parameters σ and mixing parameter η. Assuming that y takes K distinct values, and η takes M distinct values, the model is given by the equality constraints

$$P(y = k) = \sum_{m=1}^{M} f(k|\sigma, \eta = m)p_m, \text{ for } k = 1, \dots, K; \quad \sum_{m=1}^{M} p_m = 1.$$

In this example, the observed probabilities $P(y = k), k = 1, \dots, K$ are our β, and $(\sigma, \vec{p}_\eta)$ is our θ where $\vec{p}_\eta = (p_1, ..., p_M)'$. Examples of such models are the entry game with multiple equilibria in Bajari, Hahn, Hong and Ridder (2011) and the structural nonlinear panel data models in Bonhomme (forthcoming).

2 Consistent Estimation

To define the estimated set, let

$$Q(\theta, \beta; W) = g^e(\theta, \beta)' W g^e(\theta, \beta), \tag{2.1}$$

where W is a positive definite matrix. Then it clear that

$$\Theta_0 = \arg\min_{\theta \in \Theta} Q(\theta, \beta_0; W) \ s.t. \ g^{ie}(\theta) \geq 0. \tag{2.2}$$

Let $\hat{W}$ be a consistent estimator of W. The sample analogue estimator of Θ_0 is defined as

$$\hat{\Theta}_n = \arg\min_{\theta \in \Theta} Q(\theta, \hat{\beta}, \hat{W})$$
$$s.t.\ g^{ie}(\theta) \geq 0. \tag{2.3}$$

In contrast with the set estimators proposed in Chernozhukov et al. (2007) and widely recognized by the literature, our set estimator closely resembles the point estimator in a traditional point identified model. The advantage of our estimator is two-fold: (1) it is never empty and (2) it does not rely on an arbitrarily chosen "level".

A new technique is developed to prove the consistency of our estimator. The basic idea is to define a correspondence from the space of β to that of θ so that $\hat{\Theta}_n$ is the correspondence evaluated at $\hat{\beta}$. Then, we establish the continuity of the correspondence with the help of an implicit correspondence lemma. We prove this lemma by generalizing the implicit function theorem.

The consistency result is summarized in the following theorem. The detailed proof of the theorem as well as the implicit correspondence lemma are deferred to the appendix. In the theorem, $cl(A)$ denotes the closure of set A and $int(A)$ denotes the interior of set A. Let $\Theta_{ie} = \{\theta \in \Theta : g^{ie}(\theta) \geq 0\}$.

Theorem 2.1. *Suppose that*

(1) $\hat{\beta} \to_p \beta_0$ *and* $\hat{W} \to_p W$ *as* $n \to \infty$ *for some positive definite matrix* W*;*

(2) $\mathcal{B}$ *and* Θ *are compact;*

(3) $g^e(\cdot, \beta)$ *is continuously differentiable on* Θ *for all* $\beta \in \mathcal{B}$*,* g^{ie} *is continuous on* Θ*; and* either

(4) $cl(int(\Theta_{ie}) \cap \Theta_0) = \Theta_0$*, and* $\partial g^e(\theta, \beta_0)/\partial \theta'$ *has full row rank for all* $\theta \in \Theta_0$*;* or

(4*) Θ_0 *is a singleton.*

Then

$$d_H(\hat{\Theta}_n, \Theta_0) := \sup_{\theta \in \hat{\Theta}_n} \inf_{\theta_0 \in \Theta_0} ||\theta - \theta_0|| + \sup_{\theta_0 \in \Theta_0} \inf_{\theta \in \hat{\Theta}_n} ||\theta - \theta_0|| \to_p 0.$$

Proof. The proof contains four steps which we sketch below. Detailed arguments for Step 1 and Step 2 are needed and are given in the appendix.

Step 1. Let $\hat{\theta}_n$ be an arbitrary point in $\hat{\Theta}_n$ and $\theta_n \in \arg\min_{\theta \in \Theta_0} ||\theta - \hat{\theta}_n||$. We show that $||\hat{\theta}_n - \theta_n|| \to_p 0$. This implies that

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

If Θ_0 is a singleton, the proof is finished. If Θ_0 is not a singleton, the following steps are needed.

STEP 2. Let $r(\beta,\alpha) = \{\theta \in \Theta_{ie} : g^e(\theta,\beta) = \alpha\}$. Then $r(\beta,\alpha)$ is a correspondence from $\mathcal{B}\times\mathcal{A}$ to Θ_{ie} defined by the implicit function $g^e(\theta,\beta)-\alpha = 0$, where $\mathcal{A}$ is a compact R^{d_2}-ball around the origin. We show that r restricted to $\{(\beta,\alpha) \in \mathcal{B}\times\mathcal{A} : r(\beta,\alpha) \neq \emptyset\}$ is both upper and lower hemi-continuous at $(\beta,\alpha) = (\beta_0, 0)$. In this step, we make use of an Implicit Correspondence Lemma mentioned above.

STEP 3. Let $\tilde{\Theta}_n = r(\hat{\beta}, g^e(\hat{\theta}_n, \hat{\beta}))$ for an arbitrary $\hat{\theta}_n \in \hat{\Theta}_n$. Then clearly $\tilde{\Theta}_n \subseteq \hat{\Theta}_n$. The continuity of g^e implies that $g^e(\hat{\theta}_n, \hat{\beta}) \to 0$. The continuity of r shown in Step 2 then implies that $d_H(\tilde{\Theta}_n, \Theta_0) \to_p 0$.

STEP 4. Because $\tilde{\Theta}_n \subseteq \hat{\Theta}_n$, Step 3 implies that $\sup_{\theta_0\in\Theta_0} \inf_{\theta\in\hat{\Theta}_n} ||\theta-\theta_0|| \to_p 0$. □

Example. (1.1 Cont.) In the entry game example, $\hat{\beta}$ consists of the empirical frequencies of the different entry outcomes observed in the data; that is, empirical estimates of p_{11}, p_{10}, p_{01} (with $p_{00} = 1-p_{11}-p_{10}-p_{01}$). The set $\mathcal{B} = \Delta^3$ is by definition compact. The compactness of Θ is a typical assumption maintained in most extremum estimation problems. The function $g^e(\cdot,\beta) = \begin{pmatrix} g_{00}(a,\delta,\sigma) - p_{00} \\ g_{11}(a,\delta,\sigma) - p_{11} \\ \gamma - p_{01} \end{pmatrix}$ is continuously differentiable in θ as long as F is a continuous distribution and is continuously differentiable in σ. The function $g^{ie}(\theta)$ is continuous under the same condition. The assumption that the first derivative $\partial g^e(\theta,\beta_0)/\partial\theta'$ has full row rank can be verified directly because g_{00} and g_{11} are known functions given F. The assumption $cl(int(\Theta_{ie}) \cap \Theta_0) = \Theta_0$ can be verified by numerical calculation. Specifically, given any β_0, one can compute Θ_{ie} and Θ_0. By varying β_0 in a reasonable range, one can assess the shape of Θ_{ie} and Θ_0 reasonably accurately.

Remark. (a) The condition $cl(int(\Theta_{ie})\cap\Theta_0) = \Theta_0$ is worth some discussion. The condition is restrictive in the sense that it rules out (1) the case that $int(\Theta_{ie}) = \emptyset$ and (2) the case that Θ_0 contains isolated points on the boundary of Θ_{ie}. The first case can often be accommodated by a slight modification of the proof, which we discuss below.

The second case, on the other hand, has more substantive implication and should be ruled out if ones objective is Hausdorff consistency of the minimizer set of $\hat{\Theta}_n$. To see why, we give a stylized example that falls into the second case and in which Hausdorff consistency of $\hat{\Theta}_n$ fails. Consider the two-stage model with $g^e(\theta,\beta) = \begin{pmatrix} \theta_1 - \theta_2 \\ 2+\beta-\theta_1-\theta_3 \end{pmatrix}$, $g^{ie}(\theta) = \begin{pmatrix} (\theta_1-1)^2-1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$ and $\Theta = [-B,B]^3$ for a large $B > 0$. Then $\Theta_{ie} = ([-B,0] \cup [2,B]) \times [0,B] \times [0,B]$. Suppose $\beta_0 = 0$; then $\Theta_0 = \{(0,0,2),(2,2,0)\}$. The identified set falls entirely on the boundary of Θ_{ie}. Let $\hat{\beta} = -1/n$. Clearly, $\hat{\beta}$ is a consistent estimator of β_0. For any $\hat{W} \to_p W$ with W positive definite, $Q(\theta_0, \hat{\beta}, \hat{W}) = 0$ is solved uniquely at $\theta_0 = (0, 0, 2 - 1/n)$. Thus, $\hat{\Theta}_n = \{(0, 0, 2 - 1/n)\}$, and $d_H(\hat{\Theta}_n, \Theta_0) \to 2\sqrt{3} > 0$ – $\hat{\Theta}_n$ is not consistent.

(b) The most common reason that $int(\Theta_{ie}) = \emptyset$ occurs is that there are pairs of inequality restrictions which imply an equality constraint. In this case, let $g^{ie*}(\theta)$ be the remaining inequality constraints after removing the pairs that imply an equality constraint and let $\Theta^*_{ie} = \{\theta \in \Theta : g^{ie}(\theta)\}$. Stack the equality constraints extracted from $g^{ie}(\theta) \geq 0$ to the original equality constraints to form the new equality constraint $g^{e*}(\theta, \beta) = 0$. Then, the theorem above holds with g^{ie} and g^e replaced by g^{ie*} and g^{e*} respectively and with Θ_{ie} replaced by Θ^*_{ie}. To show this, one can use the proof of the above theorem except with $r(\beta, \alpha)$ replaced by by $r^*(\beta, \alpha^*) = \{\theta \in \Theta^*_{ie} : g^{e*}(\theta, \beta) = \alpha^*\}$, $\mathcal{A}$ replaced by $\mathcal{A}^*$ – a compact $R^{d^*_2}$-ball around the origin where d^*_2 is the dimension of $g^{e*}(\theta, \beta)$.

(c) Next we discuss some connection of our consistency conditions with the existing literature. To begin, we note that the existing papers consider moment equality/inequality models which are, for the most part, more complicated than the models we consider here, and that the extra complication of these models may justify the stronger assumptions made in some of these papers. Andrews, Berry, and Jia's (2004) condition $cl(int(\Theta_0)) = \Theta_0$ implies our condition $cl(int(\Theta_{ie}) \cap \Theta_0) = \Theta_0$. Ours is weaker in that we allow for Θ_0 to have empty interior as long as it lies in $int(\Theta_{ie})$ while Andrews et al. (2004) do not. Our conditions are sufficient for the degeneracy condition in Chernozhukov et al. (2007) which requires the existence of a random set Θ_n on which $Q(\theta, \hat{\beta}, \hat{W}) - \inf_{\theta \in \Theta_{ie}} Q(\theta, \hat{\beta}, \hat{W})$ vanishes (meaning $= 0$) and $d_H(\Theta_n, \Theta_0) = o_p(1)$. Clearly, our $\hat{\Theta}_n$ is such a random set. Finally, our rank conditions are quite similar to those in Yildiz (2012) but other conditions are different and nonnested.

3 Confidence Set

To define the confidence set, we choose a specific weighting matrix $\hat{W}$:

$$\hat{W}^*(\theta) = \left[G(\theta, \hat{\beta})\hat{V}_\beta G(\theta, \hat{\beta})'\right]^{-1}, \tag{3.1}$$

where $G(\theta, \beta) = \partial g^e(\theta, \beta)/\partial \beta'$, $\hat{V}_\beta$ is a consistent estimator of the asymptotic variance of $\tau_n(\hat{\beta} - \beta)$, where τ_n is a normalizing sequence, e.g. $\tau_n = \sqrt{n}$. Define the confidence set to be

$$CS_n = \{\theta : g^{ie}(\theta) \geq 0, \tau_n^2 Q(\theta, \hat{\beta}; \hat{W}^*(\theta)) \leq \chi^2_{d_2}(1 - \alpha)\}, \tag{3.2}$$

where $\chi^2_{d_2}(1-\alpha)$ is the $1-\alpha$ quantile of the chi-squared distribution with d_2 degrees of freedom and $1-\alpha \in (0,1)$ is the confidence level.

The following theorem shows that CS_n covers each point in Θ_0 with probability approaching $1-\alpha$, and if $G(\theta,\beta_0)$ does not depend on θ given that $\theta \in \Theta_0$, CS_n also covers the whole identified set with probability approaching $1-\alpha$. We note that Theorem 3.1 does not inherit the assumptions made in Theorem 2.1.

Theorem 3.1. *Suppose that $\tau_n(\hat{\beta}-\beta) \to_d Z_\beta \sim N(0,V_\beta)$, $g^e(\theta,\beta)$ is continuously differentiable in β, $G(\theta,\beta)$ is continuous in θ and β, $G(\theta,\beta_0)V_\beta G(\theta,\beta_0)'$ is invertible for all $\theta \in \Theta$ and $\hat{V}_\beta \to_p V_\beta$. Also suppose that $\Theta \times \mathcal{B}$ is compact and g^e and g^{ie} are continuous in Θ. Then*

(a) $\liminf_{n\to\infty} \inf_{\theta\in\Theta_0} \Pr(\theta \in CS_n) = \limsup_{n\to\infty} \sup_{\theta\in\Theta_0} \Pr(\theta \in CS_n) = 1-\alpha$;

(b) in addition, the following condition (***) holds

$$G(\theta_1,\beta_0) = G(\theta_2,\beta_0) \quad \text{for all } \theta_1,\theta_2 \in \Theta_0 \qquad (***)$$

then $\lim_{n\to\infty} \Pr(\Theta_0 \subseteq CS_n) = 1-\alpha$.

Remark. (a) The additional assumption (***) for part (b) is immediately satisfied if θ and β are additively separable in g^e, as they are in all the previous examples. Additive separability is likely to hold in models in which the equality restrictions take the form of "matching" empirical frequencies to outcome probabilities predicted by the model, which is a common feature of all the examples above. The condition may also be satisfied when g^e are not additively separable, but can be rewritten into an additively separable form by taking a nonlinear (e.g. logarithmic) transformation. Of course, there are models in which this additional assumption is not satisfied; for these models, part (a) still holds and can be useful.

(b) The results given in the theorem are pointwise asymptotics. It is easy to strengthen it to uniform asymptotics over a space of data generating processes, with the expense of assuming uniform convergence of $\tau_n(\hat{\beta}-\beta)$ and $\hat{V}_\beta$ and a uniform lower bound on the minimum eigenvalue of $G(\theta,\beta_0)V_\beta G(\theta,\beta_0)'$. For briefness, we do not give the formal arguments, but only point the reader to the fact that the inequality $g^{ie}(\theta) \geq 0$ does not cause trouble in deriving the uniform asymptotic theory because it is purely deterministic.

Example. [1.1 Cont.] In the entry game example without covariates, $\tau_n = \sqrt{n}$ and $V_\beta = diag(\beta) - \beta\beta'$. The Jacobian matrix $G(\theta,\beta_0) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}$, and $G(\theta,\beta_0)V_\beta G(\theta,\beta_0)' =$

$\begin{pmatrix} p_{00}(1-p_{00}) & -p_{00}p_{11} & -p_{00}p_{01} \\ -p_{00}p_{11} & p_{11}(1-p_{11}) & -p_{11}p_{01} \\ -p_{00}p_{01} & -p_{11}p_{01} & p_{01} \end{pmatrix}$ is invertible as long as $p_{00}, p_{10}, p_{11} > 0$. The compactness of $\Theta \times \mathcal{B}$ and the continuity of g^e and g^{ie} is discussed in the previous section

If there are covariates, and $p_{00}(x), ..., p_{10}(x)$ are nonparametrically estimated in the first stage, τ_n typically is the nonparametric rate of convergence. V_β should also change accordingly. The rest of the verification remains the same. If $p_{00}(x), ..., p_{10}(x)$ are estimated using a parametric model, then τ_n still may be $\sqrt{n}$.

A Proofs

The proof of Theorem 2.1 makes uses of the following implicit correspondence lemma.

Lemma A.1 (Implicit Correspondence Lemma)**.** *Let $f(x,y) : \mathcal{X} \times \mathcal{Y} \to R^{d_f}$ be a continuously differentiable function defined on the set $\mathcal{X} \times \mathcal{Y} \subseteq R^{d_x+d_y}$, where $\mathcal{X}$ is open, $\mathcal{Y}$ is compact and $cl(int(\mathcal{Y})) = \mathcal{Y}$. Let the equation $f(x,y) = 0$ define the correspondence $y(x) : x \to y$ implicitly, i.e., $y(x) = \{y \in \mathcal{Y} : f(x,y) = 0\}$. Let $\mathcal{X}_1 = \{x \in \mathcal{X} : y(x) \neq \emptyset\}$. Consider a $x_0 \in \mathcal{X}_1$. Suppose furthermore that $\partial f(x_0,y_0)/\partial y'$ has full row-rank for any $y_0 \in y(x_0) \cap int(\mathcal{Y})$ and $cl(y(x_0) \cap int(\mathcal{Y})) = y(x_0)$. Then, the correspondence $y(x)$ restricted to $\mathcal{X}_1$ is continuous at x_0.*

Proof. First, we prove the upper hemicontinuity. Consider an arbitrary sequence $\{x_m \in \mathcal{X}_1\}_{m=1}^{\infty}$ such that $\lim_{m\to\infty} x_m = x_0$ and an arbitrary converging sequence $\{y_m \in y(x_m)\}_{m=1}^{\infty}$ such that $\lim_{m\to\infty} y_m = y_\infty$. Because $f(x,y)$ is a continuous function, we have $\lim_{m\to\infty} f(x_m, y_m) = f(x_0, y_\infty)$. By the definition of the sequence $\{y_m\}$, $f(x_m, y_m) = 0$ for any m. Thus, $f(x_0, y_\infty) = 0$, i.e. $y_\infty \in y(x_0)$. This combined with the compactness of $\mathcal{Y}$ (so that every sequence $\{y_m \in \mathcal{Y}\}$ has a converging subsequence) shows the upper hemicontinuity.

The lower hemicontinuity is trickier and we show it using a combination of the implicit function theorem and normalization of parameters. Again, consider an arbitrary sequence $\{x_m \in \mathcal{X}_1\}_{m=1}^{\infty}$ such that $\lim_{m\to\infty} x_m = x_0$ and an arbitrary point $y_0 \in y(x_0)$. The lower hemicontinuity is proved if we can find a sequence $\{y_m \in y(x_m)\}_{m=1}^{\infty}$ such $\lim_{n\to\infty} y_m = y_0$. We discuss two cases below: $y_0 \in int(\mathcal{Y})$ and $y_0 \notin int(\mathcal{Y})$.

Case 1. $y_0 \in int(\mathcal{Y})$. The fact that $\partial f(x_0,y_0)/\partial y'$ has full row-rank implies that $d_f \leq d_y$. If $d_f = d_y$, then $\partial f(x_0,y_0)/\partial y'$ is invertible. By the implicit function theorem (see e.g. Theorem 9.28 of Rudin (1976)), there exists an open set $U_x \subseteq \mathcal{X}$ containing x_0, an open set $U_y \subseteq int(\mathcal{Y})$ containing y_0 and a unique $y^*(x) \in U_y$ for every $x \in U_x$ such that $y^*(x) \in y(x)$. Also, $y^*(x)$ is a continuous function on U_x by the same theorem. Simply set $y_m = y^*(x_m)$ and we have $\lim_{m\to\infty} y_m = y_0$.

If $d_f < d_y$, one cannot apply the implicit function theorem directly. But observe that when $d_f < d_y$, y_0 is “underidentified” by the equation system $f(x_0, y) = 0$. We add a few normalization equations to force y_0 to be identified. Let E be a $(d_y - d_f) \times d_y$ dimensional matrix, each row of which is an element in the standard orthogonal basis $(e_1, ..., e_{d_y})$ and the rows are orthogonal to each other and orthogonal to the rows of $\partial f(x_0,y_0)/\partial y'$. Then, $[\partial f(x_0,y_0)'/\partial y | E']'$ is invertible. We add the following normalization equations to the original equation system:

$$E \times y = E \times y_0. \tag{A.1}$$

Let $\bar{f}(x,y) = \begin{pmatrix} f(x,y) \\ E \times (y - y_0) \end{pmatrix}$. Then $\bar{f}(x,y)$ is continuously differentiable and $\partial \bar{f}(x_0,y_0)/\partial y = [\partial f(x_0,y_0)'/\partial y | E']'$ is invertible. The arguments in the previous paragraph go through with f replaced by $\bar{f}$.

Case 2. $y_0 \notin int(\mathcal{Y})$. Because $cl(y(x_0) \cap int(\mathcal{Y})) = y(x_0)$, we can find a sequence $y_n \in y(x_0) \cap int(\mathcal{Y})$ such that $\lim_{n\to\infty} y_n = y_0$. For each y_n, we can find a sequence $y_{m,n} \in y(x_m)$ such that $\lim_{m\to\infty} y_{m,n} = y_n$ by arguments in Case 1. Let n_m be such that $|y_{m,n_m} - y_0| \leq \inf_n |y_{m,n} - y_0| + 2^{-m}$. We next show that $\lim_{m\to\infty} y_{m,n_m} = y_0$, which completes the proof of lower hemicontinuity. Consider an arbitrary $\epsilon > 0$, then there exists a N such that for all $n \geq N$, $|y_n - y_0| < \epsilon/3$. Since $\lim_{m\to\infty} y_{m,N} = y_N$, there exists M_1 such that for all $m \geq M_1$, such that $|y_{m,N} - y_N| < \epsilon/3$. Let M_2 be an integer such that for all $m \geq M_2$, $2^{-m} < \epsilon/3$. Then for any $m > \max\{M_1, M_2\}$, we have

$$\begin{aligned} |y_{m,n_m} - y_0| &\leq |y_{m,N} - y_0| + 2^{-m} \\ &\leq |y_{m,N} - y_N| + |y_N - y_0| + 2^{-m} \\ &\leq \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned} \tag{A.2}$$

This shows that $\lim_{n\to\infty} y_{m,n_m} = y_0$ and by definition, $y_{m,n_m} \in y(x_m)$.

Therefore, $y(x)$ is both lower and upper hemicontinuous at x_0. The lemma is proved. □

Proof of Theorem 2.1. Here, we provide the detailed arguments underlying steps (1) and (2), which were described in the main text following the statement of the theorem.

The proof for Step 1 takes the form of a standard consistency proof. Two major component of it is the uniform convergence of Q and global identification:

$$\begin{aligned} &\sup_{\theta\in\Theta: g^{ie}(\theta)\geq 0} |Q(\theta, \hat{\beta}; W_n) - Q(\theta, \beta_0; W)| \to_p 0, \textit{ and} \\ &\forall \epsilon, \exists \delta_\epsilon > 0 \; s.t. \inf_{\theta\in\Theta: \inf_{\theta_0\in\Theta_0} ||\theta-\theta_0||>\epsilon} Q(\theta, \beta_0; W) > \delta_\epsilon. \end{aligned} \tag{A.3}$$

The uniform convergence is implied by the continuity of g^{ie} on the compact set $\Theta \times \mathcal{B}$, $\hat{\beta} \to_p \beta_0$ and $W_n \to_p W$. The global identification condition is implied by the definition of Θ_0, the continuity of $Q(\cdot, \beta_0, W)$ and the compactness of Θ. Using those two results, we have for any $\epsilon > 0$,

$$\begin{aligned}
\Pr(||\hat{\theta}_n - \theta_n|| > \epsilon) &\leq \Pr(Q(\hat{\theta}_n, \beta_0; W) > \delta_\epsilon) \\
&= \Pr(Q(\hat{\theta}_n, \beta_0; W) - Q(\hat{\theta}_n, \hat{\beta}; \hat{W}) \\
&\quad + Q(\hat{\theta}_n, \hat{\beta}; \hat{W}) - Q(\theta_n, \hat{\beta}; \hat{W}) \\
&\quad + Q(\theta_n, \hat{\beta}; \hat{W}) - Q(\theta_n, \beta_0; W) > \delta_\epsilon) \\
&\leq \Pr(\sup_{\theta \in \Theta : g^{ie}(\theta) \geq 0} |Q(\theta, \beta_0; W) - Q(\theta, \hat{\beta}; \hat{W})| > \delta_\epsilon) \quad \to 0,
\end{aligned}$$

where the first inequality holds by the second result in (A.3), the equality holds by adding and subtracting terms and by $Q(\theta_n, \beta_0, W) = 0$, the second inequality holds by $Q(\hat{\theta}_n, \hat{\beta}, \hat{W}) \leq Q(\theta_n, \hat{\beta}, \hat{W})$ and the convergence holds by the first result in (A.3). Thus, the result of Step 1 is shown.

In step (2), θ corresponds to y in Lemma A.1, $\{\theta \in \Theta : g^{ie}(\theta) \geq 0\}$ corresponds to $\mathcal{Y}$, (β, α) corresponds to x, and an arbitrary open set containing $\mathcal{B} \times \mathcal{A}$ corresponds to $\mathcal{X}$, and $g^e(\theta, \beta) - \alpha$ corresponds to $f(x, y)$. The set $\{\theta \in \Theta : g^{ie}(\theta) \geq 0\}$ is compact because Θ is compact and g^{ie} is continuous. The function $g^e(\theta, \beta) - \alpha$ is continuously differentiable because g^e is continuously differentiable. The Jacobian $\partial(g^e(\theta, \beta) - \alpha)/\partial\theta' = \partial g^e(\theta, \beta)/\partial\theta'$ has full row-rank by assumption. Therefore, Lemma A.1 applies and shows that the correspondence $r : \{(\beta, \alpha) \in \mathcal{B} \times \mathcal{A} : r(\beta, \alpha) \neq \emptyset\} \to \mathcal{B}$ is continuous at $(\beta_0, 0)$. □

Proof of Theorem 3.1. (a) By the definition of inf, there exists a sequence $\{\theta_n \in \Theta_0\}$ with

$$\liminf_{n\to\infty} \inf_{\theta \in \Theta_0} \Pr(\theta \in CS_n) = \liminf_{n\to\infty} \Pr(\theta_n \in CS_n). \tag{A.4}$$

By the definition of lim inf, there exists a subsequence $\{u_n\}$ of $\{n\}$ such that

$$\liminf_{n\to\infty} \Pr(\theta_n \in CS_n) = \lim_{n\to\infty} \Pr(\theta_{u_n} \in CS_{u_n}). \tag{A.5}$$

Because Θ is compact, there is a further subsequence $\{a_n\}$ of $\{u_n\}$ such that $\theta_{a_n} \to \theta_0$ for some $\theta_0 \in \Theta$. Because g^e and g^{ie} are continuous in θ, $\theta_0 \in \Theta_0$. We then show that

$$\tau_{a_n} Q(\theta_{a_n}, \hat{\beta}; \hat{W}^*(\theta_{a_n})) \to_d \chi^2_{d_2}. \tag{A.6}$$

To show this observe that

$$
\begin{aligned}
&\tau_{a_n}^2 Q(\theta_{a_n},\hat{\beta};\hat{W}^*(\theta_{a_n})) \\
&= \tau_{a_n}^2 [g^e(\theta_{a_n},\beta_0) + G(\theta_{a_n},\tilde{\beta})(\hat{\beta}-\beta_0)]^{'} \hat{W}^*(\theta_{a_n})[g^e(\theta_{a_n},\beta_0) + G(\theta_{a_n},\tilde{\beta})(\hat{\beta}-\beta_0)] \\
&= \tau_{a_n}^2 (\hat{\beta}-\beta_0)^{'} G(\theta_{a_n},\tilde{\beta})^{'} \hat{W}^*(\theta_{a_n}) G(\theta_{a_n},\tilde{\beta})(\hat{\beta}-\beta_0), \qquad \text{(A.7)}
\end{aligned}
$$

where $\tilde{\beta}$ lies on the line-segment between β_0 and $\hat{\beta}$. Then $\tilde{\beta} \to_p 0$. By the continuity of G, we have $G(\theta_{a_n},\tilde{\beta}) \to_p G(\theta_0,\beta_0)$. Similarly, $G(\theta_{a_n},\hat{\beta}) \to_p G(\theta_0,\beta_0)$. Thus,

$$
[\hat{W}^*(\theta_{a_n})]^{-1} \equiv G(\theta_{a_n},\hat{\beta})\hat{V}_\beta G(\theta_{a_n},\hat{\beta})^{'} \to_p G(\theta_0,\beta_0)V_\beta G(\theta_0,\beta_0)^{'}. \qquad \text{(A.8)}
$$

By the invertibility of $G(\theta_0,\beta_0)^{'} V_\beta G(\theta_0,\beta_0)$,

$$
\hat{W}^*(\theta_{a_n}) \to_p [G(\theta_0,\beta_0)V_\beta G(\theta_0,\beta_0)^{'}]^{-1}. \qquad \text{(A.9)}
$$

Therefore,

$$
\tau_{a_n}^2 Q(\theta_{a_n},\hat{\beta},\hat{W}^*(\theta_{a_n})) \to_d Z_\beta^{'} G(\theta_0,\beta_0)^{'}[G(\theta_0,\beta_0)V_\beta G(\theta_0,\beta_0)^{'}]^{-1} G(\theta_0,\beta_0) Z_\beta \sim \chi^2_{d_2}. \qquad \text{(A.10)}
$$

This implies that $\lim_{n\to\infty} \Pr(\theta_{a_n} \in CS_{a_n}) = 1-\alpha$. Then by the definition of $\{\theta_{a_n}\}$ given at the beginning of the proof, we have

$$
\liminf_{n\to\infty} \inf_{\theta\in\Theta_0} \Pr(\theta \in CS_n) = 1-\alpha. \qquad \text{(A.11)}
$$

Analogous arguments can be used to show $\limsup_{n\to\infty} \sup_{\theta\in\Theta_0} \Pr(\theta \in CS_n) = 1-\alpha$.

(b) There exists a possibly random sequence $\{\theta_n \in \Theta_0\}$ such that

$$
\sup_{\theta\in\Theta_0} \tau_{a_n}^2 Q(\theta,\hat{\beta},\hat{W}^*(\theta)) = \tau_{a_n}^2 Q(\theta_n,\hat{\beta},\hat{W}^*(\theta_n)) + o_p(1). \qquad \text{(A.12)}
$$

Like in (A.7), we can write

$$
\tau_{a_n}^2 Q(\theta_n,\hat{\beta},\hat{W}^*(\theta_n)) = \tau_{a_n}^2 (\hat{\beta}-\beta_0)^{'} G(\theta_n,\tilde{\beta})^{'} \hat{W}^*(\theta_n) G(\theta_n,\tilde{\beta})(\hat{\beta}-\beta_0). \qquad \text{(A.13)}
$$

Because $G(\theta,\beta)$ is continuous on the compact space $\Theta\times\mathcal{B}$, $G(\theta,\beta)$ is uniformly continuous on $\Theta\times\mathcal{B}$. Thus,

$$
\sup_{\theta\in\Theta_0} ||G(\theta,\tilde{\beta}) - G(\theta,\beta_0)|| \to_p 0 \textit{ and } \sup_{\theta\in\Theta_0} ||G(\theta,\hat{\beta}) - G(\theta,\beta_0)|| \to_p 0. \qquad \text{(A.14)}
$$

Let θ_0 be an arbitrary point in Θ_0. By the additional assumption that $G(\theta_1, \beta_0) = G(\theta_2, \beta_0)$ for all $\theta_1, \theta_2 \in \Theta_0$, and (A.14), for any random sequence $\{\theta_n \in \Theta_0\}$,

$$[G(\theta_n, \tilde{\beta})' \hat{W}^*(\theta_n) G(\theta_n, \tilde{\beta})] \to_p G(\theta_0, \beta_0)' [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1} G(\theta_0, \beta_0). \tag{A.15}$$

Therefore,

$$\tau_{a_n}^2 Q(\theta_n, \hat{\beta}, \hat{W}^*(\theta_n)) \to_d Z_\beta' G(\theta_0, \beta_0)' [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1} G(\theta_0, \beta_0) Z_\beta \sim \chi^2_{d_2}. \tag{A.16}$$

Combining this with (A.12), we get

$$\begin{aligned}
\Pr(\hat{\Theta}_n \subseteq CS_n) &= \Pr(\sup_{\theta \in \Theta_0} \tau_{a_n}^2 Q(\theta, \hat{\beta}, \hat{W}^*(\theta)) \leq \chi^2_{d_2}(1-\alpha)) \\
&= \Pr(\tau_{a_n}^2 Q(\theta_n, \hat{\beta}, \hat{W}^*(\theta_n)) + o_p(1) \leq \chi^2_{d_2}(1-\alpha)) \\
&\to \Pr(\chi^2_{d_2} \leq \chi^2_{d_2}(1-\alpha)) = 1 - \alpha.
\end{aligned} \tag{A.17}$$

□

References

Andrews, Donald W. K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157.

___ **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequality Models," unpublished manuscript, Department of Economics, Yale University January 2009.

___ **, Steven Berry, and Panle Jia**, "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," unpublished manuscript, Department of Economics, Yale University 2004.

Bajari, Patrick, Jinyong Hahn, Han Hong, and Geert Ridder, "A Note on Semiparametric Estimation of Finite Mixture of Discrete Choice Models with Application to Game Theoretical Models," *International Economic Review*, 2011, *52*, 807–824.

Bonhomme, Stephane, "Functional Differencing," *Econometrica*, forthcoming.

Bugni, Federico A., "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 2010, *78*, 735–753.

Canay, Ivan A., "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 2010, *156*, 408–425.

Chernozhukov, Victor, Han Hong, and Elie Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 2007, *75*, 1243–1284.

____ **, Simon Lee, and Adam Rosen**, "Inference with Intersection Bounds," unpublished manuscript, Department of Economics, University College London 2008.

Ciliberto, Federico and Elie Tamer, "Market Structure and Multiple Equilibria in the Airline Industry," *Econometrica*, 2009, *77*, 1791–1828.

Iaryczower, Matias, Xiaoxia Shi, and Matthew Shum, "Can Deliberation Trump Conflict? Partial Identification of a Deliberative Voting Model," unpublished manuscript, Princeton University 2012.

Imbens, Guido and Charles F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72*, 1845–1857.

Romano, Joseph P. and Azeem M. Shaikh, "Inference for identifiable parameters in partially identified models," *Journal of Statistical Planning and Inference*, 2008, *(Special Issue in Honor of T. W. Anderson, Jr. on the Occasion of his 90th Birthday), 138*, 2786–2807.

____ **and** ____ , "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 2010, *78*, 169–211.

Rudin, Walter, *Principles of Mathematical Analysis*, 3 ed., McGraw-Hill Companies,Inc., 1976.

Stoye, Jörg, "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2010, *77*, 1299–1315.

Yildiz, Neşe, "Consistency of Plug-in Estimators of Upper Contour and Level Sets," *Econometric Theory*, 2012, *28*, 309–327.