Econometric Theory, **31**, 2015, 493–520.
doi:10.1017/S0266466614000425

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

Xiaoxia Shi
University of Wisconsin at Madison

Matthew Shum
California Institute of Technology

This paper 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. Besides, a technical contribution is an implicit correspondence lemma which generalizes the implicit function theorem to multivalued implicit maps.

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, Hong, Tamer, 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 procedures often require simulation of either the test statistic or the critical values, as well as 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 paper, 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

We thank Yanqin Fan, Patrik Guggenberger, Bruce E. Hansen, Jack R. Porter, the editor Peter C.B. Phillips, the co-editor, and two anonymous referees for useful comments and suggestions. Xiaoxia Shi acknowledges the financial support of the Wisconsin Alumni Research Foundation via the Graduate School Fall Competition Award. Address correspondence to Xiaoxia Shi, Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI, 53706; e-mail: xshi@ssc.wisc.edu or to Matthew Shum, Division of Humanities and Social Sciences, California Institute of Technology, MC 228-77, 1200 East California Blvd., Pasadena, CA 91125; e-mail: mshum@caltech.edu.

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 analog 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; and 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 includes entry games, voting games, and discrete mixture models, all of which have been of interest to applied researchers. These models are allowed to have continuous or discrete covariates as discussed in Section 4.2. But we note that continuous covariates that enter the model parametrically are generally not allowed in our framework.

The main contribution of this paper is 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 partial identification and inequality constraints. In addition, a technical contribution of this paper 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 to multivalued implicit maps. Both the new consistency proof technique and the ICL may be useful in more general models. In particular, the ICL contributes to a literature providing generalizations of the textbook implicit function theorem (e.g., Rudin, 1976, Thm. 9.28); see, for example, Zhang and Ge (2006) and Phillips (2012). The ICL here differs from the previous results in that it applies to the cases where neither the global or the local univalence of the implicit map is guaranteed to hold.

There are a small number of papers that address the consistent estimation problem under partial identification. These include Andrews, Berry, and Jia (2004), Chernozhukov et al. (2007), Yildiz (2012), and Kaido and Santos (2011). 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 models is much larger. For a current survey, see the introduction of Andrews and Shi (2013). Our benchmark confidence set is similar in spirit to that of Andrews and Soares (2010) applying to moment equality models. Our profiled confidence set where we concentrate out the nuisance parameter offers a new and simple approach to subvector inference in a certain type of partially identified models.

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 gives conditions under which confidence sets 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. Section 4 discusses how to profile out nuisance parameters and how to handle covariates in our framework. Section 5 reports Monte Carlo simulation results for the entry game model. 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:

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

where $g^{ie} : \Theta \to R^{d_1}$ and $g^e : \Theta \times \mathcal{B} \to R^{d_2}$ are known functions, and $\theta \in \Theta \subset R^{d_\theta}$ is the unknown parameter. 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.[4]

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. Introducing the slackness parameter does not affect consistent set estimation. However, for confidence set construction this can lead to conservative inference for the "real" structural parameters. We discuss this problem in more detail and provide a solution in Section 4.1.

The two-stage model includes several useful examples which have been studied in the empirical literature on partially identified models. We describe the first example—an entry game—in detail to illustrate the applicability our framework. The other three are described only briefly, for the purpose of space. We note that our two-stage model, in general, is not a special case of moment inequality models. The first three examples below are moment inequality models, but not the last example.

Example 1.1 (Entry game)
Following Andrews et al. (2004) and Ciliberto and Tamer (2009), consider a two-firm entry game with complete information, and allow only pure strategy equilibria. 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 nonpositive, 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 $(1,1)$: $p_{00} = g_{00}(a, \delta, \sigma)$ and $p_{11} = g_{11}(a, \delta, \sigma)$ and the upper bounds for the probabilities of $(0,1)$ and $(1,0)$: $p_{01} \leq g_{01}(a, \delta, \sigma)$ and $p_{10} \leq g_{10}(a, \delta, \sigma)$, 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:

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

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 $\gamma = g_{01}(a, \delta, \sigma) - p_{01}$, 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 $\gamma \in [0, 1]$,

$$g^e(\theta, \beta) = \begin{pmatrix} g_{00}(a, \delta, \sigma) - p_{00} \\ g_{11}(a, \delta, \sigma) - p_{11} \\ g_{01}(a, \delta, \sigma) - p_{01} - \gamma \end{pmatrix}, \text{ and}$$

$$g^{ie}(\theta) = \begin{pmatrix} \gamma \\ g_{10}(a, \delta, \sigma) + g_{00}(a, \delta, \sigma) + g_{11}(a, \delta, \sigma) + g_{01}(a, \delta, \sigma) - 1 - \gamma \end{pmatrix}. \qquad \textbf{(1.4)}$$

Then the entry game model is written in the form of (1.1).[5]

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 beliefs 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 $f(y|\sigma, \eta)$ 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

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

In this example, β corresponds to the probabilities $P(y = k), k = 1, \ldots, K$, which can be estimated from the data, while $\theta = (\sigma, \vec{p}_\eta)$ where $\vec{p}_\eta = (p_1, \ldots, p_M)'$. 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 (2012).

Example 1.4 (Dynamic game)
In the dynamic game model of Bajari, Benkard, and Levin (2007) (BBL), β denotes parameters in the policy function and the law of motion of the state variables, which are estimated using flexible parametric functions in the first stage. In the second stage the remaining structural parameters θ—typically parameters in agent's utility functions, or entry/exit costs—are related to the first stage parameters through an equation arising from the equilibrium optimality conditions. Written in our notation, they have one equality constraint: $g^e(\theta, \beta) = 0$, where

$$g^e(\theta, \beta) = \frac{\partial \int \min\{h(x, \theta, \beta), 0\}^2 dH(x)}{\partial \theta}, \tag{1.5}$$

where h and H are both known and can be simulated. There are no inequality constraints. Standard theory for minimum distance estimators apply if θ_0 is point identified, but our approach allows θ_0 to be partially identified.

Remark. Example 1.1 shows how one can convert models into our framework by introducing nuisance "slackness" parameters. We discuss this procedure in more generality here. Consider a model composed of the following restrictions: $g^e(\theta^s, \beta) = 0,\ g^{ie,1}(\theta^s, \beta) \geq 0,\ g^{ie,2}(\theta^s) \geq 0$. We can introduce a γ parameter of the same dimension as $g^{ie,1}$ and convert the model into $g^e(\theta^s, \beta) = 0$, $g^{ie,1}(\theta^s, \beta) - \gamma = 0,\ g^{ie,2}(\theta^s) \geq 0,\ \gamma \geq 0$. Typically, this is the only way to convert the model. But in some models, $g^{ie,1}(\theta^s, \beta)$ and $g^e(\theta^s, \beta)$ satisfy a relationship of the form $(g^e(\theta^s, \beta)', g^{ie,1}(\theta^s, \beta)')c = r(\theta^s)$ for some constant vector c, some function $r(\theta^s)$, and all θ^s and β. This causes $(g^e(\theta^s, \hat{\beta}_n)', g^{ie,1}(\theta^s, \hat{\beta}_n)' - \gamma')'$ to have a singular asymptotic covariance matrix, preventing the application of our confidence set proposed in Section 3. In this case, one should drop one coordinate in $g^{ie,1}(\theta^s, \hat{\beta}) - \gamma$ (and hence drop one slackness parameter), solve for the dropped coordinate of γ from $(0', \gamma')c = r(\theta^s)$, and add the nonnegativity of that coordinate as an inequality restriction. Indeed, Example 1.1 illustrates this case; as we saw there, it does not make a difference which coordinate to drop since dropping a different coordinate is equivalent to a reparametrization.

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

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

Our set estimator is the argmin set of a criterion function, and in this sense closely resembles the point estimator in a traditional point identified model. This estimator has two features: (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 $\hat{\beta}_n$. 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 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 the inequality-constraint parameter space of θ be $\Theta_{ie} = \{\theta \in \Theta : g^{ie}(\theta) \geq 0\}$.

THEOREM 2.1. *Suppose that*

(1) $\hat{\beta}_n \to_p \beta_0$ *and* $\hat{W}_n \to_p W_0$ *as* $n \to \infty$ *for some positive definite matrix* W_0*;*
(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

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

The proof of the theorem is given in the Appendix. For the rest of this section, we first illustrate how to apply the theorem to Example 1.1, and then discuss some of the important assumptions of the theorem.

Example (1.1 Cont.). In the entry game example, $\hat{\beta}_n$ consists of the empirical frequencies of the 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 $g^e(\cdot,\beta) = \begin{pmatrix} g_{00}(a,\delta,\sigma) - p_{00} \\ g_{11}(a,\delta,\sigma) - p_{11} \\ g_{01}(a,\delta,\sigma) - p_{01} - \gamma \end{pmatrix}$ 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}, g_{11}, and g_{01} 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.

Remarks on Theorem 2.1 (i) Condition (4) of the theorem: $cl(int(\Theta_{ie}) \cap \Theta_0) = \Theta_0$ is worth some discussion. The condition is not restrictive relative to the seemingly similar conditions in the literature (discussed in more detail below). It is guaranteed if either (i) Θ_0 lies in the interior of Θ_{ie} or (ii) $cl(int(\Theta_0)) = \Theta_0$, but neither condition is necessary. For example, Θ_0 can be a union of sets, some of which satisfy (i) and others satisfy (ii). In particular, Θ_0 does not need to have nonempty interior.

However condition (4) does rule out two cases: (a) $int(\Theta_{ie}) = \emptyset$ and (b) Θ_0 contains isolated points on the boundary of Θ_{ie}. Case (a) occurs usually because some pair of inequality restrictions imply an equality constraint and can be handled by a slight modification of our proofs. To do so, let $g^{ie*}(\theta)$ be the remaining inequality constraints after removing the pairs that imply equality constraints and let $\Theta^*_{ie} = \{\theta \in \Theta : g^{ie*}(\theta) \geq 0\}$. Append 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 holds with g^{ie} and g^e replaced by g^{ie*} and g^{e*} respectively and with Θ_{ie} replaced by Θ^*_{ie}.

Case (b), on the other hand, has more substantive implication and should be ruled out if one aims for Hausdorff consistency of the argmin set $\hat{\Theta}_n$. 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 $\hat{\beta}_n = -1/n$. Clearly, $\hat{\beta}_n$ is a consistent estimator of β_0. For any $\hat{W}_n \to_p W$ with W positive definite, $Q(\theta_0,\hat{\beta}_n;\hat{W}_n) = 0$ is solved uniquely at $\theta_0 = (0,0,2-1/n)$. Thus, $\hat{\Theta}_n = \{(0,0,2-1/n)\}$, and $d_H(\hat{\Theta}_n,\Theta_0) \to 2\sqrt{3} > 0$, implying that $\hat{\Theta}_n$ is not consistent.

(ii) Formally testing condition (4) is possible. To do so, one first obtains a confidence set CS_β for β_0; then, for each $\beta \in CS_\beta$, we compute $C(\beta, I_{d_2}) := \arg\min_{\theta\in\Theta_{ie}} Q(\theta,\beta;I_{d_2})$ and check whether the Hausdorff distance between $int(\Theta_{ie}) \cap C(\beta,I_{d_2})$ and $C(\beta,I_{d_2})$ is zero. One rejects the null that condition (4) is violated if for all $\beta \in CS_\beta$, we have $d_H(int(\Theta_{ie})\cap C(\beta,I_{d_2}), C(\beta,I_{d_2})) = 0$. The resulting test has the same significance level as one minus the confidence level of CS_β.

If condition (4) does not hold, our argmin set estimator can be inconsistent. To retain consistency, one can follow Chernozhukov et al. (2007) to define an extended set estimator: $\hat{\Theta}_n^{\varepsilon_n} := \{\theta \in \Theta_{ie} : Q(\theta,\hat{\beta}_n;\hat{W}_n) \le \varepsilon_n\}$, where ε_n is a sequence of tuning parameters that satisfies $\tau_n^{-2}\varepsilon_n^{-1} + \varepsilon_n \to 0$ for τ_n defined in the next section. We will not discuss this well-known approach in this paper.

(iii) 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 et al.'s (2004) condition $cl(int(\Theta_0)) = \Theta_0$ implies our condition $cl(int(\Theta_{ie})\cap\Theta_0) = \Theta_0$. Kaido and Santos's (2011) conditions imply those of Andrews et al. (2004). Our conditions are weaker in that we allow Θ_0 to have empty interior as long as Θ_0 lies in $int(\Theta_{ie})$. Our conditions are lower level sufficient conditions for the degeneracy condition in Chernozhukov et al. (2007) which requires the existence of a random set Θ_n on which $Q(\theta,\hat{\beta}_n,\hat{W}_n) - \inf_{\theta\in\Theta_{ie}} Q(\theta,\hat{\beta}_n,\hat{W}_n) = 0$ 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 similar to those in Yildiz (2012) but other conditions are different and nonnested.

(iv) The full row-rank part of condition (4) is similar to the full-rank condition for global univalence of implicit maps (see e.g., Phillips, 2012), but is weaker than the latter because the number of rows of the Jacobian matrix here is allowed to be smaller than the numbers of columns. In typical applications of our approach, the number of equalities (i.e., number of rows) indeed are smaller than the dimension of the parameter (i.e., number of columns), in which case the global univalence does not hold, and the implicit function theorems for univalued implicit maps do not apply.

3. CONFIDENCE SET

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

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

where $G(\theta, \beta) = \partial g^e(\theta, \beta)/\partial\beta'$ and $\hat{V}_{\beta,n}$ is a consistent estimator of the asymptotic variance, V_β, of $\tau_n(\hat{\beta}_n - \beta_0)$, where τ_n is a normalizing sequence, e.g., $\tau_n = \sqrt{n}$. Let the confidence set be

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

where $\chi^2_{d_2}(1-\alpha)$ is the $1-\alpha$ quantile of the χ^2 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 $\tau_n(\hat{\beta}_n - \beta_0) \to_d Z_\beta \sim N(0, V_\beta)$, $g^e(\theta, \beta)$ is continuously differentiable in β in $\mathcal{B}$ for all $\theta \in \Theta$, $G(\theta, \beta)$ is continuous in θ for all β in $\mathcal{B}$, $G(\theta, \beta_0)V_\beta G(\theta, \beta_0)'$ is invertible for all $\theta \in \Theta$ and $\hat{V}_\beta \to_p V_\beta$. Also suppose that Θ is compact and g^e and g^{ie} are continuous in θ in Θ for all β in $\mathcal{B}$. 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, if the following condition* (***) *holds*

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

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

Remark. (i) The additional assumption (***) for part (b) is immediately satisfied if θ and β are additively separable in g^e. 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 Examples 1.1–1.3.[6] Of course, there are models in which this additional assumption is not satisfied; for these models, part (a) still holds and can be useful.

(ii) Part (a) shows that our confidence sets have asymptotic correct coverage probability uniformly over points in the identified set. It is straightforward to strengthen this to uniform validity over a space of data generating processes, at the cost of assuming uniform convergence of $\tau_n(\hat{\beta}_n - \beta_0)$ and $\hat{V}_{\beta,n}$ and a uniform lower bound on the minimum eigenvalue of $G(\theta, \beta_0)V_\beta G(\theta, \beta_0)'$. This is because both types of uniformity can be established using the same sort of local asymptotic arguments and address the same sort of discontinuity of the pointwise asymptotic distribution of the test statistic.[7] For briefness, we do not give the formal arguments.

Example (1.1 Cont.). In the entry game example without covariates, $\tau_n = \sqrt{n}$ and $V_\beta = diag(\beta) - \beta\beta'$. The 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} are discussed in the previous section.

4. EXTENSIONS AND DISCUSSIONS

In this section, we discuss some practical issues with the methods proposed above.

4.1. Concentrating out Nuisance Parameters

Both the argmin set estimator and the confidence set proposed above are designed for the full vector θ_0. An immediate question is: what if only a subvector of θ_0, say θ_0^s, is of interest to the applied researcher? The answer depends on the objective and the nuisance parameter.

If the objective is to obtain a consistent estimator for the identified set of θ_0^s, then one only needs to obtain the projection of the argmin set $\hat{\Theta}_n$ to the space of θ_0^s. This projection is a good consistent set estimator, in the sense that it does not involve an arbitrarily chosen tuning parameter.

If the objective is to obtain a confidence set for either the true value of the subvector or the identified set of it, one can obtain the projection of CS_n to the space of θ_0^s. The projection, denoted $CS_n^{s,proj}$, will inherit the properties of CS_n described in the "lim sup" part of Theorem 3.1(a) and in 3.1(b), except with "$= 1-\alpha$" replaced by "$\geq 1-\alpha$". The "$\geq$" means that the $CS_n^{s,proj}$ may be conservative.

On the other hand, obtaining a nonconservative confidence set for the subvector is a more complicated problem. A general treatment is beyond the scope of this paper.[8] However, simplifications arise in the case (as in Example 1.1) in which the nuisance parameters are the slackness parameters γ which we introduced earlier in order to transform a model into the form of (1.1). The solution is a new profiled confidence set which we now describe.

The new confidence set is defined for the parameter θ^s for the following model

$$\begin{aligned} g^e(\theta^s,\beta) &= 0 \\ g^{ie,1}(\theta^s,\beta) - \gamma &= 0 \\ g^{ie,2}(\theta^s,\gamma) \geq 0 \text{ and } \gamma &\in R_+^{d_\gamma}, \end{aligned} \tag{4.1}$$

where $g^e(\theta^s,\beta)$, $g^{ie,1}(\theta^s,\beta)$, and $g^{ie,2}(\theta^s,\gamma)$ are respectively R^{d_e}, R^{d_γ}, and $R^{d_{ie,2}}$-valued known functions, β is the parameter that is estimated in the first step, and γ is the nuisance slackness parameter. Let Θ^s be the parameter space of θ^s. For any $\theta^s \in \Theta^s$, let the inequality-restricted parameter space of γ: $\Gamma(\theta^s) = \{\gamma \in R_+^{d_\gamma} : g^{ie,2}(\theta^s,\gamma) \geq 0\}$ and let the inequality-restriced parameter space of θ: $\Theta_{ie}^s = \{\theta^s \in \Theta^s : \Gamma(\theta^s) \neq \emptyset\}$. Let the identified set of θ^s be $\Theta_0^s = \{\theta^s \in \Theta^s : g^e(\theta^s,\beta_0) = 0 \text{ and } g^{ie,1}(\theta^s,\beta_0) \in \Gamma(\theta^s)\}$.

The new confidence set is based on the following criterion function

$$Q(\theta^s,\hat\beta_n) = \min_{\gamma\in\Gamma(\theta^s)} \begin{pmatrix} g^e(\theta^s,\hat\beta_n) \\ g^{ie,1}(\theta^s,\hat\beta_n)-\gamma \end{pmatrix}' \hat W_n^*(\theta^s) \begin{pmatrix} g^e(\theta^s,\hat\beta_n) \\ g^{ie,1}(\theta^s,\hat\beta_n)-\gamma \end{pmatrix}, \quad \textbf{(4.2)}$$

where the weight matrix $\hat W_n^*(\theta^s) = [G^s(\theta^s,\hat\beta_n)\hat V_{\beta,n} G^s(\theta^s,\hat\beta_n)']^{-1}$, with $\hat V_{\beta,n}$ defined as in the previous section and

$$G^s(\theta^s,\beta) = \frac{\partial}{\partial\beta'} \begin{pmatrix} g^e(\theta^s,\beta) \\ g^{ie,1}(\theta^s,\beta) \end{pmatrix}. \quad \textbf{(4.3)}$$

The critical value, $c_n(\theta^s, 1-\alpha)$, is the $1-\alpha$ quantile of the following random variable

$$J_n(\theta^s) := \min_{t\in\Gamma(\theta^s)-\hat\gamma(\theta^s)} \begin{pmatrix} Z_n^e \\ Z_n^{ie}-\kappa_n^{-1}\tau_n t \end{pmatrix}' \hat W_n^*(\theta^s) \begin{pmatrix} Z_n^e \\ Z_n^{ie}-\kappa_n^{-1}\tau_n t \end{pmatrix}, \quad \textbf{(4.4)}$$

where the random vector $(Z_n^{e,\prime}, Z_n^{ie,\prime})' \sim N(0,\hat W_n^*(\theta^s)^{-1})$, $\hat\gamma(\theta^s) = \arg\min_{\gamma\in\Gamma(\theta^s)} \|\gamma - g^{ie,1}(\theta^s,\hat\beta_n)\|$, and $\{\kappa_n\}$ is a sequence of tuning parameters that diverges to ∞.

The following theorem shows that the confidence set $CS_n^s = \{\theta^s \in \Theta_{ie}^s : \tau_n^2 Q(\theta^s,\hat\beta_n) \leq c_n(\theta^s,1-\alpha)\}$ asymptotically covers the true value of θ^s with probability no less than $1-\alpha$, and under additional assumptions, equal to $1-\alpha$.

THEOREM 4.1. *Suppose that* $\tau_n(\hat\beta_n - \beta_0) \to_d Z_\beta \sim N(0,V_\beta)$, $G^s(\theta^s,\beta)$ *in* (4.3) *is well defined and continuous in* θ^s *and* β *for* β *in* $\mathcal{B}$, $G^s(\theta^s,\beta_0)V_\beta G^s(\theta^s,\beta_0)'$ *is invertible for all* $\theta^s \in \Theta^s$ *and* $\hat V_{\beta,n} \to_p V_\beta$. *Also suppose that* Θ^s *is compact,* g^e, $g^{ie,1}$ *are continuous in* $\theta^s \in \Theta^s$ *for all* β *in* $\mathcal{B}$ *and* $g^{ie,2}$ *is continuous in* θ *and* γ. *Lastly assume that* $\Gamma(\theta^s)$ *is convex for all* $\theta^s \in \Theta_0^s$ *and that* $d_e \geq 1$. *Then*

(a) *if* $\kappa_n \to \infty$, *then* $\liminf_{n\to\infty} \inf_{\theta^s\in\Theta_0^s} \Pr(\theta^s \in CS_n^s) \geq 1-\alpha$;

(b) *if in addition,* $\kappa_n^{-1}\tau_n \to \infty$, *then we have* $\lim_{n\to\infty} \inf_{\theta^s\in\Theta_0^s} \Pr(\theta^s \in CS_n^s) = 1-\alpha$.

Remark. (i) The special condition imposed for this theorem (compared to Theorem 3.1) is the convexity of $\Gamma(\theta^s)$ and $d_e \geq 1$. This condition guarantees that $J_n(\theta^s)$ converges to a distribution that first order stochastically dominates the asymptotic distribution of $Q(\theta^s, \hat{\beta}_n)$, which is important for establishing the lower bound for the limiting coverage probability. The convexity of $\Gamma(\theta^s)$ is typically satisfied if one follows the guideline in the remark at the end of Section 1 to transform a raw model into our framework. The $d_e \geq 1$ condition is important for $J_n(\theta^s)$ to have continuous asymptotic distribution.

(ii) The new profiled confidence set CS_n^s is always weakly smaller than $CS_n^{s,proj}$, the projection of CS_n to the space of θ^s. This can be seen easily by first observing that

$$CS_n^{s,proj} = \{\theta^s \in \Theta_{ie}^s : \tau_n^2 Q(\theta^s, \hat{\beta}_n) \leq \chi^2_{d_e+d_\gamma}(1-\alpha)\}. \quad \textbf{(4.5)}$$

The critical value $\chi^2_{d_e+d_\gamma}(1-\alpha)$ is weakly bigger than $c_n(\theta^s, 1-\alpha)$ because $0 \in \Gamma(\theta^s) - \hat{\gamma}_n(\theta^s)$ and the random variable following "min" in (4.4) is $\chi^2_{d_e+d_\gamma}$ distributed when $t = 0$. Thus, $CS_n^s \subseteq CS_n^{s,proj}$. The profiling approach shrinks the confidence set because it uses a critical value that is adaptive to the size of each element of γ, rather than a larger χ^2 bound. The cost of course is that one needs to choose κ_n. A smaller κ_n shrinks the confidence set more because $c_n(\theta^s, 1-\alpha)$ is monotonically increasing in κ_n. Yet, one cannot use too small a κ_n because κ_n^{-1} must be small enough to dominate an asymptotically Gaussian term to guarantee good coverage probability. We use $\kappa_n = \sqrt{\log(n)}$ in the Monte Carlo in Section 5 and it performs well.

(iii) Remark (ii) on Theorem 3.1 applies here too, except the content in the footnote therein.

4.2. Introducing Covariates

Up to now, we have not considered models with covariates. However, this framework is broad enough to accommodate covariates in several ways.

The framework can allow any covariate to enter fully nonparametrically. Consider a covariate X with support $\mathcal{X}$. Suppose that in the first stage, the function $\beta(x) : \mathcal{X} \to \mathcal{B}$ can be estimated, and that the structural parameter, $\theta(x) : \mathcal{X} \to \Theta$, satisfies:

$$g^e(\theta(x), \beta(x)) = 0, \text{ and } g^{ie}(\theta(x)) \geq 0, \forall x \in \mathcal{X}. \quad \textbf{(4.6)}$$

Then the argmin set estimator and the confidence set proposed above apply pointwise for the function $\theta(x)$. This fully nonparametric approach cannot make use of any shape restriction on $\theta(x)$ directly, although shape restrictions can be imposed on $\beta(x)$ in the first stage estimation.

The framework can allow covariates with finite support points to enter either nonparametrically or parametrically. Consider a covariate X with finite support set $\mathcal{X} = \{x_1, \ldots, x_m\}$, where $m < \infty$. Again, assume that the first stage function $\beta(x) : \mathcal{X} \to \mathcal{B}$ can be estimated; for $\theta(x)$, we assume a parametric functional form; for instance, we may have $\theta(x) = \theta(x, \lambda)$, where

$\lambda \in \Lambda \subseteq R^{d_\lambda}$ and $\theta(\cdot,\cdot)$ is known. Let $\vec{\beta} = (\beta(x_1)', \ldots, \beta(x_m)')'$, $g^e(\lambda, \vec{\beta}) = (g^e(\theta(x_1,\lambda), \beta(x_1)), \ldots, g^e(\theta(x_k,\lambda), \beta(x_k)))'$, and $g^{ie}(\lambda, \vec{\beta}) = (g^{ie}(\theta(x_1,\lambda)), \ldots, g^{ie}(\theta(x_k,\lambda)))'$. Then we have

$$g^e(\lambda, \vec{\beta}) = 0 \text{ and } g^{ie}(\lambda) \geq 0. \tag{4.7}$$

Then the set estimator and the confidence sets proposed above can be used for λ.[9]

Example (1.1 Cont). We illustrate the parametric approach of allowing for covariates in the context of the entry game example. Suppose that X is some market and/or firm characteristics on which the profit function depends. Suppose that X takes m values. A plausible model assumes that $a_j(x) = x\lambda_j$, $\delta_j(x) = \delta_j$, and $\sigma_j(x) = \sigma_j$, that is, the monopoly profit depends on X linearly and the competition effects and the error variances are constants. Let $p_{00}(x) = \Pr(0,0|X = x), \ldots, p_{10}(x) = \Pr(1,0|X = x)$ be the conditional probabilities of each entry outcome. These conditional probabilities can be estimated using their sample counterparts. Then the model falls into the framework of (4.7) with $\lambda = (\lambda_1', \lambda_2', \delta_1, \delta_2, \sigma_1, \sigma_2, \gamma(x_1), \ldots, \gamma(x_m))'$ and $\vec{\beta} = (p_{00}(x_1), \ldots, p_{10}(x_1), \ldots, p_{00}(x_m), \ldots, p_{10}(x_m))'$.

5. MONTE CARLO SIMULATION

In this section, we present Monte Carlo results for the entry game example to illustrate the performance of our argmin set estimator and confidence sets. Our results show that (i) the Hausdorff distance between the argmin set estimator and the identified set declines at an encouraging speed as the sample size increases and (ii) the confidence sets have good coverage probabilities as the theory predicts. Because the entry game example is also a moment inequality/equality model, an alternative to ours is the Andrews and Soares (2010) (hereafter "AS") confidence sets. We compare the coverage probability and false coverage probability of our confidence sets to the AS ones and find that the performance of ours is competitive.

We consider the entry game described in Example 1.1. Let $(\varepsilon_1, \varepsilon_2) \sim N(0, I_2)$ so that $F(e_1, e_2) = \Phi(e_1)\Phi(e_2)$. The full set of model parameters are $\theta = (\theta^{s,\prime}, \gamma)'$, but the structural parameters of interest are just $\theta^s := (a_1, a_2, \delta_1, \delta_2)'$. We write out the functions $g_{00}, g_{11}, g_{01}, g_{10}$ as follows:

$$\begin{aligned}
g_{00}(a,\delta) &= (1 - \Phi(a_1))(1 - \Phi(a_2)) \\
g_{11}(a,\delta) &= \Phi(a_1 + \delta_1)\Phi(a_2 + \delta_2) \\
g_{10}(a,\delta) &= \Phi(a_1)(1 - \Phi(a_2 + \delta_2)) \\
g_{01}(a,\delta) &= \Phi(a_2)(1 - \Phi(a_1 + \delta_1)).
\end{aligned} \tag{5.1}$$

The two-step model to estimate is (1.4) with $g_{00}, g_{11}, g_{10}, g_{01}$ in the above display.

The first stage estimators $\hat{p}_{00}, \hat{p}_{11}, \hat{p}_{10}, \hat{p}_{01}$ are frequencies of the equilibrium outcomes for sample size n. For both the consistent estimation and the confidence sets, we use the weight matrix $W = \hat{W}_n^* = [\hat{V}_{\beta,n}]^{-1}$, where

$$\hat{V}_{\beta,n} = \begin{pmatrix} \hat{p}_{00}(1-\hat{p}_{00}) & -\hat{p}_{00}\hat{p}_{11} & -\hat{p}_{00}\hat{p}_{01} \\ -\hat{p}_{00}\hat{p}_{11} & \hat{p}_{11}(1-\hat{p}_{11}) & -\hat{p}_{11}\hat{p}_{01} \\ -\hat{p}_{00}\hat{p}_{01} & -\hat{p}_{11}\hat{p}_{01} & (1-\hat{p}_{01})\hat{p}_{01} \end{pmatrix}.$$

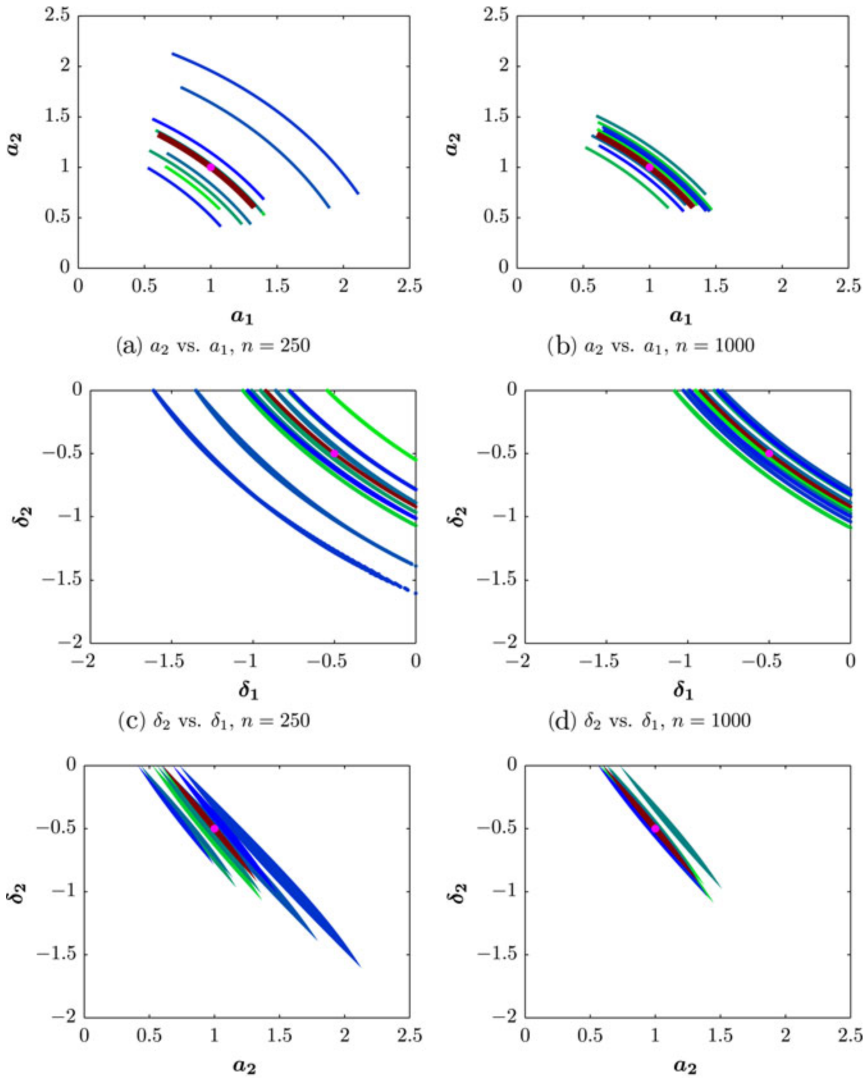
In generating the sample, we set the parameters $(a_{10}, a_{20}) = (1,1)$ and $(\delta_{10}, \delta_{20}) = (-0.5, -0.5)$. In case of multiple equilibria, the equilibrium that maximizes the joint profit is selected.

In Figure 1, we plot the two-dimensional projections for the identified set (IDset) and various set estimators for the parameters $(\alpha_{10}, \alpha_{20}, \delta_{10}, \delta_{20})$, for sample sizes $n = 250$ and $n = 1000$. We see that (1) the identified set does not have nonempty interior because the projection onto the a_1-a_2 space is a line, (2) the estimated sets have similar shape as the identified set, and (3) as the sample size increases from 250 to 1000, the estimated sets become notably more concentrated around the identified set.

To formally check the consistency of our set estimators, we compute the Hausdorff distance between the estimated set and the identified set for 5000 samples and report several measures of closeness: the median Hausdorff distance (HD) between the estimated set and the identified set, the 90% quantile of HD, and the probability that the HD exceeds some fixed levels. Table 1 shows the results. It is easy to see that the HD decreases both at the median and the 90% quantile as the sample size n grows. The last three columns directly confirm the consistency result of Theorem 2.1.[10]

Next, we investigate the finite sample performance of the confidence sets. Figure 2 shows the two-dimensional projections of one finite sample realization of CS_n and the profiled confidence set CS_n^s, as well as the realization in the same sample of AS's SUM/PA and SUM/GMS confidence sets.[11] As the figure shows, all four confidence sets have very similar shape and cover the identified set. AS's GMS confidence set is slightly smaller than their PA version; our profile confidence set CS_n^s is smaller than our projection confidence set $CS_n^{s,proj}$. In the sample with $n = 250$, our confidence sets are strictly smaller than the AS ones. But this is not generally true. For example, in the sample with $n = 1000$, our confidence sets are nonnested with the AS ones.

To formally check the finite sample coverage properties of the confidence sets, we compute various types of coverage probabilities using 5000 Monte Carlo repetitions. The first type of coverage probability is that of the true parameter $\theta_0^s = (1, 1, -0.5, -0.5)'$. We abbreviate this coverage probability by "CP-Point". The second type of coverage probability is that of parameter values outside the identified set. We consider two such points: $\theta_1^s = (0.85, 1.15, -0.5, -0.5)'$ and $\theta_2^s = (1.18, 0.60, -0.92, 0.00)'$, both approximately 0.14 units away from the identified set.[12] We abbreviate these false coverage probabilities by "FCP-θ_1^s" and "FCP-θ_2^s", respectively. We compute the first two types of coverage probabilities for $CS_n^{s,proj}$, CS_n^s, AS-SUM/PA, and AS-SUM/GMS. The third type of coverage



(a) a_2 vs. a_1, $n = 250$ (b) a_2 vs. a_1, $n = 1000$

(c) δ_2 vs. δ_1, $n = 250$ (d) δ_2 vs. δ_1, $n = 1000$

(e) δ_2 vs. a_2, $n = 250$ (f) δ_2 vs. a_2, $n = 1000$

FIGURE 1. Two-dimensional projections of the identified set and estimated set (dark red line/area: identified set; pink dot: true parameter; blueish-greenish lines/areas: estimated sets from 10 random samples).

probability is that of the identified set. We denote it by "CP-IDset" and compute it for $CS_n^{s,proj}$ only because this is the only confidence set for θ^s that is designed to cover the identified set with a prespecified probability.

For our profiling confidence set CS_n^s, the baseline tuning parameter value κ_n is chosen to be $\sqrt{\log n}$, a choice adopted from Andrews and Soares (2010).

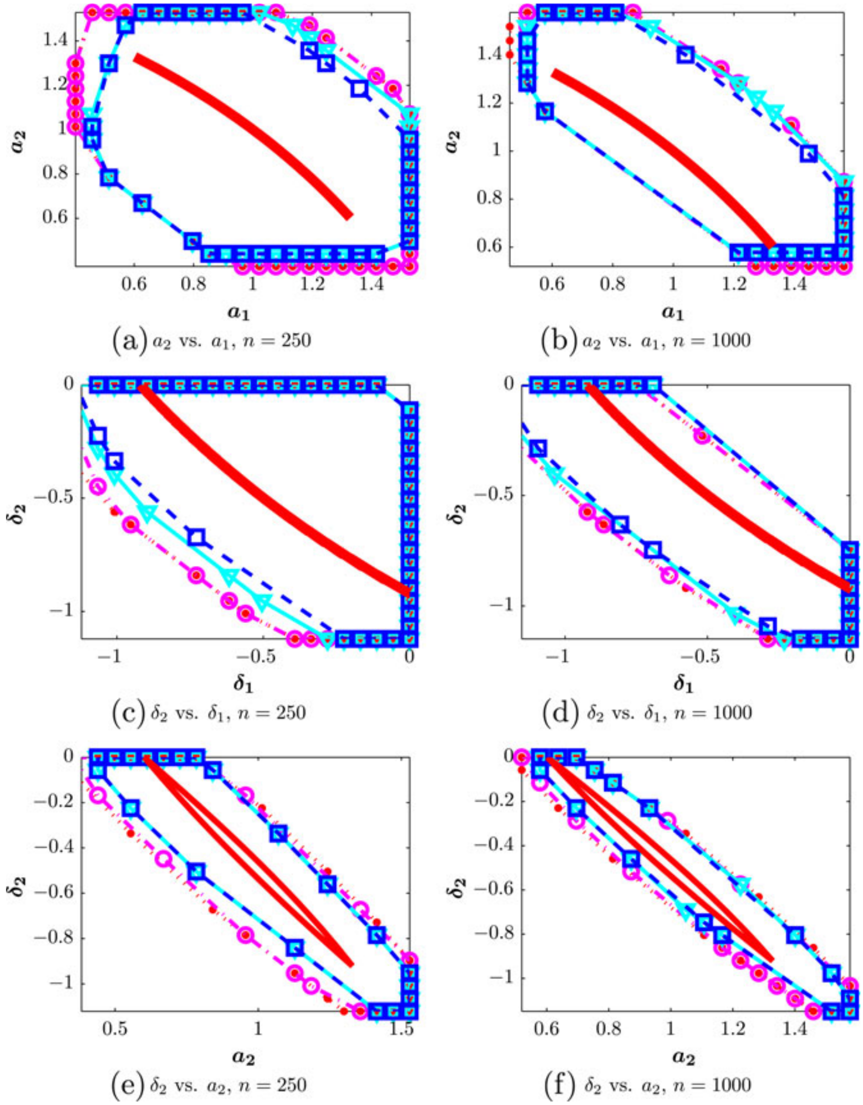


FIGURE 2. Two-dimensional projections of the confidence sets. (Area enclosed by the dotted line with solid dot markers: AS-PA; area enclosed by the dash-dotted line with circle markers: AS-GMS; area enclosed by the solid line with triangle markers: CS_n; area enclosed by the dashed line with square markers: CS_n^s; area enclosed by the thick solid line with no marker: Θ_0.).

In addition, we also report results for two other κ_n values which are respectively half and twice the baseline value. The purpose of this is to get some idea about how sensitive our approach is to different tuning parameter choices.

TABLE 1. Consistency of the set estimators

n	Median HD	90% quantile	Pr(HD > 0.5)	Pr(HD > 0.4)	Pr(HD > 0.3)
250	.269	.559	.141	.250	.424
500	.194	.397	.041	.098	.224
750	.162	.324	.010	.038	.134
1,000	.149	.283	.002	.018	.079
10,000	.084	.128	.000	.000	.000

TABLE 2. Coverage probability of true parameter value ($1-\alpha=90\%$)

		$n=250$	$n=500$	$n=750$	$n=1000$
CP-Point	$CS_n^{s,proj}$	.905	.923	.938	.941
	$CS_n^s:\kappa_n=\sqrt{logn}$	.894	.912	.919	.922
	$CS_n^s:\kappa_n=\frac{\sqrt{logn}}{2}$	.889	.901	.914	.912
	$CS_n^s:\kappa_n=2\sqrt{\log n}$	.899	.917	.928	.930
	AS-PA	.904	.921	.938	.936
	AS-GMS	.898	.915	.934	.931
CP-IDset	$CS_n^{s,proj}$	.872	.877	.886	.890

The results are reported in Tables 2 and 3. As Table 2 shows, the CP-Points are close to or bigger than the nominal level (90%) for confidence sets considered. Our $CS_n^{s,proj}$ also covers the identified set with probability close to 90%. For our profiling confidence set, the coverage probability of the true parameter value changes slightly with different choices of κ_n. As predicted by theory, smaller κ_n reduces the CP, while larger κ_n increases it.

In terms of FCPs, our profiled confidence set CS_n^s is slightly better than the projection $CS_n^{s,proj}$ for θ_1^s and is about the same as the latter for θ_2^s.[13] For AS's confidence set, the GMS version strictly improves upon the PA version uniformly across all cases. For our profiling confidence set, the effect of the tuning parameter choice on FCP is similar to that on the CP of the true parameter value.

It is notable that both of our confidence sets have uniformly lower (better) FCPs than the AS confidence sets across the two points and across sample sizes. This implies that introducing the slackness parameter γ to fit the moment inequality/equality model into our framework does not hurt the power, at least not in this example. It appears that the contrary is true. One intuition for the finding is that writing the model into our framework allows us to use the inverse of the covariance matrix as the weight matrix, and such a weight matrix works better than the diagonal weight matrix used in the AS SUM test statistic.

TABLE 3. False coverage probability ($1-\alpha = 90\%$)

		$n=250$	$n=500$	$n=750$	$n=1000$
FCP-θ_1^s	$CS_n^{s,proj}$	.502	.222	.089	.033
	$CS_n^s : \kappa_n = \sqrt{logn}$	.477	.194	.074	.025
	$CS_n^s : \kappa_n = \frac{\sqrt{logn}}{2}$	.461	.184	.071	.024
	$CS_n^s : \kappa_n = 2\sqrt{logn}$	.488	.204	.079	.028
	AS-PA	.718	.482	.294	.160
	AS-GMS	.666	.399	.219	.109
FCP-θ_2^s	$CS_n^{s,proj}$	.550	.313	.158	.076
	CS_n^s	.551	.313	.159	.078
	$CS_n^s : \kappa_n = \frac{\sqrt{logn}}{2}$	.551	.313	.159	.078
	$CS_n^s : \kappa_n = 2\sqrt{logn}$	.551	.313	.159	.078
	AS-PA	.684	.479	.304	.185
	AS-GMS	.662	.452	.274	.158

NOTES

1. e.g., Chernozhukov et al. (2007), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2008, 2010), and Chernozhukov, Lee and Rosen (2013).
2. See e.g., Stoye (2010), Andrews and Soares (2010), and Chernozhukov et al. (2013).
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.
4. If the parameter is point identified, there are no inequality constraints, and θ_0 lies in the interior of Θ, then the classical minimum distance estimator is consistent and asymptotically normal. In this case, our set estimator will be a singleton and coincide with the classical minimum distance estimator. However, our confidence set projected to each coordinate of θ is wider than the standard profile likelihood ratio confidence interval for the same coordinate because the critical value $\chi^2_{d_2}(1-\alpha)$ exceeds the univariate critical value $(\chi^2_1(1-\alpha))$.
5. Generalizing the example to a game with more than 2 players can be done following Ciliberto and Tamer (2009).
6. The condition may also be satisfied when g^e is not additively separable, but can be rewritten into an additively separable form by taking a nonlinear (e.g., logarithmic) transformation.
7. In fact, the pointwise asymptotic distribution of the test statistic $\tau_n Q(\theta, \hat{\beta}_n; \hat{W}_n^*(\theta))$ is constant—$\chi^2_{d_2}$—regardless of the data generating process or the true value of θ, and thus does not have a discontinuity. This is because the inequality constraints are by design not dependent on the estimated quantity $\hat{\beta}_n$ and do not enter the test statistic.
8. See Bugni, Canay, and Shi (2012) for a general treatment of the moment inequality model.
9. It would be interesting to extend the current framework to allow continuous X to enter the model parametrically. But the extension requires substantial change to the current method and we plan to explore it in future work.
10. As it turned out, the HD numbers for $\theta \equiv (\theta^{s,\prime}, \gamma)^\prime$ and those for θ^s are identical up to the accuracy level we report. Thus, Table 1 can be seen as the results for both θ and θ^s. The reason that the numbers are identical is that the γ values in both the estimated set and the identified set are of much smaller magnitude than the a and δ values.

11. The AS's SUM/PA confidence set uses their SUM test statistic and the plug-in asymptotic (PA) critical value. The AS's SUM/GMS confidence set uses their SUM test statistic and the generalized moment selection critical value. The GMS critical value uses the $\varphi(x) = +\infty 1(x > 1)$ moment selection function (assuming $\infty \cdot 0 = 0$), and $\kappa_n = \sqrt{\log(n)}$. The $\kappa_n = \sqrt{\log(n)}$ is recommended by Andrews and Soares (2010).

12. We pick the first point to be a simple deviation from the true value and the second point to be a simple deviation from one extreme point of the identified set. Our results are not likely dependent on the points picked.

13. In theory, the FCPs of CS_n^s should not be bigger than those of $CS_n^{s,proj}$, as discussed in Remark(ii) of Theorem 4.1. Yet in the table, the former appear to be slightly bigger than the latter at $n = 250, 750$, and 1000. This is due to simulation error in the critical value procedure for the profiled confidence set. The critical value for CS_n (and thus for its projection $CS_n^{s,proj}$) is $\chi_3^2(90\%)$ and does not involve simulation error.

REFERENCES

Andrews, D.W.K., S. Berry, & 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.

Andrews, D.W.K. & X. Shi (2013) Inference based on conditional moment inequalities. *Econometrica* 81, 609–666.

Andrews, D.W.K. & G. Soares (2010) Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78, 119–157.

Bajari, P., J. Hahn, H. Hong, & G. Ridder (2011) A note on semiparametric estimation of finite mixture of discrete choice models with application to game theoretical models. *International Economic Review* 52, 807–824.

Bajari, P., C. Lanier Benkard, & J. Levin (2007) Estimating dynamic models of imperfect competition. *Econometrica* 75, 1331–1370.

Bonhomme, S. (2012) Functional differencing. *Econometrica* 80, 1337–1385.

Bugni, F.A. (2010) Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* 78, 735–753.

Bugni, F., I. Canay, & X. Shi (2012) Specification Test for Partially Identified Models. Working paper, Duke University.

Canay, I.A. (2010) EL inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics* 156, 408–425.

Chernozhukov, V., H. Hong, & E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75, 1243–1284.

Chernozhukov, V., S. Lee, & A. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81, 667–737.

Ciliberto, F. & E. Tamer (2009) Market structure and multiple equilibria in the airline industry. *Econometrica* 77, 1791–1828.

Corbae, D., M.B. Stinchcombe, & J. Zeman (2009) *An Introduction to Mathematical Analysis for Economic Theory and Econometrics*, 1st ed. Princeton University Press.

Iaryczower, M., X. Shi, & M. Shum (2012) Words Get in the Way? The Effect of Deliberation in Collective Decision-Making. Working paper, Princeton University.

Imbens, G. & C.F. Manski (2004) Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.

Kaido, H. & A. Santos (2011) Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities. Working paper, Boston University.

Phillips, P.C.B. (2012) Folklore theorems, implicit maps, and indirect inference. *Econometrica* 80, 425–454.

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

Romano, J.P. & A.M. Shaikh (2010) Inference for the identified set in partially identified econometric models. *Econometrica* 78, 169–211.

Rudin, W. (1976) *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill Companies, Inc.

Stoye, J. (2010) More on confidence intervals for partially identified parameters. *Econometrica* 77, 1299–1315.

van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Yildiz, N. (2012) Consistency of plug-in estimators of upper contour and level sets. *Econometric Theory* 28, 309–327.

Zhang, W. & S.S. Ge (2006) A global implicit function theorem without initial point and its applications to control of non-affine systems of high dimensions. *Journal of Mathematical Analysis and Applications* 313, 251–261.

APPENDIX: 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 continuous, 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 $(e_1, \ldots, e_{d_y})$ 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 its Jacobian $\partial \bar{f}(x_0, y_0)/\partial y = [\partial f(x_0, y_0)'/\partial y|E']'$ 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 $|y_{m,n_m} - y_0| \le \inf_n |y_{m,n} - y_0| + 2^{-m}$. 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 $n \ge N$, $|y_n - y_0| < \epsilon/3$. Since $\lim_{m\to\infty} y_{m,N} = y_N$, there exists M_1 such that for all $m \ge M_1$, such that $|y_{m,N} - y_N| < \epsilon/3$. Let M_2 be an integer for all $m \ge M_2$, $2^{-m} < \epsilon/3$. Then for any $m > \max\{M_1, M_2\}$, we have

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

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. The proof makes use of the following set-valued function: $C : \mathcal{B} \times \mathcal{W} \to \mathcal{K}(\Theta)$ defined by $C(\beta, W) = \arg\min_{\theta\in\Theta_{ie}} Q(\theta, \beta, W)$, where $\mathcal{W}$ is the set of $d_2 \times d_2$ positive definite matrices and $\mathcal{K}(\Theta)$ is the set of subsets of Θ. Clearly, $\Theta_0 = C(\beta_0, W_0)$, and $\hat{\Theta}_n = C(\hat{\beta}_n, \hat{W}_n)$. Consider the metric spaces $(\mathcal{K}(\Theta), d_H)$ and $(\mathcal{B} \times \mathcal{W}, \|\cdot\|)$ where $\|(\beta_1, W_1)\|^2 = \|\beta_1\|^2 + trace(W_1' W_1)$. Below we show that the mapping $C : \mathcal{B} \times \mathcal{W} \to \mathcal{K}(\Theta)$ is continuous at (β_0, W_0) under the metrics we consider. Once the continuity is established, the continuous mapping theorem (e.g., van der Vaart and Wellner, 1996, Thm. 1.9.5) applies and yields the desired result.

We now prove the continuity of $C(\cdot, \cdot)$. Consider a (deterministic) sequence $(\beta_n, W_n) \in \mathcal{B} \times \mathcal{W}$ such that $(\beta_n, W_n) \to (\beta_0, W_0)$ and we want to show that

$$d_H(C(\beta_n, W_n), C(\beta_0, W_0)) \to 0. \tag{A.3}$$

The proof of the continuity of (A.3) contains four steps. For clarity, we first sketch the steps and afterward give detailed arguments for each step.

STEP 1. Let $\bar{\theta}_n$ be an arbitrary point in $C(\beta_n, W_n)$ and $\theta_n \in \arg\min_{\theta\in\Theta_0} ||\theta - \bar{\theta}_n||$. We show that $||\bar{\theta}_n - \theta_n|| \to 0$. This implies that

$$\sup_{\theta\in C(\beta_n, W_n)} \inf_{\theta_0\in\Theta_0} ||\theta - \theta_0|| \to 0. \tag{A.4}$$

If Θ_0 is a singleton, the proof is finished. Otherwise, the following steps are needed.

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

STEP 3. Let $\tilde{\Theta}_n = r(\beta_n, g^e(\bar{\theta}_n, \beta_n))$ for an arbitrary $\bar{\theta}_n \in C(\beta_n, W_n)$. Then clearly $\tilde{\Theta}_n \neq \emptyset$ and $\tilde{\Theta}_n \subseteq C(\beta_n, W_n)$. Also, by (A.4), there exists a sequence $\theta_n \in \Theta_0$ such that $\|\bar{\theta}_n - \theta_n\| \to 0$. By this, $\beta_n \to \beta_0$, $g^e(\theta_n, \beta_0) = 0$, and the uniform continuity of g^e on $\Theta \times \mathcal{B}$, we have $g^e(\bar{\theta}_n, \beta_n) \to 0$. This combined with $\beta_n \to \beta_0$ and the continuity of r shown in STEP 2 implies that $d_H(\tilde{\Theta}_n, \Theta_0) \to 0$.

STEP 4. Because $\tilde{\Theta}_n \subseteq C(\beta_n, W_n)$, STEP 3 implies that $\sup_{\theta_0 \in \Theta_0} \inf_{\theta \in C(\beta_n, W_n)} ||\theta - \theta_0|| \to 0$. This combined with (A.4) shows (A.3).

Now we give detailed arguments for STEPS 1–3. STEP 4 is self-evident.

The proof for STEP 1 takes the form of a standard consistency proof. Two major components of it are the uniform convergence of Q and global identification:

$$\sup_{\theta \in \Theta_{ie}} |Q(\theta, \beta_n; W_n) - Q(\theta, \beta_0; W_0)| \to 0, \textit{ and}$$
$$\forall \epsilon > 0, \exists \delta_\epsilon > 0 \; s.t. \inf_{\theta \in \Theta : \inf_{\theta_0 \in \Theta_0} ||\theta - \theta_0|| > \epsilon} Q(\theta, \beta_0; W_0) > \delta_\epsilon. \tag{A.5}$$

The uniform convergence is implied by the continuity of g^e on the compact set $\Theta \times \mathcal{B}$, $\beta_n \to \beta_0$, and $W_n \to W_0$. In more detail, observe that

$$\begin{aligned}
&\sup_{\theta \in \Theta_{ie}} |Q(\theta, \beta_n; W_n) - Q(\theta, \beta_0; W_0)| \\
&\leq \sup_{\theta \in \Theta_{ie}} |(g^e(\theta, \beta_n) - g^e(\theta, \beta_0))' W_n g^e(\theta, \beta_n)| \\
&\quad + \sup_{\theta \in \Theta_{ie}} |(g^e(\theta, \beta_n) - g^e(\theta, \beta_0))' W_n g^e(\theta, \beta_0)| + \sup_{\theta \in \Theta_{ie}} |g^e(\theta, \beta_0)'(W_n - W_0) g^e(\theta, \beta_0)| \\
&\leq 2 \sup_{\theta \in \Theta_{ie}} \|g^e(\theta, \beta_n) - g^e(\theta, \beta_0)\| \|W_n\| \sup_{\theta \in \Theta_{ie}, \beta \in \mathcal{B}} \|g^e(\theta, \beta)\| \\
&\quad + \sup_{\theta \in \Theta_{ie}} \|g^e(\theta, \beta_0)\|^2 \|W_n - W_0\|.
\end{aligned} \tag{A.6}$$

The first summand on the right-hand side converges to zero in probability because $g^e(\theta,\beta)$ is uniformly continuous on $\Theta_{ie} \times \mathcal{B}$ (which holds because it is continuous and $\Theta_{ie} \times \mathcal{B}$ is compact) and because $\|W_n\| = O(1)$ and $\sup_{\theta \in \Theta_{ie}, \beta \in \mathcal{B}} \|g^e(\theta,\beta)\| < \infty$ (which holds because g^e is continuous and $\Theta_{ie} \times \mathcal{B}$ is compact). The second summand on the right-hand side converges to zero due to similar argument. This shows the uniform convergence.

The global identification condition is implied by the definition of Θ_0, the continuity of $Q(\cdot, \beta_0; W_0)$, and the compactness of Θ.

Using the uniform convergence result, we have

$$\begin{aligned}
Q(\bar{\theta}_n, \beta_0; W_0) &= Q(\bar{\theta}_n, \beta_0; W_0) - Q(\bar{\theta}_n, \beta_n; W_n) \\
&\quad + Q(\bar{\theta}_n, \beta_n; W_n) - Q(\theta_n, \beta_n; W_n) \\
&\quad + Q(\theta_n, \beta_n; W_n) - Q(\theta_n, \beta_0; W_0) \\
&\leq 2 \sup_{\theta \in \Theta_{ie}} |Q(\theta, \beta_0; W_0) - Q(\theta, \beta_n; W_n)| \to 0.
\end{aligned} \tag{A.7}$$

Above, the equality holds by adding and subtracting terms and by $Q(\theta_n, \beta_0, W) = 0$ (which holds because $\theta_n \in \Theta_0$). The inequality holds because $Q(\bar{\theta}_n, \beta_n, W_n) \le Q(\theta_n, \beta_n, W_n)$ (by $\bar{\theta}_n \in C(\beta_n, W_n)$) and because both $\bar{\theta}_n$ and θ_n are in Θ_{ie}. Finally, the convergence holds by the first line of (A.5). This combined with the second line of (A.5) shows $\|\bar{\theta}_n - \theta_n\| \to 0$, which in turn shows (A.4).

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

In STEP 3, first by the upper hemicontinuity of $r(\cdot)$, for any $\varepsilon > 0$, there exists a N large enough such that for all $n \ge N$, we have $\tilde{\Theta}_n \equiv r(\beta_n, g^e(\bar{\theta}_n, \beta_n)) \subseteq r(\beta_0, 0)^\varepsilon := \{\theta \in \Theta : \inf_{\theta^* \in r(\beta_0, 0)} \|\theta - \theta^*\| \le \varepsilon\}$. This implies that

$$\lim_{n\to\infty} \sup_{\theta \in \tilde{\Theta}_n} \inf_{\theta^* \in r(\beta_0, 0)} \|\theta - \theta^*\| = 0. \tag{A.8}$$

Now consider $\theta_n^* \in r(\beta_0, 0)$ such that $\inf_{\theta \in \tilde{\Theta}_n} \|\theta_n^* - \theta\| = o(1) + \sup_{\theta^* \in r(\beta_0, 0)} \inf_{\theta \in \tilde{\Theta}_n} \|\theta^* - \theta\|$. Let $\{u_n\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} \inf_{\theta \in \tilde{\Theta}_{u_n}} \|\theta_{u_n}^* - \theta\| = \limsup_{n\to\infty} \inf_{\theta \in \tilde{\Theta}_n} \|\theta_n^* - \theta\|$ and $\theta_{u_n}^* \to \theta_\infty^*$ for some $\theta_\infty^* \in \Theta$. Such a subsequence always exists by the property of lim sup and the compactness of Θ. By the lower-semicontinuity of $r(\cdot)$, for any $\varepsilon > 0$ and the open ball $\mathcal{N}_\varepsilon(\theta_\infty^*)$, there exists N large enough such that for all $n \ge N$, $\tilde{\Theta}_{u_n} \cap \mathcal{N}_\varepsilon(\theta_\infty^*) \neq \emptyset$, which implies that $\inf_{\theta \in \tilde{\Theta}_{u_n}} \|\theta - \theta_\infty^*\| < \varepsilon$. Therefore, $\lim_{n\to\infty} \inf_{\theta \in \tilde{\Theta}_{u_n}} \|\theta - \theta_\infty^*\| = 0$. But by the definition of θ_∞^*, we have

$$\begin{aligned} \limsup_{n\to\infty} \sup_{\theta^* \in r(\beta_0, 0)} \inf_{\theta \in \tilde{\Theta}_n} \|\theta^* - \theta\| &= \lim_{n\to\infty} \inf_{\theta \in \tilde{\Theta}_{u_n}} \|\theta_{u_n}^* - \theta\| \\ &= \lim_{n\to\infty} \inf_{\theta \in \tilde{\Theta}_{u_n}} \|\theta - \theta_\infty^*\| = 0. \end{aligned} \tag{A.9}$$

Lastly observe that $r(\beta_0, 0) = \Theta_0$. This along with (A.8) and (A.9) implies $d_H(\tilde{\Theta}_n, \Theta_0) \to 0$. ■

Proof of Theorem 3.1. (a) By the definition of inf, there exists a sequence $\{\theta_n \in \Theta_0\}$ such that $\Pr(\theta_n \in CS_n) = \inf_{\theta \in \Theta_0} \Pr(\theta \in CS_n) + o(1)$. Thus,

$$\liminf_{n\to\infty} \inf_{\theta \in \Theta_0} \Pr(\theta \in CS_n) = \liminf_{n\to\infty} \Pr(\theta_n \in CS_n). \tag{A.10}$$

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

$$\liminf_{n\to\infty} \Pr(\theta_n \in CS_n) = \lim_{n\to\infty} \Pr(\theta_{u_n} \in CS_{u_n}). \tag{A.11}$$

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

$$\tau_{a_n}^2 Q(\theta_{a_n}, \hat{\beta}_{a_n}; \hat{W}_{a_n}^*(\theta_{a_n})) \to_d \chi_{d_2}^2. \tag{A.12}$$

To show this observe that

$$\tau_{a_n}^2 Q(\theta_{a_n}, \hat{\beta}_{a_n}; \hat{W}_{a_n}^*(\theta_{a_n})) = \tau_{a_n}^2 [g^e(\theta_{a_n}, \hat{\beta}_{a_n}) - g^e(\theta_{a_n}, \beta_0)]' \hat{W}_{a_n}^*(\theta_{a_n}) \times [g^e(\theta_{a_n}, \hat{\beta}_{a_n}) - g^e(\theta_{a_n}, \beta_0)]. \quad \textbf{(A.13)}$$

By the continuity of G, we have $G(\theta_{a_n}, \hat{\beta}_{a_n}) \to_p G(\theta_0, \beta_0)$. Thus,

$$[\hat{W}_{a_n}^*(\theta_{a_n})]^{-1} \equiv G(\theta_{a_n}, \hat{\beta}_{a_n}) \hat{V}_{\beta, a_n} G(\theta_{a_n}, \hat{\beta}_{a_n})' \to_p G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)'. \quad \textbf{(A.14)}$$

Because $G(\theta_0, \beta_0)' V_\beta G(\theta_0, \beta_0)$ is invertible, by Slutsky's theorem,

$$\hat{W}_{a_n}^*(\theta_{a_n}) \to_p [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1}. \quad \textbf{(A.15)}$$

Also, by mean-value expansion,

$$\tau_{a_n}(g^e(\theta_{a_n}, \hat{\beta}_{a_n}) - g^e(\theta_{a_n}, \beta_0)) = \tau_{a_n} G(\theta_{a_n}, \tilde{\beta}_{a_n})(\hat{\beta}_{a_n} - \beta_0) \to_d G(\theta_0, \beta_0) Z_\beta \sim N(0, G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)', \quad \textbf{(A.16)}$$

where the convergence holds by the continuity of $G(\cdot, \cdot)$, $\tau_{a_n}(\hat{\beta}_{a_n} - \beta_0) \to Z_\beta$ and the continuous mapping theorem. Therefore,

$$\tau_{a_n}^2 Q(\theta_{a_n}, \hat{\beta}_{a_n}, \hat{W}_{a_n}^*(\theta_{a_n})) \to_d Z_\beta' G(\theta_0, \beta_0)' [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1} G(\theta_0, \beta_0) Z_\beta \sim \chi_{d_2}^2. \quad \textbf{(A.17)}$$

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

$$\liminf_{n\to\infty} \inf_{\theta \in \Theta_0} \Pr(\theta \in CS_n) = 1 - \alpha. \quad \textbf{(A.18)}$$

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

$$\sup_{\theta \in \Theta_0} \tau_n^2 Q(\theta, \hat{\beta}_n, \hat{W}_n^*(\theta)) = \tau_n^2 Q(\theta_n, \hat{\beta}_n, \hat{W}_n^*(\theta_n)) + o_p(1). \quad \textbf{(A.19)}$$

Like in (A.13), we can write

$$\tau_n^2 Q(\theta_n, \hat{\beta}_n, \hat{W}_n^*(\theta_n)) = \tau_n^2 (\hat{\beta}_n - \beta_0)' G(\theta_n, \tilde{\beta}_n)' \hat{W}_n^*(\theta_n) G(\theta_n, \tilde{\beta}_n)(\hat{\beta}_n - \beta_0). \quad \textbf{(A.20)}$$

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,

$$\sup_{\theta\in\Theta_0} \|G(\theta, \tilde{\beta}_n) - G(\theta, \beta_0)\| \to_p 0 \; and \; \sup_{\theta\in\Theta_0} \|G(\theta, \hat{\beta}_n) - G(\theta, \beta_0)\| \to_p 0. \quad \textbf{(A.21)}$$

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.21), for any random sequence $\{\theta_n \in \Theta_0\}$,

$$[G(\theta_n, \tilde{\beta}_n)' \hat{W}_n^*(\theta_n) G(\theta_n, \tilde{\beta}_n)] \to_p G(\theta_0, \beta_0)' [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1} G(\theta_0, \beta_0). \quad \textbf{(A.22)}$$

Therefore,

$$\tau_n^2 Q(\theta_n, \hat{\beta}_n, \hat{W}_n^*(\theta_n)) \\ \to_d Z_\beta' G(\theta_0, \beta_0)' [G(\theta_0, \beta_0) V_\beta G(\theta_0, \beta_0)']^{-1} G(\theta_0, \beta_0) Z_\beta \sim \chi^2_{d_2}. \quad \textbf{(A.23)}$$

Combining this with (A.19), we get

$$\begin{aligned} \Pr(\hat{\Theta}_n \subseteq CS_n) &= \Pr(\sup_{\theta \in \Theta_0} \tau_n^2 Q(\theta, \hat{\beta}_n, \hat{W}_n^*(\theta)) \leq \chi^2_{d_2}(1-\alpha)) \\ &= \Pr(\tau_n^2 Q(\theta_n, \hat{\beta}_n, \hat{W}_n^*(\theta_n)) + o_p(1) \leq \chi^2_{d_2}(1-\alpha)) \\ &\to \Pr(\chi^2_{d_2} \leq \chi^2_{d_2}(1-\alpha)) = 1 - \alpha. \end{aligned} \quad \textbf{(A.24)}$$

■

The proof of Theorem 4.1 makes use of the following lemma:

LEMMA A.2 (Continuity of Extreme Value). *Consider metric spaces* $(\mathcal{X}, d_\mathcal{X})$, $(\mathcal{Y}, d_\mathcal{Y})$ *and a function* $f(x, y) : \mathcal{X} \times \mathcal{Y} \to R \cup \{\infty\}$. *Suppose that* $f(x, y)$ *is continuous in* $\mathcal{X} \times \mathcal{Y}$ *at any* $(x, y) \in \mathcal{X} \times \mathcal{Y}$; *that is, for any sequence* $x_n \to x$ *and* $y_n \to y$, *we have* $f(x_n, y_n) \to f(x, y)$, *whether* $f(x, y) \in R$ *or* $f(x, y) = \infty$. *Also suppose that* $(\mathcal{X}, d_\mathcal{X})$ *is compact. Let* $\{X_n\}$ *be a sequence of compact subsets of* $(\mathcal{X}, d_\mathcal{X})$ *and* $\{y_n\}$ *be a sequence of elements in* $\mathcal{Y}$, *such that* $X_n \to X$ *in the Hausdorff distance based on* $d_\mathcal{X}$ *and* $y_n \to y$ *in* $d_\mathcal{Y}$, *for some compact subset* X *of* $(\mathcal{X}, d_\mathcal{X})$ *and some* $y \in \mathcal{Y}$.

Then

$$\min_{x \in X_n} f(x, y_n) \to \min_{x \in X} f(x, y). \quad \textbf{(A.25)}$$

Proof. First, because X_n is compact and $f(\cdot, y_n)$ is continuous on X_n, there exists $x_n^* \in X_n$ such that $f(x_n^*, y_n) = \min_{x \in X_n} f(x, y_n)$ by the extreme value theorem. Let $\{u_n\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} f(x_{u_n}^*, y_{u_n}) = \liminf_{n\to\infty} f(x_n^*, y_n)$. Such a subsequence always exists by the definition of lim inf. Let $\{v_n\}$ be a further subsequence of $\{u_n\}$ such that $x_{v_n}^* \to x^*$ for some $x^* \in X$. Such a subsequence also exists because $(\mathcal{X}, d_\mathcal{X})$ is compact and $X_n \to X$ in Hausdorff distance. Then

$$\begin{aligned} \liminf_{n\to\infty} \min_{x \in X_n} f(x, y_n) &= \liminf_{n\to\infty} f(x_n^*, y_n) \\ &= \lim_{n\to\infty} f(x_{v_n}^*, y_{v_n}) = f(x^*, y) \geq \min_{x \in X} f(x, y). \end{aligned} \quad \textbf{(A.26)}$$

If $\min_{x \in X} f(x, y) = \infty$, then the proof is done. If $\min_{x \in X} f(x, y) \in R$, we also need to show $\limsup_{n\to\infty} \min_{x \in X_n} f(x, y_n) \leq \min_{x \in X} f(x, y)$.

Let x^* be a point in X such that $f(x^*, y) = \min_{x \in X} f(x, y)$. Then there exists a sequence $\{x_n^* \in X_n\}$ such that $x_n^* \to x^*$. By construction, for every n, $\min_{x \in X_n} f(x, y_n) \leq f(x_n^*, y_n)$. Therefore,

$$\begin{aligned} \limsup_{n\to\infty} \min_{x \in X_n} f(x, y_n) &\leq \limsup_{n\to\infty} f(x_n^*, y_n) \\ &= \lim_{n\to\infty} f(x_n^*, y_n) = f(x^*, y) = \min_{x \in X} f(x, y). \end{aligned} \quad \textbf{(A.27)}$$

This concludes the proof. ■

Proof of Theorem 4.1. (a) Similar to the proof of Theorem 3.1(a), consider a sequence $\{\theta_n^s \in \Theta_0^s\}$ and a subsequence $\{u_n\}$ of $\{n\}$ such that $\liminf_{n\to\infty} \inf_{\theta^s \in \Theta_0^s} \Pr(\theta \in CS_n^s) = \lim_{n\to\infty} \Pr(\theta_{u_n} \in CS_{u_n}^s)$. Such a subsequence always exists. Then, because Θ^s is compact, there must exist a further subsequence $\{v_n\}$ of $\{u_n\}$ such that $\theta_{v_n}^s \to \theta_\infty^s$ for some $\theta_\infty^s \in \Theta^s$. By the continuity of $g^e, g^{ie,1}, g^{ie,2}$, the limit point θ_∞^s must be a point in Θ_0^s.

Define the metric d_Φ on $[-\infty,\infty]^{d_\gamma}$ as

$$d_\Phi(\ell,\ell^*) = \|\bar\Phi(\ell) - \bar\Phi(\ell^*)\|, \ \forall \ell,\ell^* \in [-\infty,\infty]^{d_\gamma} \quad \textbf{(A.28)}$$

where $\bar\Phi(\ell) = (\Phi(\ell_1),\ldots,\Phi(\ell_{d_\gamma}))'$ for $\ell = (\ell_1,\ldots,\ell_{d_\gamma})'$, and $\Phi(x)$ is the standard normal cumulative distribution function (cdf) function for $x \in R$ and $\Phi(\infty) = 1$ and $\Phi(-\infty) = -1$. It is easy to verify that the metric space $([-\infty,\infty]^{d_\gamma}, d_\Phi)$ is compact and so are its subspaces (L_n, d_Φ) and (L_n', d_Φ), where

$$\begin{aligned} L_n &:= \{\tau_n(\gamma - \gamma_0(\theta_n^s)) : \gamma \in \Gamma(\theta_n^s),\ \gamma_0(\theta_n^s) = g^{ie,1}(\theta_n^s, \beta_0)\} \\ L_n' &:= \{\kappa_n^{-1}\tau_n(\gamma - \gamma_0(\theta_n^s)) : \gamma \in \Gamma(\theta_n^s),\ \gamma_0(\theta_n^s) = g^{ie,1}(\theta_n^s, \beta_0)\}. \end{aligned} \quad \textbf{(A.29)}$$

Then by Theorem 6.1.16 of Corbae, Stinchcombe, and Zeman (2009), there exists a subsequence $\{a_n\}$ of $\{v_n\}$ and sets $L_\infty, L_\infty' \subseteq [0,\infty]^{d_\gamma}$ compact under the d_Φ metric such that

$$d_{H,\Phi}(L_{a_n}, L_\infty) := \max\left\{ \sup_{\ell \in L_{a_n}} \inf_{\ell^* \in L_\infty} d_\Phi(\ell,\ell^*),\ \sup_{\ell^* \in L_\infty} \inf_{\ell \in L_{a_n}} d_\Phi(\ell,\ell^*) \right\} \to 0, \text{ and}$$
$$d_{H,\Phi}(L_{a_n}', L_\infty') \to 0. \quad \textbf{(A.30)}$$

Next we focus on this subsequence $\{a_n\}$ and derive the asymptotic distribution of $\tau_{a_n}^2 Q(\theta_{a_n}^s, \hat\beta_{a_n})$. Observe that

$$\begin{aligned} \tau_{a_n}^2 Q(\theta_{a_n}^s, \hat\beta_{a_n}) &= \min_{\gamma \in \Gamma(\theta_{a_n}^s)} \begin{pmatrix} \tau_{a_n}(g^e(\theta_{a_n}^s, \hat\beta_{a_n}) - g^e(\theta_{a_n}^s, \beta_0)) \\ \tau_{a_n}(g^{ie,1}(\theta_{a_n}^s, \hat\beta_{a_n}) - g^{ie,1}(\theta_{a_n}^s, \beta_0)) - \tau_{a_n}(\gamma - \gamma_0(\theta_{a_n}^s)) \end{pmatrix}' \\ &\quad \times \hat W_{a_n}^*(\theta_{a_n}^s) \begin{pmatrix} \tau_{a_n}(g^e(\theta_{a_n}^s, \hat\beta_{a_n}) - g^e(\theta_{a_n}^s, \beta_0)) \\ \tau_{a_n}(g^{ie,1}(\theta_{a_n}^s, \hat\beta_{a_n}) - g^{ie,1}(\theta_{a_n}^s, \beta_0)) - \tau_{a_n}(\gamma - \gamma_0(\theta_{a_n}^s)) \end{pmatrix} \\ &= \min_{\ell \in L_{a_n}} \begin{pmatrix} \hat Z_{a_n}^e \\ \hat Z_{a_n}^{ie} - \ell \end{pmatrix}' \hat W_{a_n}^*(\theta_{a_n}^s) \begin{pmatrix} \hat Z_{a_n}^e \\ \hat Z_{a_n}^{ie} - \ell \end{pmatrix}, \end{aligned} \quad \textbf{(A.31)}$$

where $\hat Z_{a_n}^e = \tau_{a_n}(g^e(\theta_{a_n}^s, \hat\beta_{a_n}) - g^e(\theta_{a_n}^s, \beta_0))$, and $\hat Z_{a_n}^{ie} = \tau_{a_n}(g^{ie,1}(\theta_{a_n}^s, \hat\beta_{a_n}) - g^{ie,1}(\theta_{a_n}^s, \beta_0))$. It is clear from the last line of the above display that $\tau_{a_n}^2 Q(\theta_{a_n}^s, \hat\beta_{a_n})$ is following functional evaluated at $\left(L_{a_n}, \hat Z_{a_n}^e, \hat Z_{a_n}^{ie}, \hat W^*(\theta_{a_n}^s)\right)$:

$$\varphi\left(L, z^e, z^{ie}, W\right) = \min_{\ell \in L} \begin{pmatrix} z^e \\ z^{ie} - \ell \end{pmatrix}' W \begin{pmatrix} z^e \\ z^{ie} - \ell \end{pmatrix}. \quad \textbf{(A.32)}$$

By Lemma A.2, for any sequence $(L_n, z_n^e, z_n^{ie}, W_n) \in \mathcal{K}([-\infty,\infty]^{d_\gamma}) \times R^{d_e} \times R^{d_\gamma} \times \mathcal{W}$, where $\mathcal{K}$ is the set of compact subsets of $([-\infty,\infty]^{d_\gamma}, d_\Phi)$, if $(L_n, z_n^e, z_n^{ie}, W_n) \to$

$(L, z^e, z^{ie}, W_n) \in \mathcal{K}([-\infty,\infty]^{d_\gamma}) \times R^{d_e} \times w \times R^{d_\gamma}$, then $\varphi(L_n, z_n^e, z_n^{ie}, W_n) \to \varphi(L, z^e, z^{ie}, W)$. Arguments identical to those that show (A.15) and (A.16) can be used to show:

$$(\hat{Z}_{an}^{e,\prime}, \hat{Z}_{an}^{ie,\prime})' \to_d (Z^{e,\prime}, Z^{ie,\prime})' \sim N(0, W^*(\theta_\infty^s)^{-1}),$$
$$\hat{W}_{a_n}^*(\theta_{a_n}^s) \to_p W^*(\theta_\infty^s). \tag{A.33}$$

With these results and (A.30), the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Thm. 1.11.1) applies and shows that

$$\tau_{a_n}^2 Q(\theta_{a_n}^s, \hat{\beta}_{a_n}) \to_d \varphi(L_\infty, Z^e, Z^{ie}, W^*(\theta_\infty^s)). \tag{A.34}$$

Similar to (A.31), we can write

$$J_n(\theta_n^s) = \min_{\ell \in L_n'} \begin{pmatrix} Z_n^e \\ Z_n^{ie} + \kappa_n^{-1}(\hat{\gamma}(\theta^s) - \gamma_0(\theta^s)) - \ell \end{pmatrix}' \hat{W}^*(\theta_n^s) \times \begin{pmatrix} Z_n^e \\ Z_n^{ie} + \kappa_n^{-1}(\hat{\gamma}(\theta^s) - \gamma_0(\theta^s)) - \ell \end{pmatrix}'. \tag{A.35}$$

We see that $J_n(\theta_n^s) = \varphi(L_n', Z_n^e, Z_n^{ie} + \kappa_n^{-1}(\hat{\gamma}(\theta^s) - \gamma_0(\theta^s)), \hat{W}^*(\theta_n^s))$. Also,

$$\|\kappa_n^{-1}(\hat{\gamma}(\theta^s) - \gamma_0(\theta^s))\| \le \|\kappa_n^{-1}(g^{ie,1}(\theta^s, \hat{\beta}_n) - \gamma_0(\theta^s)\| = o_p(1), \tag{A.36}$$

where the inequality holds due to the convexity of $\Gamma(\theta^s)$ and the definition of $\hat{\gamma}(\theta^s)$ (below equation (4.4)), and the equality holds by $\kappa_n \to \infty$ and (A.33). Then by similar arguments as those for (A.34), we can show that

$$J_{a_n}(\theta_{a_n}^s) \to_d \varphi(L_\infty', Z^e, Z^{ie}, W^*(\theta_\infty^s)). \tag{A.37}$$

Because $d_e > 0$, the cdf of $\varphi(L_\infty', Z^e, Z^{ie}, W^*(\theta_\infty^s))$ is continuous and strictly increasing. Thus, by standard arguments,

$$c_{a_n}(\theta_{a_n}^s, 1-\alpha) \to_p q_{1-\alpha}, \tag{A.38}$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of $\varphi(L_\infty', Z^e, Z^{ie}, W^*(\theta_\infty^s))$.

Because $\Gamma(\theta^s)$ is convex and contains $\gamma_0(\theta^s)$, the sets L_n is closed to multiplication of numbers from $[0, 1]$. That is, for any $\ell \in L_n$, $c \times \ell \in L_n$ as long as $c \in [0, 1]$. This implies that $L_n' \subset L_n$ for all large enough n because $\kappa_n^{-1} \to 0$. This then implies that $L_\infty' \subseteq L_\infty$. Then, by the definition of the function φ, with probability one,

$$\varphi(L_\infty', Z^e, Z^{ie}, W^*(\theta_\infty^s)) \ge \varphi(L_\infty, Z^e, Z^{ie}, W^*(\theta_\infty^s)). \tag{A.39}$$

Therefore,

$$\begin{aligned} \liminf_{n\to\infty} \Pr(\theta_n^s \in CS_n^s) &= \lim_{n\to\infty} \Pr(\tau_{a_n}^2 Q(\theta_{a_n}^s, \hat{\beta}_{a_n}) \le c_{a_n}(\theta_{a_n}^s, 1-\alpha)) \\ &= \Pr(\varphi(L_\infty, Z^e, Z^{ie}, W^*(\theta_\infty^s)) \le q_{1-\alpha}) \\ &\ge \Pr(\varphi(L_\infty', Z^e, Z^{ie}, W^*(\theta_\infty^s)) \le q_{1-\alpha}) = 1-\alpha. \end{aligned} \tag{A.40}$$

(b) To show part (b), we consider a fixed $\theta^s \in \Theta_0^s$. Let L_n and L'_n be defined as in (A.29) except with θ_n^s replaced by the fixed θ^s. Then both $\{L_n\}$ and $\{L'_n\}$ are increasing sequences of sets. Let $L_\infty = cl(\cup_{n=1}^\infty L_n)$ and $L'_\infty = cl(\cup_{n=1}^\infty L'_n)$. It is straightforward to show that $d_{H,\Phi}(L_n, L_\infty) \to 0$ and $d_{H,\Phi}(L'_n, L'_\infty) \to 0$.

Then, all the steps in part (a) from (A.31) to (A.38) go through with $\{a_n\}$ replaced by $\{n\}$ and θ_n^s and θ_∞^s replaced by the fixed θ^s. Below we show $L_\infty = L'_\infty$. With this, we have

$$\begin{aligned}\limsup_{n\to\infty} \inf_{\theta^s \in \Theta_0^s} \Pr(\theta^s \in CS_n) &\le \limsup_{n\to\infty} \Pr(\tau_n^2 Q(\theta^s, \hat{\beta}_n) \le c_n(\theta_n^s, 1-\alpha)) \\ &= \Pr(\varphi(L_\infty, Z^e, Z^{ie}, W^*(\theta_\infty^s)) \le q_{1-\alpha}) \\ &= \Pr(\varphi(L'_\infty, Z^e, Z^{ie}, W^*(\theta_\infty^s)) \le q_{1-\alpha}) = 1-\alpha. \qquad \textbf{(A.41)}\end{aligned}$$

To show that $L_\infty = L'_\infty$, consider a point $\ell \in L_\infty$. Then there is a sequence $\ell_n \in L_n$ such that $d_\Phi(\ell_n, \ell) \to 0$. By the definition of L_n, there is a $\gamma_n \in \Gamma(\theta^s)$ such that $\tau_n(\gamma_n - \gamma_0(\theta^s)) = \ell_n$ for every n. For each n, let a_n be the smallest integer such that $\kappa_{a_n}^{-1}\tau_{a_n} \ge \tau_n$. Such an a_n always exists because $\kappa_a^{-1}\tau_a \to \infty$ as $a \to \infty$. Then by the definition of L'_{a_n}, $\ell'_{a_n} = \kappa_{a_n}^{-1}\tau_{a_n}(\gamma_n - \gamma_0(\theta^s)) \in L'_{a_n}$. Also, because $\kappa_{a_n}^{-1}\tau_{a_n} \ge \tau_n$, the vector ℓ_n lies on the line segment connecting ℓ'_{a_n} and 0. Because the set $\Gamma(\theta^s)$ is convex, which implies that L'_{a_n} is convex, we have $\ell_n \in L'_{a_n}$. This holds for every n. Thus, we have found a sequence of points in L'_{a_n} such that the sequence converges to ℓ. Because $d_{H,\Phi}(L'_{a_n}, L'_\infty) \to 0$, we have $\ell \in L'_\infty$. This shows that $L_\infty \subseteq L'_\infty$. This combined with $L'_\infty \subseteq L_\infty$ shown in part (a) gives us $L_\infty = L'_\infty$. ■