Robust confidence regions for incomplete models

Larry G. Epstein
Hiroaki Kaido
Kyoungwon Seo

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP20/15

Robust Confidence Regions for Incomplete Models*

Larry G. Epstein Hiroaki Kaido Kyoungwon Seo

April 22, 2015

Abstract

Call an economic model incomplete if it does not generate a probabilistic prediction even given knowledge of all parameter values. We propose a method of inference about unknown parameters for such models that is robust to heterogeneity and dependence of unknown form. The key is a Central Limit Theorem for belief functions; robust confidence regions are then constructed in a fashion paralleling the classical approach. Monte Carlo simulations support tractability of the method and demonstrate its enhanced robustness relative to existing methods.

* *Corresponding author*: Kyoungwon Seo, Korea Advanced Institute of Science and Technology (KAIST), kseo@kaist.ac.kr. Epstein and Kaido are at Boston University, lepstein@bu.edu and hkaido@bu.edu. We gratefully acknowledge the financial support of KAIST (Seo) and the National Science Foundation SES-1216339 (Epstein) and SES-1357643 (Kaido). We are grateful also for comments from Andres Aradillas-Lopez, Gabriel Carroll, Victor Chernozhukov, Ivan Fernandez-Val, Alfred Galichon, Marc Henry, Yuhei Miyauchi, Whitney Newey, Joris Pinkse, Jack Porter, Zhongjun Qu, and from audiences at Penn State, MIT, BU, Wisconsin, UNC, Korea U. and Sungkyunkwan U.

1. Introduction

1.1. Objectives and outline

In a wide class of structural models, when the analyst is not willing to make assumptions that are driven by convenience rather than by economic theory, the resulting economic structures are *incomplete* in the sense that they do not yield unique reduced forms. In this paper, we consider the class of such models that can be represented as follows: given a structural parameter $\theta \in \Theta \subset \mathbb{R}^d$ and the realization $u \in U \subset \mathbb{R}^{d_u}$ of an unobservable random variable, the model predicts a nonsingleton set, denoted $G(u|\theta)$, of values for the outcome variable; that is, $G(u|\theta)$ is a subset of the (finite) outcome space S. Examples include discrete game models with multiple Nash equilibria (Jovanovic 1989, Tamer 2003), where the modeler is agnostic about the way in which selection from among multiple equilibria operates; auction models (Haile and Tamer 2003, Aradillas-Lopez 2008), where the modeler is agnostic about the precise game form underlying the auction data in her sample and/or she is willing to adopt only weak assumptions about bidders' behavior; and a model of job and skill heterogeneity (Galichon and Henry 2009), where the relation between the unobservable skill level and observable job characteristics is not well understood. A number of other applied models also fit into the class we consider. These include discrete choice models with social interactions (Soetevent and Kooreman 2007), matching models with externalities (Uetake and Watanabe 2012), and some network formation models (Sheng 2014, Miyauchi 2014). An incomplete structure arises also in a single-agent discrete choice model when multiple alternatives maximize the agent's utility with positive probability and a tie-breaking rule is not specified.

The lack of a unique reduced form implies that a conventional identification analysis based on a (single) likelihood cannot be applied, which has motivated recent research on identification and inference in incomplete models. An important objective of this literature is expressed by Ciliberto and Tamer (2009, p. 1800), who write in the context of entry games: "*This [selection] mechanism is not specified, and one objective of the methodology in this paper is to examine the question of what can be learned when researchers remain agnostic about this selection function.*"

A common assumption in the literature is the availability of i.i.d. samples of outcomes. To elaborate, think of a number of experiments, or random events, indexed by $i = 1, 2, ...$, each of which may be described as above, for a common

Θ, G and S;[1] then each infinite sequence of unobserved variables $u^\infty \equiv (u_1, u_2, ...)$ generates a sample $(s_1, s_2, ...)$ of outcomes, where $s_i \in G(u_i \mid \theta)$ for all i. Though seemingly standard and innocuous, the assumption that (s_i) is an i.i.d. sample becomes subtle given incompleteness of the model and the declared agnosticism about the selection mechanism. This is because if the selection mechanism in each market is not understood, then there is no basis for taking a stand on how such selections are related to each other across experiments. To emphasize this point further, think of the nonsingleton nature of $G(u_i \mid \theta)$ in terms of "omitted variables:" a complete theory may exist in principle in that it may be possible to explain and predict selection given a suitable set of explanatory variables. However, the analyst's theory does not identify these omitted variables. They are not only unobservable to her, as are the latent variables captured by U–more fundamentally, their identity is unknown. Consequently, there is no basis for understanding how selection, and thus realized outcomes, may differ or be related across experiments.

In this paper, we develop a new inference method that is robust to *heterogeneity and dependence of an unknown form.* We outline our approach here leaving technical details and formal results for the sequel. The first step is to specify the set of outcome sequences that are consistent with what is known according to the analyst's theory. For each given θ, robustness to an unknown form of dependence implies that if for each i, s_i is a conceivable outcome in the ith experiment (in isolation) given u_i, then $(s_1, s_2, ...)$ is a conceivable sequence given $(u_1, u_2, ...)$. Thus, without further assumptions, the model predicts that the (selected) outcomes $(s_1, s_2, \cdots)$ take their values in the Cartesian product of $G(u_i|\theta)$, $i = 1, 2, ...$, and we define:

$$G^\infty (u_1, ..., u_i, ... \mid \theta) \equiv \Pi_{i=1}^{\infty} G(u_i \mid \theta). \tag{1.1}$$

Note that experiments are indistinguishable in the sense that the same correspondence $G(\cdot \mid \theta)$ applies to each experiment. However, even if $G(u_i \mid \theta) = G(u_j \mid \theta)$, as when $u_i = u_j$, any outcome in $G(u_i \mid \theta)$ is possible in experiment i and any possibly different outcome is possible in experiment j. Therefore, the model, expanded in this way to sequences, does not restrict how selection might differ or be related across experiments.

The second step is to add a suitable stochastic structure that again leaves the heterogeneity and dependence structure of selections unrestricted. Fix θ. Assume

[1] For example, each experiment could correspond to a different market where an entry game is played.

that u^∞ jointly follows a parametric distribution m_θ^∞, the i.i.d. product of the measure m_θ on U. For each given u^∞, any probability distribution P_{u^∞} supported on $G^\infty(u^\infty \mid \theta)$ is a valid conditional distribution of the sequence of outcomes; and the implied distribution of outcomes is $P = \int P_{u^\infty} dm_\theta^\infty$. Accordingly, we consider the entire set $\mathcal{P}_\theta$ of distributions over outcomes given by

$$\mathcal{P}_\theta = \left\{ P \in \Delta(S^\infty) : P = \int_{U^\infty} P_{u^\infty} dm_\theta^\infty(u^\infty),\ P_{u^\infty} \in \Delta(G^\infty(u^\infty \mid \theta))\ \ m_\theta^\infty(\cdot)\text{-}a.s. \right\}.$$

Note that because $\Delta(G^\infty(u^\infty \mid \theta))$ equals the entire simplex of distributions on $\Pi_{i=1}^\infty G(u_i \mid \theta)$, including both nonidentical product measures and nonproduct measures, the set $\mathcal{P}_\theta$ accommodates many forms of heterogeneity and dependence across experiments even given u^∞.

Though sets of probability measures may not seem to be convenient vehicles for conducting inference, the set $\mathcal{P}_\theta$ has a special structure that makes it tractable: its lower envelope, $\nu_\theta^\infty(\cdot)$ defined (for every measurable $B \subset S^\infty$) by

$$\nu_\theta^\infty(B) = \inf_{P \in \mathcal{P}_\theta} P(B), \tag{1.2}$$

is a *belief function* on S^∞.[2] We exploit this and prove a (new) central limit theorem (CLT) for each belief function ν_θ^∞ and thus indirectly also for each set $\mathcal{P}_\theta$. Then we show how this CLT can be used to construct suitably robust confidence regions for the unknown parameter θ.

A confidence region $\mathcal{C}_n$ is a set of parameter values constructed from a finite number of observations $s_1, ..., s_n$ such that, for each θ, the coverage probability is asymptotically at least at a prespecified level $1-\alpha$ *under any probability distribution in* $\mathcal{P}_\theta$. We construct $\mathcal{C}_n$ using a statistic based on the empirical frequencies $n^{-1}\sum_{i=1}^n I(s_i \in A_j)$ for a class $\{A_j\}_{j=1}^J$ of subsets of S. Then we use the CLT to prove that $\nu_\theta^\infty(\{\theta \in \mathcal{C}_n\}) \to 1-\alpha$, which implies that $\mathcal{C}_n$ controls the asymptotic coverage probability uniformly over $\mathcal{P}_\theta$. Furthermore, we show that the coverage is uniform over the generalized parameter space $\mathcal{F} = \{(\theta, P) : P \in \mathcal{P}_\theta, \theta \in \Theta\}$; that is, our confidence region satisfies

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{F}} P(\theta \in \mathcal{C}_n) \geq 1-\alpha.$$

[2] Belief functions are special cases of capacities (or "non-additive probabilities"), sometimes referred to as totally, completely, or infinitely monotone capacities. They originated in Dempster (1967) and Shafer (1976). Definitions for more general settings can be found, for example, in Philippe, Debs and Jaffray (1999), and Molchanov (2005).

A notable feature of our confidence region is that, in contrast to existing methods, its construction does not require tuning parameters. This is due to the different procedure used to approximate the (worst-case) probability that the confidence region covers θ. As we show below, the model implies that asymptotically the probability of any set of outcomes $A \subset S$ lies in a probability interval $[\nu_\theta(A), \nu_\theta^*(A)]$ that depends on θ. Under the assumption adopted in existing methods that the outcomes s_i are i.i.d., the empirical frequency $n^{-1}\sum_{i=1}^n I(s_i \in A)$ converges to a unique probability $p(A)$ asymptotically; and the pointwise limiting distributions of the test statistics used to construct confidence regions change depending on whether $p(A)$ equals $\nu_\theta(A)$ or $\nu_\theta^*(A)$, or is in the interior of the interval.[3] This creates a discontinuity of the limiting distribution in the underlying data generating process. A sequence of tuning parameters is commonly used to handle this discontinuity. However, though the choice of tuning parameters often affects the performance of existing methods in non-trivial ways, their optimal choice remains a difficult problem.

In contrast, we do not presume the existence of such unique limits. Even so, inference on the structural parameter is possible because if θ is the true parameter, then the empirical frequency cannot deviate from the above probability interval asymptotically. Our CLT provides a normal approximation to the distribution of deviations from this restriction in finite samples. This normal approximation is expressed in terms of the lower envelope over all possible data generating processes, and thus the true data generating process does not affect the approximation given θ. This eliminates the discontinuity of the limiting distribution.

After describing some links to the literature in the remainder of this introduction, the paper proceeds as follows. Section 2 lays out the formal framework. The latter is used in Section 3 which presents our results regarding inference. Examples and some Monte Carlo simulation results follow in the next two sections. To this point, the analysis is carried out under the assumption that there is no

[3] For example, a commonly used test statistic $T_n(\theta) = \sqrt{n}\max\{\nu_\theta(A) - n^{-1}\sum_{i=1}^n I(s_i \in A), n^{-1}\sum_{i=1}^n I(s_i \in A) - \nu_\theta^*(A)\}$ converges in distribution to

$$T(\theta) = \begin{cases} -Z & \text{if } \nu_\theta(A) = p(A) < \nu_\theta^*(A) \\ -\infty & \text{if } \nu_\theta(A) < p(A) < \nu_\theta^*(A) \\ Z & \text{if } \nu_\theta(A) < p(A) = \nu_\theta^*(A) \\ \max\{-Z, Z\} & \text{if } \nu_\theta(A) = p(A) = \nu_\theta^*(A), \end{cases}$$

where Z is the limiting distribution of $\sqrt{n}(n^{-1}\sum_{i=1}^n I(s_i \in A) - p(A))$ under the i.i.d. assumption.

observable heterogeneity across experiments. Section 6 describes an extension to include covariates. Appendices contain proofs as well as an outline of an extension that robustifies the treatment of latent variables, and also details regarding implementation.

1.2. Relation to the literature

Using the theory of random sets, the existing literature has shown that the stochastic behavior of s_i in each experiment can be characterized by capacities. Capacities have been employed to characterize the set of parameter values that are identifiable from the observed variable (Galichon and Henry 2011, Beresteanu, Molchanov, and Molinari 2011). For example, Galichon and Henry (2011) use the capacity defined by $\mu_\theta(A) \equiv m_\theta(G(u|\theta) \cap A \neq \emptyset)$, $A \subset S$, as a primitive object to conduct their identification analysis. This functional gives, for each single experiment, the upper envelope of the probability of A over the set of distributions compatible with the model. Here we focus on the entire sequence of experiments jointly, and we use another capacity, the belief function. This choice is made because the belief function gives the lower envelope, which is relevant for studying the robust control of the asymptotic coverage probability.

Our approach to inference is related to Beresteanu and Molinari (2008) in the sense that we both use generalized limit theorems. But theirs is for set-valued random variables having probabilistic distributions (Molchanov 2005), while we use limit laws for capacities that are generated by set-valued random variables; this difference accords with their focus on inference about the identified set as opposed to the true parameter. Another difference is that they assume that, translated into our setting, the entire set of outcomes is observed for each experiment rather than merely the selected outcome (for example, outcomes are interval-valued). In addition, they adopt the counterpart of the i.i.d. assumption discussed above. Galichon and Henry (2009, 2011) study inference using a statistic based on capacities, but they also maintain the i.i.d. assumption.

In various incomplete models, structural parameters often satisfy model restrictions that take the form of moment inequalities. Therefore, econometric tools for moment inequalities have been used to make inference for incomplete models (Chernozhukov, Hong, and Tamer 2007, Andrews and Soares 2010, Bugni 2009, Andrews and Shi 2013). Although these methods do not preclude data heterogeneity and dependence per se,[4] it is commonly assumed that data are generated iden-

[4] Andrews and Soares (2010) extend their framework and give conditions under which their

tically and independently across experiments which precludes robustness against potential heterogeneity and dependence due to model incompleteness. Though the method we develop here is applicable to the narrower class of incomplete structural models, it has the advantage of being robust.

Bresnahan and Reiss (1990, 1991) consider an identification and estimation method that is robust to the multiplicity of equilibria. Their strategy is to transform the outcome variable so that the model becomes complete after the transformation. Since this transformation aggregates some of the outcomes that can be selected from multiple equilibria, it incurs a loss of information.

Belief functions play a central role in Epstein and Seo (2015), who describe a Bayesian-style approach to doing inference in incomplete models. Besides their subjectivist as opposed to frequentist approach, their paper differs also in its focus on axiomatic decision-theoretic foundations.

In the literature, much attention is paid to the "identified set." Because readers may wonder why it does not play a role here, we discuss it briefly. Following Manski (2003), the identified set is taken to be the set of parameters compatible with what is revealed asymptotically by the sampling process. Given the structure $(S, U, G, \Theta; m)$ augmented by the assumption that outcome sequences are distributed i.i.d. with some measure $p \in \Delta(S)$, then empirical frequencies converge almost surely to p, rendering p observable. The identified set, denoted $\Theta^I(p)$, consists of all θ such that there exists a (suitably measurable) selection rule $u \longmapsto p_u \in \Delta(G(u \mid \theta))$ satisfying[5]

$$p(\cdot) = \int_U p_u(\cdot)\, dm_\theta(u),$$

which equates true and predicted empirical frequencies. A number of papers describe (finite sample) estimators for Θ^I; see, for example, Ciliberto and Tamer (2009). From our perspective, such a focus on $\Theta^I(p)$ is unjustified since both its definition and interpretation presume that outcomes are i.i.d. which we have argued is problematic when the analyst's model is incomplete. When robustness with respect to unknown forms of heterogeneity and dependence is sought, it is apparent that the appropriate definition of an identified set should be formulated in

inference method is applicable to dependent data. However, in our understanding the main goal of this extension is to handle more general data (e.g. time series data) rather than to make inference robust to the heterogeneity and dependence due to incompleteness.

[5] See Beresteanu, Molchanov and Molinari (2011) and Galichon and Henry (2011), for example. The latter show that $\Theta^I(p)$ is equal to the set of all parameters θ such that p is contained in the core of the belief function ν_θ on S–see(2.4) below.

the space of outcome sequences. However, we do not pursue such a definition here because it does not seem vital for studying inference about the true parameter.

2. The framework

Consider a setting with an infinite sequence of experiments (or random events), where $S_i = S$ (finite) denotes the set of possible outcomes for the ith experiment. The economic model of each single experiment is described by $(S, U, G, \Theta; m)$ with the following interpretation and restrictions. Θ is a set of structural parameters. The true parameter is common to all experiments but is unknown to the analyst. Each u in U describes the unobservable characteristics of the single experiment under consideration. In alternative terminology, S and U capture endogenous and latent variables respectively; an extension to include covariates describing observable heterogeneity is provided in Section 6. We assume that U is a Polish (complete separable metric) space. Latent variables are distributed according to the Borel probability measure m_θ, which may depend on θ; let $m = (m_\theta)_{\theta\in\Theta}$. Finally, for each $\theta \in \Theta$, $G(\cdot \mid \theta) : U \rightsquigarrow S$ is a correspondence that describes the set of outcomes for each given u and parameter θ. The multi-valued nature of G gives one sense in which the analyst's model (or theory) is incomplete: for each single experiment, and given the structural parameter, theory prescribes only a set of possible outcomes, with no indication of which outcomes in the set are more or less likely to be selected. We assume that, for each θ, $G(\cdot \mid \theta)$ is weakly measurable.[6]

The analyst observes outcomes in some experiments and wishes to draw inferences, via the construction of confidence regions for the structural parameters. To address inference, we extend the above formal structure to accommodate the entire sequence of experiments. Accordingly, consider the tuple $(S^\infty, U^\infty, G^\infty, \Theta; m^\infty)$. The meaning of and rationale for S^∞ and U^∞ are clear;[7] they have generic elements $s^\infty = (s_1, s_2, ...)$ and $u^\infty = (u_1, u_2, ...)$ respectively. By m^∞, an abbreviation for $(m_\theta^\infty)_{\theta\in\Theta}$, we mean that, conditional on θ, unobserved variables are distrib-

[6] A correspondence $\Gamma : U \rightsquigarrow X$, where X is metric, is weakly measurable if $\{u : \Gamma(u) \subset A\}$ is a (Borel) measurable subset of U for every closed $A \subset X$. If Γ is compact-valued, then weak measurability is equivalent to the property that $\{u : \Gamma(u) \subset A\}$ is measurable for every open $A \subset X$ (Aliprantis and Border 2006, Lemma 18.2).

[7] For any metric space X, we endow X^∞ with the product metric and corresponding Borel σ-algebra. (Then S^∞ is separable compact metric, and hence Polish). We denote by $\Delta(X)$ the set of Borel (countably additive) probability measures on X.

uted i.i.d. across experiments according to m_θ. The parameter set Θ remains unchanged and parameters are assumed to be constant across experiments. The remaining component G^∞, a key to our approach, is, for each θ, a correspondence $G^\infty(\cdot \mid \theta) : U^\infty \rightsquigarrow S^\infty$ defined as in (1.1). As described there, the Cartesian product structure in (1.1) imposes no restrictions on how selection might differ or be related across experiments. This is another important sense of model incompleteness. Note that $G^\infty(\cdot \mid \theta)$ is weakly measurable by Aliprantis and Border (2006, Lemma 18.4); it is also compact-valued.

In seeking robust inferences, the analyst takes into account ALL probability distributions $P \in \Delta(S^\infty)$ that are consistent with the given $(S^\infty, U^\infty, G^\infty, \Theta; m^\infty)$, that is, for each given θ, she considers the set $\mathcal{P}_\theta$ defined in the introduction and repeated here for convenience:

$$\mathcal{P}_\theta = \left\{ P \in \Delta(S^\infty) : P = \int_{U^\infty} P_{u^\infty} dm_\theta^\infty(u^\infty),\ P_{u^\infty} \in \Delta(G^\infty(u^\infty \mid \theta)) \quad m_\theta^\infty(\cdot)\text{-}a.s. \right\}. \tag{2.1}$$

Each indicated conditional distribution P_{u^∞} is assumed to be such that $u^\infty \longmapsto P_{u^\infty}(B)$ is measurable for every measurable $B \subset S^\infty$, and is referred to as a *selection rule*. When the analyst's model is complete, ($G^\infty(\cdot \mid \theta)$ is single-valued), then $\mathcal{P}_\theta = \{P_\theta\}$ is a singleton and P_θ is the i.i.d. product of the measure on S induced by m_θ and $G(\cdot \mid \theta) : U \to S$. However, in general, she considers all (including non i.i.d.) selection rules consistent with her incomplete theory.

Related structures appear, for example, in Koopmans and Reiersol (1950) and Jovanovic (1989), and are employed by Galichon and Henry (2009, 2013) in constructing confidence regions given partial identification. *These papers differ from ours in how they use the single experiment structure* $(S, U, G, \Theta; m)$ *when considering sequences.* In particular, the inference procedures described by Galichon and Henry (and in most of the ambient literature) rely on the assumption that there is a true probability law p^* on S that, though unknown, can be approximated arbitrarily well by the empirical distribution for large samples *because experimental outcomes are taken to be i.i.d. according to* p^*. Such approaches do not deliver robustness against incompleteness of the analyst's model: she cannot rely on a single probability law over S^∞ because, for example, if there are omitted variables that influence selection, then the distribution of experimental outcomes will depend on how those omitted variables play out across experiments, for which the analyst's theory provides no guidance.

More formally, an i.i.d. law on S^∞ can be justified if one limits attention to selection rules P_{u^∞} in (2.1) of the form:[8]

$$P_{u^\infty} = \otimes_{i=1}^{\infty} p_{u_i},$$

where each $p_u \in \Delta\left(G\left(u \mid \theta\right)\right)$ is a measure on S that describes probabilistic selection within each market. Then the induced measure P on S^∞ is i.i.d. because

$$P = \int_{U^\infty} \left(\otimes_{i=1}^{\infty} p_{u_i}\right) dm_\theta^\infty\left(u^\infty\right) = \otimes_{i=1}^{\infty}\left(\int_U p_{u_i} dm_\theta\left(u_i\right)\right),$$

which is the i.i.d. product measure generated by $\int_U p_u dm_\theta\left(u\right)$ (it plays the role of p^* above). Importantly, this i.i.d. property is derived, in particular, from the assumption that selection in each market i depends on u_i but not otherwise on the identity of the market, that is, $p_{u_i} = p_{u_j}$ if $u_i = u_j$, thereby precluding that selection might be affected by omitted variables. Such an assumption is convenient–as the literature has demonstrated, it permits one to do inference. However, we show that robust inference procedures also exist.[9]

The structure of the set $\mathcal{P}_\theta$ defined in (2.1) implies a form of symmetry across experiments that warrants explicit mention. Roughly, it indicates that the ordering of experiments has no significance in the following sense. For any finite permutation π of the indices $1, 2, ...,$ and any probability measure P on S^∞, denote by πP the unique probability measure satisfying (for all rectangles) $\left(\pi P\right)\left(A_1 \times A_2 \times ...\right) = P\left(A_{\pi^{-1}(1)} \times A_{\pi^{-1}(2)} \times ...\right)$. Then it is easy to see that

$$P \in \mathcal{P}_\theta \iff \pi P \in \mathcal{P}_\theta. \tag{2.2}$$

Such symmetry seems more natural in a cross-sectional setting where experiments are resolved simultaneously than in a time-series context where experiments are differentiated because they are ordered temporally. Accordingly, though the formal results that follow do not require the cross-sectional interpretation, we think of our approach to inference as particularly relevant to cross-sectional data. When considering symmetry, keep in mind that currently we are ruling out observable differences between experiments. When these are included and modeled via covariates as in Section 6, then the implied symmetry is suitably conditional–roughly,

[8] $\otimes_{i=1}^{\infty} p_{u_i}$ is the product measure defined by the p_{u_i}s.

[9] Menzel (2011) assumes that outcome samples are drawn from an exchangeable rather than i.i.d. distribution, which delivers some robustness. However, he restricts selection to depend only on variables that affect payoffs, and thus his method is not robust against the effects of unknown omitted variables.

(2.2) is weakened so as to apply only to permutations that permute experiments having common covariate values.

Finally, a feature of $\mathcal{P}_\theta$ that we exploit heavily is its connection to a belief function, which we now explain. Define $\nu_\theta^\infty(\cdot)$ to be the lower envelope of $\mathcal{P}_\theta$ as in (1.2). Then ν_θ^∞ can also be expressed in the form: For every measurable $B \subset S^\infty$,

$$\nu_\theta^\infty (B) \equiv m_\theta^\infty (\{u^\infty \in U^\infty : G^\infty (u^\infty \mid \theta) \subset B\}) . \tag{2.3}$$

Thus ν_θ^∞ is the capacity on measurable subsets of S^∞ induced by the correspondence $G^\infty (\cdot \mid \theta)$ and the probability measure m_θ^∞ on U^∞, which is in the form of one of the common definitions of a belief function.

Remark 1. *Here are some details supporting the preceding claims. Because these are well-known in the literature (see, for example, Aliprantis and Border (2006, Ch. 18) and Philippe, Debs and Jaffray (1999)), we provide only an outline here rather than a formal lemma. The set* $\{u^\infty \in U^\infty : G^\infty (u^\infty \mid \theta) \subset B\}$ *in (2.3) is in general not measurable for every Borel measurable* B*. However, it is universally measurable, and moreover, each Borel measure* m_θ^∞ *has a unique extension to a probability measure (also denoted* m_θ^∞*) on the collection of all universally measurable subsets of* S^∞*. This renders the RHS of (2.3) well-defined. In addition, it follows from Philippe, Debs and Jaffray (1999, Theorem 3) that (2.3) and (1.2) provide equivalent definitions of* ν_θ^∞*.*

One can proceed similarly when considering a single experiment in isolation. Then the set of all probability laws on any single experiment that are consistent with θ and the given structure $(S, U, G, \Theta; m)$ is given by

$$\left\{ p \in \Delta (S) : p = \int_U p_u dm_\theta (u) \, , \, p_u (G (u \mid \theta)) = 1 \ m_\theta\text{-}a.s. \right\} .$$

If we define ν_θ on S as the lower envelope of this set, then

$$\nu_\theta (A) \equiv m_\theta (\{u \in U : G (u \mid \theta) \subset A\}) \, , \, A \subset S, \tag{2.4}$$

from which we can conclude that ν_θ is a belief function on S. The upper envelope of the set of consistent measures is also of interest. Thus define also the conjugate of ν_θ, denoted ν_θ^*, by

$$\nu_\theta^* (A) = 1 - \nu_\theta (S \backslash A) . \tag{2.5}$$

Then $\nu_\theta^*(A)$ is the maximum probability of A consistent with the model. Of course,

$$\nu_\theta(\cdot) \leq \nu_\theta^*(\cdot).$$

A final comment is that, in common with all the surrounding literature, our framework treats asymmetrically the uncertainty generated by latent variables as opposed to the uncertainty regarding selection–the former is described by a single probability measure (for each θ) while there is complete ignorance about the latter. One may question the assumption of extensive knowledge of latent variables particularly since they are not observed by the analyst. However, contrary to appearances, our framework also permits the analyst to have an incomplete model of latent variables. Formally, one can take each m_θ to be a belief function on U, and the approach to inference that follows carries through. See Appendix E for details.

3. Inference

Here we construct confidence regions for the unknown parameters that are robust to the limitations of the analyst's model. The approach largely mimics the classical approach used when $\mathcal{P}_\theta$ is a singleton i.i.d. measure, where the classical CLT can be used to construct desired confidence regions. We show that a corresponding procedure can be adopted also when the analyst's model is incomplete. The first step is to establish (in Theorem 3.1) a CLT for belief functions ν_θ^∞. The coverage property of our confidence regions is then established in Theorem 3.2.

3.1. A central limit theorem

Belief functions aid tractability because they permit extensions of some basic tools of probability theory, namely the LLN and CLT. The former is taken from Maccheroni and Marinacci (2005), while the CLT is original to this paper and is described shortly. Both rely on the fact that, in a suitable sense, ν_θ^∞ is an "i.i.d. product" of ν_θ, which explains also our notation ν_θ^∞ for the belief function on S^∞.

Let $\Psi_n(s^\infty)(\cdot)$ be the empirical frequency measure in the first n experiments along the sample $s^\infty = (s_1, s_2, ...)$, that is,

$$\Psi_n(s^\infty)(A) = \frac{1}{n}\Sigma_{i=1}^n I(s_i \in A), \text{ for every } A \subset S.$$

Though empirical frequencies need not converge, the LLN asserts certainty that asymptotically $\Psi_n(s^\infty)(A)$ lies in the interval $[\nu_\theta(A), \nu_\theta^*(A)]$:

$$\nu_\theta^\infty\{s^\infty : [\liminf \Psi_n(s^\infty)(A), \limsup \Psi_n(s^\infty)(A)] \subset [\nu_\theta(A), \nu_\theta^*(A)]\} = 1; \quad (3.1)$$

and this condition is tight in the sense that

$$\begin{aligned} \nu_\theta^\infty(\{s^\infty : \nu_\theta(A) < \liminf \Psi_n(s^\infty)(A)\}) &= 0, \text{ and} \\ \nu_\theta^\infty(\{s^\infty : \limsup \Psi_n(s^\infty)(A) < \nu_\theta^*(A)\}) &= 0. \end{aligned} \quad (3.2)$$

In light of the lower envelope condition (1.2), the LLN asserts that the event in (3.1) has unit probability according to *every* measure in $\mathcal{P}_\theta$, while each event appearing in (3.2) has arbitrarily small probability according to *some* measure in $\mathcal{P}_\theta$.

Turn to the CLT. For any positive semidefinite matrix $\Lambda \in \mathbb{R}^{J \cdot J}$, $\boldsymbol{N}_J(.;\Lambda)$ denotes the J-dimensional normal cdf with zero mean and covariance matrix Λ– for any $c = (c_1, ..., c_J) \in \mathbb{R}^J$, $\boldsymbol{N}_J(c;\Lambda)$ is the probability mass associated with values less than or equal to c (in the vector sense), that is, with the closed lower orthant at c. Of primary interest will be covariance matrices constructed as follows. Fix J events, $A_1, ..., A_J$, subsets of S, and for any θ, let

$$cov_\theta(A_i, A_j) = \nu_\theta(A_i \cap A_j) - \nu_\theta(A_i)\,\nu_\theta(A_j), \quad (3.3)$$

$$var_\theta(A_i) = \nu_\theta(A_i)(1 - \nu_\theta(A_i)) = cov_\theta(A_i, A_i).$$

Denote by Λ_θ the $J \times J$ symmetric and positive semidefinite matrix $(cov_\theta(A_i, A_j))$.[10]

Theorem 3.1. *Let $\Lambda_{\theta_n} \to \Lambda \in \mathbb{R}^{J \cdot J}$ and $c_n \to c \in \mathbb{R}^J$. Then*

$$\nu_{\theta_n}^\infty\left(\cap_{j=1}^J \left\{s^\infty : \sqrt{n}\left[\nu_{\theta_n}(A_j) - \Psi_n(s^\infty)(A_j)\right] \le c_{nj}\right\}\right) \to \boldsymbol{N}_J(c;\Lambda). \quad (3.4)$$

See Appendix A for a proof.[11]

Though the inequalities in (3.4) place only a lower bound on empirical frequencies, upper bounds are also accommodated. To demonstrate this and to facilitate interpretation of the CLT, suppose that $J = 2I$ and that $A_{I+i} = S \backslash A_i$ for each

[10] Positive semidefiniteness is proven in the theorem.

[11] Marinacci (1999, Theorem 16) proves a central limit theorem for a class of capacities that he calls "controlled," which property neither implies nor is implied by being a belief function. Thus the CLTs are not comparable. Marinacci does not study confidence regions.

$i=1,...,I$, that is, each event A_i is accompanied by its complement A_{I+i}; in this case we refer to $\{A_j\}$ as being "complement-closed." Then the event appearing in (3.4) is

$$\cap_{i=1}^{I}\left\{-c_{ni}/\sqrt{n}+\nu_{\theta_n}\left(A_i\right)\leq\Psi_n\left(s^{\infty}\right)\left(A_i\right)\leq\nu_{\theta_n}^{*}\left(A_i\right)+c_{n(I+i)}/\sqrt{n}\ \right\}, \tag{3.5}$$

where $\nu_{\theta_n}^{*}$ is the conjugate belief function defined as in (2.5). For greater clarity, suppose further that $(\theta_n,c_n)=(\theta,c)$ for all n. Then, rather than certainty that the empirical frequency of A_i in an infinite sample lies in the interval $[\nu_\theta\left(A_i\right),\nu_\theta^{*}\left(A_i\right)]$, as in the LLN, the CLT describes, as an approximation, the distribution of deviations from that restriction in finite samples. In particular, when c_i and c_{I+i} are positive, the empirical frequency can be smaller than $\nu_\theta\left(A_i\right)$ or larger than $\nu_\theta^{*}\left(A_i\right)$, and the distribution of such deviations according to ν_θ^{∞} is approximately normal.

When each ν_{θ_n} is additive and hence a probability measure, then the variances and covariances defined in (3.3) are the usual notions applied to indicator functions $I(s\in A_i)$ and $I(s\in A_j)$ and the CLT reduces to (a special case of) the classical triangular CLT (see, for example, White (2001, Theorem 5.11)). Other special cases of the theorem are also immediate implications of classical results. For example, if $J=1$, then the CLT provides an approximation to

$$\nu_{\theta_n}^{\infty}\left(\left\{-c_{n1}/\sqrt{n}+\nu_{\theta_n}\left(A_1\right)\leq\Psi_n\left(s^{\infty}\right)\left(A_1\right)\right\}\right). \tag{3.6}$$

But it can be shown that for this event the minimum in (1.2) is achieved at an i.i.d. measure P_n^{*}.[12] Thus one can invoke a classical triangular CLT. However, in general, reduction to the classical additive case is not elementary because even if minimizing measures exist, they are not easily determined nor is there any reason to expect that they are i.i.d. measures.

The proof of our general result exploits the close connection between belief functions and probability measures expressed in (2.3), and also the Cartesian product structure of G^{∞} given in (1.1). Together they permit, for each θ_n, transforming our assertion about belief functions into one about i.i.d. probability measures $m_{\theta_n}^{\infty}$ as follows:

$$\begin{aligned}&\nu_{\theta_n}^{\infty}\left(\sqrt{n}\left(\nu_{\theta_n}\left(A_j\right)-\Psi_n\left(s^{\infty}\right)\left(A_j\right)\right)\leq c_{nj}\text{ for each }j\right)\\&\qquad=m_{\theta_n}^{\infty}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nu_{\theta_n}\left(A_j\right)-X_{ni}^{j}\right)\leq c_{nj}\text{ for each }j\right),\end{aligned} \tag{3.7}$$

[12] P_n^{*} is the i.i.d. product of $p_n^{*}\in\Delta\left(S\right)$ such that $p_n^{*}\left(A_1\right)=\nu_{\theta_n}\left(A_1\right)$. When a minimizer exists in (1.2) for an event, refer to it as a minimizing or worst-case measure for that event.

where for each j, $X_{ni}^j = I(G(u_i|\theta) \subset A)$, $i = 1, \cdots, n$, is an i.i.d. sequence of random variables. Then the classical CLT can be applied. Note that despite the fact that the distribution of the sequence of outcomes involves incidental parameters P_{u^∞} describing selection, the fact that selection can vary arbitrarily across markets does not affect our limit theorem. This is because each belief function $\nu_{\theta_n}^\infty$ is a lower envelope (1.2) as one varies over all possible selections, which set is described by the i.i.d. set-valued random variable $G(\cdot \mid \theta_n)$. Consequently, the (selection) incidental parameters do not enter into the representation of belief functions as in (2.3).

We also note that the assumption that $m_{\theta_n}^\infty$ is i.i.d. (for each θ_n) may be relaxed, that is, one can establish a CLT similar to Theorem 3.1 while allowing for heterogeneity and dependence of a known form for $m_{\theta_n}^\infty$. This is because, in light of (3.7), as long as the sequence of random vectors $X_{ni} = (X_{ni}^1, \cdots, X_{ni}^J)'$, $i = 1, \cdots, n$, obey a suitable central limit theorem under $m_{\theta_n}^\infty$, such an extended result becomes available.[13]

3.2. Confidence regions

Fix $0 < \alpha < 1$ and $A_1, \cdots, A_J$, subsets of S. For each θ, let Λ_θ be the $J \times J$ covariance matrix defined as above, and let

$$\sigma_\theta \equiv \left(\sqrt{var_\theta (A_1)}, ..., \sqrt{var_\theta (A_J)} \right). \tag{3.8}$$

Our confidence region $\mathcal{C}_n$ is given by

$$\mathcal{C}_n = \left\{ \theta \in \Theta : \nu_\theta (A_j) - \Psi_n (s^\infty) (A_j) \leq c_\theta \sqrt{var_\theta (A_j)/n},\ j = 1, ..., J \right\}, \tag{3.9}$$

where[14]

$$c_\theta = \min \left\{ c \in \mathbb{R}_+ : \boldsymbol{N}_J (c\sigma_\theta; \Lambda_\theta) \geq 1 - \alpha \right\}. \tag{3.10}$$

Note that $\mathcal{C}_n$ is based on a normalized Kolmogorov-Smirnov-type statistic, because it equals $\{\theta \in \Theta : T_n(\theta) \leq c_\theta\}$, where $T_n(\theta)$ is the maximum of the normalized empirical frequencies $T_{j,n}(\theta) \equiv (\nu_\theta (A_j) - \Psi_n (s^\infty) (A_j)) / \sqrt{var_\theta(A_j)/n}$, $j =$

[13]For example, Jenizh and Prucha's (2009) central limit theorem for arrays of random fields allows variables to have spatial correlations.

[14]The proof of the next theorem shows that c_θ is well-defined. If $\sigma_\theta = 0$, then $\boldsymbol{N}_J (0; \Lambda_\theta)$ refers to a degenerate distribution at the mean, which is 0, and thus $\boldsymbol{N}_J (c\sigma_\theta; \Lambda_\theta) = 1$ for all $c \geq 0$, and $c_\theta = 0$.

$1, \cdots, J$, where we take $1/0 = \infty$, $0/0 = 0$ and $-1/0 = -\infty$. Here, $var_\theta(A_j)$ is equal to 0 if and only if $\nu_\theta(A_j) = 0$ or 1. If $\nu_\theta(A_j) = 0$, then $T_{j,n}(\theta) = -\infty$ and event A_j does not provide any restriction on θ. If $\nu_\theta(A_j) = 1$, then θ is excluded from the confidence region whenever $\Psi_n(s^\infty)(A_j) < 1$, ($T_{j,n}(\theta) = \infty$ in this case), while it is included in the confidence region if $\Psi_n(s^\infty)(A_j) = 1$ ($T_{j,n}(\theta) = 0$ in this case) and $T_{k,n}(\theta) \leq c_\theta$ for all $k \neq j$.

The asymptotic coverage property of $\mathcal{C}_n$ is established by the following theorem.

Theorem 3.2. *Let $0 < \alpha < 1$. Then*

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \nu_\theta^\infty \left(\{s^\infty : \theta \in \mathcal{C}_n\}\right) \geq 1 - \alpha. \tag{3.11}$$

Further, there is equality in (3.11) if $\alpha < 1/2$ and $\Lambda_\theta \neq 0$ for some $\theta \in \Theta$.

Since $\mathcal{P}_\theta$ is the set of all probability laws consistent with the model and θ and since ν_θ^∞ gives the lower envelope of $\mathcal{P}_\theta$, the theorem establishes that if θ is the "true value" of the parameter, then, in the limit for large samples, $\mathcal{C}_n$ contains θ with probability at least $1 - \alpha$ *according to every probability law that is consistent with the model and θ*. Moreover, (3.11) can also be stated as $\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{F}} P(\theta \in \mathcal{C}_n) \geq 1 - \alpha$, where $\mathcal{F} = \{(\theta, P) : P \in \mathcal{P}_\theta, \theta \in \Theta\}$. Thus our coverage statement is uniform on the general parameter space $\mathcal{F}$. Finally, the noted coverage is tight in the sense of equality in (3.11) if (as one would expect) $\alpha < 1/2$, and if we exclude the very special case where $\sigma_\theta = 0$ for all $\theta \in \Theta$, that is, where $\nu_\theta(A_j) \in \{0, 1\}$ for all j and θ.[15]

The confidence regions and their coverage properties are discussed further in the next section in the context of examples.

4. Examples

4.1. Discrete normal form games

A widely studied class of games in the applied literature is the class of entry games with multiple Nash equilibria. Here we focus on the canonical example from Jovanovic (1989), because it illustrates simply the main issues and because it is used widely for that purpose in the ambient literature. However, the reader

[15] Note that because Λ_θ is positive semidefinite, $\sigma_\theta = 0$ if and only if $\Lambda_\theta = 0$.

will likely realize that our analysis accommodates a much broader class of games– more on this after outlining how the Jovanovic game is accommodated.

In the Jovanovic entry game, in each market two firms play the entry game described by the following payoff matrix:

	out	in
out	$0,0$	$0,-u_2$
in	$-u_1,0$	$\theta^{1/2}-u_1,\theta^{1/2}-u_2$

The parameter θ lies in $[0,1]$ and $u=(u_1,u_2)$ is observed by players but not by the analyst. She views θ as fixed and common across markets and u as uniformly distributed on $[0,1]^2$ and i.i.d. across markets. Her theory is that the outcome in each market is a pure strategy Nash equilibrium. However, her theory is incomplete because she does not understand equilibrium selection. Thus the translation into our set up has: $S=\{0,1\}$, where 0 (1) indicates that no (both) firms enter the market; $\Theta=[0,1]$; $U=[0,1]^2$; m independent of θ and uniform on $[0,1]^2$; and G equal to the (pure strategy) Nash equilibrium correspondence given by

$$G\left(u_1,u_2\mid\theta\right)=\begin{cases}\{0,1\} & \text{if } 0\le u_1,u_2\le\theta^{1/2}\\ \{0\} & \text{otherwise.}\end{cases}\tag{4.1}$$

The implied set of distributions over S consists of all probability measures for which the probability of $s=1$ lies in $[0,\theta]$. This interval of probabilities is equivalently represented by the belief function ν_θ, where

$$\nu_\theta\left(1\right)=0,\ \nu_\theta\left(0\right)=1-\theta,\ \nu_\theta\left(\{0,1\}\right)=1.$$

Turn to inference about θ. Suppose first that $J=1$ and $A_1=\{1\}$. Then, for all θ, $\nu_\theta\left(1\right)=0$ and $\sigma_\theta=0$. It follows that $\mathcal{C}_n=\Theta=[0,1]$. Thus without making use of the (implied) sample frequency of $s=0$, observations of $s=1$ alone do not provide any information about the unknown parameter θ.

Suppose, however, that ($J=2$ and) we use also the sample frequency of $A_2=\{0\}$. Then, for each θ, $\nu_\theta\left(0\right)=1-\theta$ and $\sigma_\theta=\left(0,\left[\theta\left(1-\theta\right)\right]^{1/2}\right)$, and therefore,

$$\mathcal{C}_n=\{\theta\in[0,1]:\Psi_n\left(s^\infty\right)\left(1\right)\le\theta+c_\theta\left[\theta\left(1-\theta\right)\right]^{1/2}/\sqrt{n}\},$$

where, $c_\theta=0$ if $\theta=0$ or 1, and otherwise c_θ is the critical value for the standard normal variable and satisfies $\boldsymbol{N}_1\left(c_\theta;1\right)\ge1-\alpha$.[16] Thus the interval constraint

[16]The reduction to a univariate distribution is a consequence of the fact that $var_\theta\left(\{1\}\right)=0$ for all θ.

imposed by the LLN (see the appropriate form of (3.1)), whereby asymptotically the empirical frequency of $s = 1$ is bounded above by θ, is relaxed here to the degree expressed by $c_\theta \left[\theta\left(1-\theta\right)\right]^{1/2} / \sqrt{n}$. In particular, $c_\theta = 1.645$ if $\alpha = .05$.

It must be noted that the identical confidence region can arise also if the analyst completes her model and assumes that selections are i.i.d. across markets, and that when there are multiple equilibria then the equilibrium where both firms enter ($s = 1$) is selected with probability 1.[17] Then s_i is a Bernoulli random variable with parameter θ which is the largest (unconditional) probability consistent with the incomplete model. The MLE for θ is then $\hat{\theta} \equiv \Psi_n(s^\infty)(1)$. Assuming that the CLT for i.i.d. samples applies, $\hat{\theta}$ has the limiting normal distribution with mean 0 and variance $\theta(1-\theta)$, and the identical set $\mathcal{C}_n$ arises.

The preceding begs the questions "why does the noted procedural equivalence arise?" and "when does incompleteness make a difference?" The key observation is that in this example, for any given θ,

$$\begin{aligned}
\nu_\theta^\infty\left(\left\{s^\infty : \theta \in \mathcal{C}_n\right\}\right) &= \nu_\theta^\infty\left(\left\{s^\infty : \Psi_n\left(s^\infty\right)(1) \leq \theta + c_\theta\left[\theta\left(1-\theta\right)\right]^{1/2} / \sqrt{n}\right\}\right) \\
&= \min_{P \in \mathcal{P}_\theta} P\left(\left\{s^\infty : \Psi_n\left(s^\infty\right)(1) \leq \theta + c_\theta\left[\theta\left(1-\theta\right)\right]^{1/2} / \sqrt{n}\right\}\right),
\end{aligned}$$

and that a minimizing (or worst-case) measure exists as pointed out in the discussion surrounding (3.6)–a worst-case scenario for an event defined by an upper bound on the frequency of $s = 1$ is that the probability that $s = 1$ in each market is maximal (hence equal to θ) and is independent across markets. Thus the confidence region generated by the 'completed' model as above is also robust to all the scenarios arising from model incompleteness.

However, the scope of such procedural equivalence is limited. Indeed, it fails once both upper and lower bounds on the empirical frequency are relevant as in the next more general example.

Though we have focussed on the Jovanovic game, it is evident that our analysis can be applied also to any normal form game having finitely many pure strategies and where pure strategy Nash equilibria exist, that is, the equilibrium correspondence $G(\cdot \mid \theta)$ is nonempy-valued for every parameter θ. It may also be evident that the framework accommodates also games where players do not necessarily play equilibrium strategies. For example, if the analyst is willing to assume only that outcomes correspond to rationalizable strategy profiles, then the correspondence $G(\cdot \mid \theta)$ can be defined accordingly and inference can proceed as described

[17] We are not claiming that this is the most natural way to complete the model–just that the identical confidence region can arise also with some complete model featuring i.i.d. selection.

above.[18] However, the restriction to pure strategies is important. If we allowed mixed strategies, then the equilibrium correspondence $G(\cdot \mid \theta)$ would map into subsets of the probability simplex $\Delta(S)$ and ν_θ would be a belief function on $\Delta(S)$ rather than on S. Our formal results can be extended to this case in principle (though we have not studied the generalization of the CLT to infinite state spaces such as $\Delta(S)$). However, the corresponding CLT would refer to the empirical frequencies of mixed strategies, which are unobservable, rather than to the observable frequencies of realized pure strategies. Thus it seems that mixed strategies are beyond the scope of our approach to inference.

4.2. Binary experiments

This is a slight generalization of the Jovanovic example where the minimum probability is not fixed to equal 0; it corresponds also to a natural generalization of coin-tossing that incorporates an incomplete theory about the coin. Thus take $S = \{0,1\}$. The set of structural parameters is $\Theta = \{\theta = (\theta_1, \theta_2) \in [0,1]^2 : \theta_1 \leq \theta_2\}$, where θ_1 and θ_2 are interpreted as the minimal and maximal probabilities for the outcome $s = 1$. For (U, m), take any nonatomic probability space (with U Polish and $m_\theta = m$ for all θ). Finally, define $G(\cdot \mid \theta) : U \rightsquigarrow S$ by

$$G(u \mid \theta) = \begin{cases} \{1\} & \text{if } u \in U_{\theta 1} \\ \{0\} & \text{if } u \in U_{\theta 2} \\ \{1,0\} & \text{otherwise,} \end{cases}$$

where $U_{\theta 1}$ and $U_{\theta 2}$ are disjoint (Borel measurable) subsets of U such that $m(U_{\theta 1}) = \theta_1$ and $m(U_{\theta 2}) = 1 - \theta_2$. Then each θ induces the belief function ν_θ on S, where $\nu_\theta(1) = \theta_1$ and $\nu_\theta(0) = 1 - \theta_2$.

For inference about θ, take $J = 2$, $A_1 = \{1\}$ and $A_2 = \{0\}$. Then

$$\mathcal{C}_n = \left\{\theta : \theta_1 - c_\theta \left[\theta_1 (1-\theta_1)/n\right]^{1/2} \leq \Psi_n(s^\infty)(1) \leq \theta_2 + c_\theta \left[\theta_2(1-\theta_2)/n\right]^{1/2}\right\}, \tag{4.2}$$

which is the set of all $\theta_1 \leq \theta_2$ in the unit square that are either consistent with the interval restriction (3.1) due to the LLN, (here asserting that all limit points of $\Psi_n(s^\infty)(1)$ lie in $[\theta_1, \theta_2]$), or that permit the indicated small deviations from it. The region excludes θs for which θ_1 is "too large," but all sufficiently small θ_1

[18] Every Nash equilibrium profile is rationalizable and the converse is false in general. All profiles are rationalizable in the Jovanovic example, but in some games rationalizability rules out many profiles. See Chapters 4 and 5 of Osborne and Rubinstein (1994).

satisfy the first indicated inequality. This is because θ_1 is a minimum probability, and a small minimum cannot be contradicted by a larger empirical frequency for $s = 1$ which is attributed by the model to the vagaries of selection. Similarly, the confidence region excludes values of θ_2 that are too small relative to the empirical frequency, but all sufficiently large values are included.

A noteworthy feature of $\mathcal{C}_n$, that reflects the robustness of our approach, is that the critical value c_θ is scaled differently on the two extreme sides of the inequalities. The intuition is as follows. While (4.2) can be understood as describing a relaxation of the LLN to accommodate finite samples, the issue is how much to relax each inequality; for example, how much smaller than θ_1 can the empirical frequency be and still be seen as consistent with θ_1? This amounts to deciding on how much sampling variability to allow for $\Psi_n(s^\infty)(1)$. Since any probability law in $\mathcal{P}_\theta$ may apply, a conservative approach is to use the worst-case scenario, which, as in the Jovanovic example, is the i.i.d. law with the minimum probability for $s = 1$, namely θ_1. The associated variance is thus $\theta_1(1-\theta_1)$, as above. Similarly, for the upper bound on $\Psi_n(s^\infty)(1)$, for which the worst-case scenario has the maximum probability, namely θ_2, for $s = 1$, and thus a conservative approach leads to the variance $\theta_2(1-\theta_2)$ for the second inequality in (4.2). The resulting difference in scaling factors is implicit in the Jovanovic example because $\theta_1 = 0$ there.

There is another way to see why, in contrast with the preceding example, model incompleteness makes a difference here for confidence regions. Roughly speaking, our confidence regions provide coverage at least $1-\alpha$ according to every measure in $\mathcal{P}_\theta$, and thus are driven by the least favorable scenarios for the events $\{s^\infty : \theta \in \mathcal{C}_n\} =$

$$\left\{s^\infty : \theta_1 - c_\theta\left[\theta_1(1-\theta_1)/n\right]^{1/2} \leq \Psi_n(s^\infty)(1) \leq \theta_2 + c_\theta\left[\theta_2(1-\theta_2)/n\right]^{1/2}\right\}. \tag{4.3}$$

Because of the two-sided constraint on the frequency $\Psi_n(s^\infty)(1)$, these scenarios are not i.i.d., but rather feature "positive correlation" across markets which makes extreme values for the empirical frequency likely. We cannot be more precise about the nature of these unfavorable scenarios, in particular, we cannot identify particular parametric forms of dependence.[19] However, our confidence regions provide the desired coverage no matter what form that dependence might take.

Fix $\alpha = .05$. The critical value c_θ depends on θ according to (3.10). Though closed-forms are not available for all θ, the following can be shown by elementary

[19] Dependence in a cross-sectional context is often modeled by various parametric copulas.

arguments applied to the bivariate normal distribution (Appendix C):

$$\begin{aligned} c_{(0,0)} &= c_{(0,1)} = c_{(1,1)} = 0 \\ c_{(\theta_1,1)} &= 1.645 \text{ if } 0 < \theta_1 < 1 \\ c_{(0,\theta_2)} &= 1.645 \text{ if } 0 < \theta_2 < 1 \\ c_{(\theta_1,\theta_2)} &= 1.96 \text{ if } 0 < \theta_1 = \theta_2 < 1 \\ \{c_\theta &: 0 < \theta_1 < \theta_2 < 1\} = \{c : 1.955 < c < 1.96\}. \end{aligned} \tag{4.4}$$

In addition, $c_{(\theta_1,\theta_2)}$ is (strictly) increasing in θ_1 and decreasing in θ_2 on the domain $\{0 < \theta_1 < \theta_2 < 1\}$.

One may compare our confidence region to those in the moment inequalities (MI) literature. Below, we discuss a confidence region that assumes i.i.d. sampling. Under this assumption, the standard LLN and CLT imply that $\Psi_n(s^\infty)(1)$ converges in probability to $p(1) = p(s = 1)$ and that the studentized empirical frequency $\sqrt{n}(\Psi_n(s^\infty)(1) - p(1))/[\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))]^{1/2}$ converges in distribution to the standard normal distribution. Thus let

$$\begin{aligned} \mathcal{C}_n^{MI} = &\left\{\theta \in \Theta : \theta_1 - \tilde{c}_{n,\theta}[\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))/n]^{1/2} \le \Psi_n(s^\infty)(1) \right. \\ &\left. \le \theta_2 + \tilde{c}_{n,\theta}[\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))/n]^{1/2}\right\}. \end{aligned}$$

The critical value $\tilde{c}_{n,\theta}$ is given by:[20]

$$\tilde{c}_{n,\theta} = \begin{cases} 1.645 & \text{if } \hat{l}_{1n}(\theta) \le \kappa_n \text{ and } \hat{l}_{2n}(\theta) > \kappa_n \\ 1.645 & \text{if } \hat{l}_{1n}(\theta) > \kappa_n \text{ and } \hat{l}_{2n}(\theta) \le \kappa_n \\ 1.96 & \text{if } \hat{l}_{1n}(\theta) \le \kappa_n \text{ and } \hat{l}_{2n}(\theta) \le \kappa_n \\ 0 & \text{if } \hat{l}_{1n}(\theta) > \kappa_n \text{ and } \hat{l}_{2n}(\theta) > \kappa_n \end{cases}, \tag{4.5}$$

where $\{\kappa_n\}$ is a sequence of positive constants such that $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$ and

$$\hat{l}_{1,n}(\theta) \equiv \frac{\sqrt{n}(\Psi_n(s^\infty)(1) - \theta_1)}{[\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))]^{1/2}}, \quad \hat{l}_{2,n}(\theta) \equiv \frac{\sqrt{n}(\theta_2 - \Psi_n(s^\infty)(1))}{[\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))]^{1/2}}. \tag{4.6}$$

$\mathcal{C}_n^{MI}$ is a confidence region based on moment inequalities.[21] The studentized moments $\hat{l}_{j,n}$ are used to select those constraints to enter into calculation of the

[20] For comparison purposes, we use the critical value based on an asymptotic normal approximation instead of bootstrap approximations commonly used in the literature.

[21] One may view $\mathcal{C}_n^{MI}$ as Galichon and Henry's (2009) inference method with studentized moments. It also belongs to the general class of confidence regions studied by Andrews and Soares (2010).

critical value. For example, when $\hat{l}_{1,n}(\theta) \le \kappa_n$, the MI approach interprets this as indicating that the corresponding population constraint $p(1) - \theta_1 \ge 0$ is close to being binding, and hence retains this constraint in calculating the critical value; when $\hat{l}_{1,n}(\theta) > \kappa_n$, this constraint is not used.

The two confidence regions $\mathcal{C}_n$ and $\mathcal{C}_n^{MI}$ differ in terms of their critical values and scaling factors. As opposed to our method, the MI approach scales its critical value by the square root of $\Psi_n(s^\infty)(1)(1 - \Psi_n(s^\infty)(1))$. This is because their inference is based on the LLN and CLT with the i.i.d. assumption, under which the studentized empirical frequency converges in distribution to the standard normal distribution. Second, while $\tilde{c}_{n,\theta}$ and $c_{(\theta_1,\theta_2)}$ both take values between 0 and 1.96, the ways these critical values switch between distinct values are different: $\tilde{c}_{n,\theta}$ switches between 0, 1.645, and 1.96 depending on the number of constraints selected by the procedure, while $c_{(\theta_1,\theta_2)}$ changes its values depending on the covariance of the bivariate normal distribution.

The MI approach uses $\tilde{c}_{n,\theta} = 1.96$ when the two inequalities are locally binding, that is, $\hat{l}_{1n}(\theta) \le \kappa_n$ and $\hat{l}_{2n}(\theta) \le \kappa_n$. This is likely to occur when the interval $[\theta_1, \theta_2]$ is short, meaning that its length is comparable to the order $O(n^{-1/2})$ of the sampling variation of $\Psi_n(s^\infty)(1)$. Heuristically, $\Psi_n(s^\infty)(1)$ can then fall on either side of the interval, which motivates the two-sided critical value.[22] The value $\tilde{c}_{n,\theta} = 1.645$ is used when only one of the constraints is selected, which occurs when $\Psi_n(s^\infty)(1)$ is close to one of the end points, say θ_1 but not to θ_2. The MI approach interprets this as the length of the interval being large relative to the sampling variation and $p(1)$ being close to θ_1 but not to θ_2. Hence, if the empirical frequency is convergent to $p(1)$, then asymptotically it may fall to the left of θ_1 but not to the right of θ_2. Therefore, the problem reduces to a one-sided problem, which motivates $\tilde{c}_{n,\theta} = 1.645$. Finally, $\tilde{c}_{n,\theta} = 0$ is used when both constraints are considered slack, which occurs when the interval is long and $p(1)$ is not close to either endpoint. Since the MI approach assumes that $\Psi_n(s^\infty)(1)$ converges to $p(1)$ in the interior of the interval, the probability of it falling outside the interval tends to 0, which motivates $\tilde{c}_{n,\theta} = 0$.

In our framework, $\Psi_n(s^\infty)(1)$ does not necessarily converge. Hence, except in the special cases discussed below, $\Psi_n(s^\infty)(1)$ may fall on either side of the interval even asymptotically. Using our CLT, we approximate the minimum probability of the event where the empirical frequency is in an enlarged interval (in (4.3)) by a bivariate normal distribution. Therefore, the critical value $c_{(\theta_1,\theta_2)}$ depends on θ through the parameters in the bivariate normal distribution according to

[22] This was pointed out previously by Imbens and Manski (2004) and Stoye (2009).

(3.10). Accordingly, as stated in (4.4), $c_{(\theta_1,\theta_2)} = 1.96$ when $0 < \theta_1 = \theta_2 < 1$. This is because the two moments have a perfect (negative) correlation in this case, and the coverage probability reduces to $\Psi_n(s^\infty)(1)$'s two-sided variation around a common point $\theta_1 = \theta_2$. The value $c_{(\theta_1,\theta_2)} = 1.645$ is used when either θ_1 or θ_2 is on the boundary of the parameter space. For example, when $\theta_1 = 0$, there is no room for $\Psi_n(s^\infty)(1)$ to the left of θ_1; hence, it suffices to consider $\Psi_n(s^\infty)(1)$'s variation around θ_2, which motivates the one-sided critical value. Finally, $c_{(\theta_1,\theta_2)} = 0$ when both θ_1 and θ_2 are on the boundary. For example when $(\theta_1, \theta_2) = (0, 1)$, there is no room for $\Psi_n(s^\infty)(1)$ on the left of θ_1 or on the right of θ_2, which motivates 0 as the critical value. When $(\theta_1, \theta_2) = (0, 0)$ or $(1, 1)$, $\Psi_n(s^\infty)(1)$ does not involve any randomness and there is no need to relax any of the inequalities.

5. Monte Carlo simulations

We conduct Monte Carlo simulations to illustrate the performance of our inference method. For comparison purposes, we also include the results of existing procedures.[23]

Simulations are based on the binary experiment, slightly specialized so that $U = [0, 1]$, m is uniform on $[0, 1]$, $\Theta = \{(\theta_1, \theta_2) \in [0, 1]^2 : \theta_1 \le \theta_2\}$, and

$$G(u|\theta) = \begin{cases} \{1\} & \text{if } u < \theta_1 \\ \{0\} & \text{if } u > \theta_2 \\ \{0,1\} & \text{if } u \in [\theta_1, \theta_2]. \end{cases}$$

Thus each θ induces the belief function ν_θ on $\{0, 1\}$ given by

$$\nu_\theta(1) = \theta_1, \quad \text{and} \quad \nu_\theta(0) = 1 - \theta_2. \tag{5.1}$$

We consider two specifications for the equilibrium selection mechanism. In both specifications, $s_i = 1$ is selected from $\{0, 1\}$ when $u_i \in [\theta_1, \theta_2]$ and a binary latent variable v_i takes 1. The first specification is an i.i.d. selection mechanism, in which v_i is generated as an i.i.d. Bernoulli random variable independent of u_i with $prob(v_i = 1) = \tau$ for some $\tau \in [0, 1]$.

The second specification is a non-i.i.d. selection mechanism. For this, let N_k, $k = 1, 2, \cdots$, be an increasing sequence of integers. The set $\{i : N_{k-1} < i \le N_k\}$

[23] The MATLAB code for our simulations is available at: http://sites.google.com/site/seo8240.

defines a cluster of markets. We impose a common selection mechanism within each cluster. Let $h(i)=N_k$ if $N_{k-1}<i\leq N_k$ and define

$$v_i=\begin{cases}1 & \Psi^G_{h(i)}(u^\infty)>\frac{\theta_1}{\theta_1+(1-\theta_2)}\\ 0 & \Psi^G_{h(i)}(u^\infty)\leq\frac{\theta_1}{\theta_1+(1-\theta_2)}\end{cases},\quad\text{where }\Psi^G_n(u^\infty)=\frac{\sum_{i=1}^n I[G(u_i|\theta)=\{1\}]}{\sum_{i=1}^n I[G(u_i|\theta)\neq\{0,1\}]}. \tag{5.2}$$

The non-i.i.d. specification selects $s_i=1$ from $\{0,1\}$ when $\Psi^G_n(u^\infty)$, the relative frequency of the event where the model predicts $\{1\}$ as a unique outcome, crosses a threshold. Otherwise, $s_i=0$ is selected. This mechanism applies to all i belonging to the k-th cluster for which multiple equilibria are present.

Our inference procedure is implemented as follows. Since the belief function has a closed form (see (5.1)), computing the statistic and components of the covariance matrix Λ_θ is straightforward. To compute the critical value c_θ, one needs to evaluate a CDF of a multivariate normal distribution with covariance matrix Λ_θ. We do so by using simulated draws from the Geweke-Hajivassiliou-Keane (GHK) simulator and approximating the CDF $\boldsymbol{N}_J(\cdot;\Lambda_\theta)$ by Monte Carlo integration.[24] The critical value is then computed according to (3.10) replacing $\boldsymbol{N}_J(\cdot;\Lambda_\theta)$ by its approximation. Throughout this section, we denote our confidence region by $\mathcal{C}_n^{\mathrm{EKS}}$.

We compare the performance of the robust confidence region with that of existing methods. For each θ, let $\bar{M}_{n,\theta}\equiv(\nu^*_\theta(1)-\Psi_n(s^\infty)(1),\nu^*_\theta(0)-\Psi_n(s^\infty)(0))'$. Confidence regions in the moment inequalities (MI) literature take the form:

$$CS_n=\left\{\theta\in\Theta:\Gamma(\sqrt{n}\bar{M}_{n,\theta},\hat{\Sigma}_{n,\theta})\leq\tilde{c}_{n,\theta}(\kappa_n)\right\},$$

where $\Gamma:\mathbb{R}^J\times\mathbb{R}^{J\cdot J}\to\mathbb{R}$ is a function that aggregates (normalized) moment functions, and $\hat{\Sigma}_{n,\theta}$ is an estimator of the asymptotic variance of $\sqrt{n}\bar{M}_{n,\theta}$. $\tilde{c}_{n,\theta}$ is a critical value that depends on a possibly data-dependent tuning parameter κ_n.

We consider two confidence regions that belong to this class. The first, denoted $\mathcal{C}_n^{\mathrm{MI}}$, based on Galichon and Henry (2009) and Andrews and Soares (2010), uses

[24] See simulation procedure 2 in Appendix F for details on the implementation of our procedure. In the present simulations, $J=2$ and we need to compute bivariate normal CDF values. There are faster and more accurate algorithms for the bivariate case, (see Genz (2004), for example), but we adopt the GHK method because it may be used for applications with larger J.

the following criterion function and estimator of the asymptotic variance:

$$\begin{aligned}\Gamma(M,\Sigma) &= \max_{j=1,\cdots,J}(-\Sigma_{jj}^{-1/2}M_j)\\ \hat{\Sigma}_{n,\theta} &= \frac{1}{n}\sum_{i=1}^{n}(M_\theta(s_i)-\bar{M}_{n,\theta})(M_\theta(s_i)-\bar{M}_{n,\theta})',\end{aligned}$$

where $M_\theta(s) \equiv (\nu_\theta^*(1) - I(s_i = 1), \nu_\theta^*(0) - I(s_i = 0))'$. We then compute $\tilde{c}_{n,\theta}$ via bootstrap combined with a generalized moment selection (GMS) procedure. This method selects the moments that are relevant for inference by comparing sample moments to a tuning parameter κ_n provided by the researcher. Specifically, for each j, let $\hat{l}_{j,n}(\theta) = \bar{M}_{j,n,\theta}/[\Psi_n(s^\infty)(1)(1-\Psi_n(s^\infty)(1))]^{1/2}$ be the studentized moment and let $\varphi_{n,\theta}$ be a $J \times 1$ vector whose j-th component satisfies

$$\varphi_{j,n,\theta} = \begin{cases} 0 & \text{if } \hat{l}_{j,n}(\theta) \leq \kappa_n \\ \infty & \text{if } \hat{l}_{j,n}(\theta) > \kappa_n \ . \end{cases}$$

The critical value is then calculated as the $1-\alpha$ quantile of the bootstrapped statistic $\Gamma(\bar{M}^*_{n,\theta} + \varphi_{n,\theta}, \hat{\Sigma}^*_{n,\theta})$, where $(\bar{M}^*_{n,\theta}, \hat{\Sigma}^*_{n,\theta})$ is a bootstrap counterpart of $(\bar{M}_{n,\theta}, \hat{\Sigma}_{n,\theta})$.[25]

The second confidence region, denoted $\mathcal{C}_n^{\mathrm{AB}}$, based on Andrews and Barwick (2012), uses the test statistic $T_n(\theta) = \Gamma(\sqrt{n}\bar{M}_{n,\theta}, \tilde{\Sigma}_{n,\theta})$ with the following criterion function and regularized estimator of the asymptotic variance:

$$\begin{aligned}\Gamma(M,\Sigma) &= \inf_{t\in\bar{\mathbb{R}}_+^J}(M-t)'^{-1}(M-t)\\ \tilde{\Sigma}_{n,\theta} &\equiv \hat{\Sigma}_{n,\theta} + \max(0.012 - \det(\hat{\Omega}_{n,\theta}), 0)\hat{D}_{n,\theta},\end{aligned}$$

where $\hat{D}_{n,\theta} = \mathrm{diag}(\hat{\Sigma}_{n,\theta})$ and $\hat{\Omega}_{n,\theta} = \hat{D}_{n,\theta}^{-1/2}\hat{\Sigma}_{n,\theta}\hat{D}_{n,\theta}^{-1/2}$. Their critical value requires three tuning parameters including κ_n, which we set following their recommendations.

Table 5.1 reports the coverage probabilities of the three confidence regions $\mathcal{C}_n^{\mathrm{EKS}}, \mathcal{C}_n^{\mathrm{MI}}$, and $\mathcal{C}_n^{\mathrm{AB}}$ under alternative values of (θ_1, θ_2) for a nominal level of 0.95. We set $\tau = 0.5$ and 1 for the i.i.d. selection mechanism, and $N_k =$

[25] See Andrews and Soares (2010) for details on general GMS procedures that include $\varphi_{n,\theta}$ as a special case. The moment selection tuning parameter κ_n here corresponds to $\sqrt{n}$ times the tuning parameter h_n in Galichon and Henry (2009).

Table 5.1: Coverage Probabilities of Confidence Regions

Eq. Sel.	Sample Size	Robust	MI		Robust	MI	
	n	$\mathcal{C}_n^{\mathrm{EKS}}$	$\mathcal{C}_n^{\mathrm{MI}}$	$\mathcal{C}_n^{\mathrm{AB}}$	$\mathcal{C}_n^{\mathrm{EKS}}$	$\mathcal{C}_n^{\mathrm{MI}}$	$\mathcal{C}_n^{\mathrm{AB}}$
		A: $(\theta_1, \theta_2) = (0.4, 0.6)$			B: $(\theta_1, \theta_2) = (0.49, 0.51)$		
i.i.d. ($\tau = 0.5$)	100	1.000	0.999	0.999	0.963	0.934	0.966
	256	1.000	1.000	1.000	0.983	0.946	0.979
	400	1.000	1.000	1.000	0.979	0.949	0.974
	10000	1.000	1.000	1.000	1.000	1.000	1.000
	65536	1.000	1.000	1.000	1.000	1.000	1.000
i.i.d. ($\tau = 1$)	100	0.981	0.961	0.959	0.959	0.932	0.964
	256	0.977	0.960	0.959	0.973	0.936	0.970
	400	0.981	0.950	0.954	0.973	0.941	0.978
	10000	0.973	0.940	0.941	0.969	0.945	0.943
	65536	0.974	0.941	0.947	0.976	0.950	0.952
non-i.i.d.	100	0.952	0.919	0.926	0.952	0.905	0.954
	256	0.955	0.919	0.914	0.949	0.893	0.938
	400	0.984	0.967	0.964	0.962	0.923	0.959
	10000	0.973	0.953	0.946	0.962	0.922	0.923
	65536	0.969	0.918	0.925	0.958	0.909	0.913

Note: We simulate 1000 datasets for each setting. For the non-i.i.d case, $N_k = 2^{2^k} \in \{4, 16, 256, 65536\}$. $\mathcal{C}_n^{\mathrm{MI}}$ uses the generalized moment selection procedure with the tuning parameter value $\kappa_n = \ln \ln n$. $\mathcal{C}_n^{\mathrm{AB}}$ uses the tuning parameter values recommended by Andrews and Barwick (2012).

$2^{2^k} \in \{4, 16, 256, 65536\}$ for the non-i.i.d. selection mechanism. We report simulation results based on samples of size $n \in \{100, 256, 400, 10000, 65536\}$. $\mathcal{C}_n^{\mathrm{MI}}$ uses the generalized moment selection procedure with the tuning parameter value $\kappa_n = \ln \ln n$. $\mathcal{C}_n^{\mathrm{AB}}$ uses the tuning parameter values recommended by Andrews and Barwick (2012).[26]

We note that the non-i.i.d. selection mechanism becomes less favorable to controlling the coverage probability when n is close to N_k for some k. This can be understood as follows. When the empirical frequency of the event $G(u_i|\theta) = \{1\}$, i.e. 1 being predicted as a unique outcome, crosses the threshold in (5.2), then the selection mechanism additionally selects $s_i = 1$ across all markets in cluster k where multiple equilibria are predicted. This increases the empirical frequency

[26] The moment selection tuning parameter κ_n and size correction factors (η_{1n}, η_{2n}) are selected from Table I in Andrews and Barwick (2012) based on the smallest off-diagonal element of $\hat{\Omega}_{n,\theta}$.

Table 5.2: Volume of Confidence Regions

Eq. Sel.	Sample Size n	Robust $\mathcal{C}_n^{\text{EKS}}$	MI $\mathcal{C}_n^{\text{MI}}$	MI $\mathcal{C}_n^{\text{AB}}$	Robust $\mathcal{C}_n^{\text{EKS}}$	MI $\mathcal{C}_n^{\text{MI}}$	MI $\mathcal{C}_n^{\text{AB}}$
		A: $(\theta_1, \theta_2) = (0.4, 0.6)$			B: $(\theta_1, \theta_2) = (0.49, 0.51)$		
i.i.d. $(\tau = 0.5)$	100	0.360	0.340	0.326	0.360	0.341	0.327
	256	0.314	0.303	0.299	0.314	0.304	0.298
	400	0.300	0.293	0.290	0.300	0.293	0.290
	10000	0.262	0.261	0.261	0.262	0.261	0.261
	65536	0.262	0.261	0.261	0.262	0.261	0.261
i.i.d. $(\tau = 1)$	100	0.350	0.329	0.317	0.360	0.341	0.327
	256	0.305	0.294	0.289	0.314	0.304	0.299
	400	0.290	0.282	0.280	0.300	0.292	0.290
	10000	0.252	0.251	0.251	0.257	0.255	0.256
	65536	0.252	0.251	0.252	0.250	0.250	0.250
non-i.i.d.	100	0.346	0.326	0.314	0.359	0.340	0.326
	256	0.300	0.290	0.285	0.314	0.303	0.298
	400	0.293	0.285	0.283	0.300	0.292	0.290
	10000	0.252	0.251	0.251	0.257	0.255	0.255
	65536	0.252	0.252	0.252	0.250	0.250	0.250

Note: We simulate 1000 datasets for each setting. For the non-i.i.d case, $N_k = 2^{2^k} \in \{4, 16, 256, 65536\}$. $\mathcal{C}_n^{\text{MI}}$ uses the generalized moment selection procedure with the tuning parameter value $\kappa_n = \ln\ln n$. $\mathcal{C}_n^{\text{AB}}$ uses the tuning parameter values recommended by Andrews and Barwick (2012).

of $\{1\}$, and thus lowers the probability of the statistic being dominated by the critical value.

Overall, our confidence region controls the coverage probability properly across all specifications even in small samples. This confirms the robustness of our procedure. The coverage probabilities of the two other confidence regions depend on the equilibrium selection specifications.

Panel A in Table 5.1 shows the results for the case $(\theta_1, \theta_2) = (0.4, 0.6)$. Under the i.i.d. selection mechanism with $\tau = 0.5$, the coverage probabilities of all confidence regions are close to 1. This is because, under this specification the empirical frequency converges to a point ($p = 0.5$) in the interior of the probability interval $[\theta_1, \theta_2]$ whose length is long relative to the sampling variation of the empirical frequency. When $\tau = 1$, the empirical frequency $\Psi_n(1)$ converges to $\nu_\theta^*(1)$. All confidence regions control the coverage probabilities reasonably well under this specification. Under the non-i.i.d. specification, the empirical frequency does not have a unique limit point. As discussed above, size control becomes more difficult when n is close to N_k for some k. The coverage probabilities of $\mathcal{C}_n^{\mathrm{MI}}$ and $\mathcal{C}_n^{\mathrm{AB}}$ in such settings are below the nominal level, for example, they equal 0.919 and 0.914 respectively when $n = 256$. Even when $n = 65536$, their respective coverage probabilities equal 0.918 and 0.925, thus exhibiting size distortions even in large samples due to the non-i.i.d. (highly dependent) nature of the selection mechanism.

Panel B in Table 5.1 reports coverage probabilities for $(\theta_1, \theta_2) = (0.49, 0.51)$. In this setting, the probability interval has a shorter length. Overall, under the i.i.d. specifications, existing methods control size reasonably well although the coverage probability for $\mathcal{C}_n^{\mathrm{MI}}$ is slightly below the nominal level in small samples.[27] For the non-i.i.d. specification, however, we again see that they have size distortions when the sample size equals N_k for some k. For example, the coverage probabilities of $\mathcal{C}_n^{\mathrm{MI}}$ and $\mathcal{C}_n^{\mathrm{AB}}$ are 0.909 and 0.913 respectively when $n = 65536$. In addition, there are size distortions even when sample sizes are not close to N_k (e.g. their coverage probabilities are 0.922 and 0.923 respectively when $n = 10000$).

Finally, we examine the cost of robustness by comparing the volume of the robust confidence region to the volumes in existing methods. Table 5.2 shows the average volume of the different confidence regions. Overall, the robust confidence region has a slightly higher volume than the other methods especially in small samples. However, this difference becomes very small as the sample size gets large. These features hold under both i.i.d. and non-i.i.d. specifications.

6. Covariates

This section describes an extension of our approach to accommodate covariates that model observable heterogeneity. Because interpretations follow closely those for the stripped-down model, we keep discussions brief and focussed on the new features.

The model of each individual experiment is now described by $(S, X, U, G, \Theta; q, m)$, where: S, U, Θ, m are as before, and X is the finite set of covariate values. Covariates are stochastic and distributed according to the full support measure $q \in \Delta(X)$, independently from u. Model predictions take the form of a (weakly measurable) correspondence $G(\cdot \mid \theta, x) : U \rightsquigarrow S$, for each $\theta \in \Theta$ and $x \in X$. The

[27] Under the i.i.d. specification with $\tau = 0.5$, the coverage probabilities of all confidence regions are now below 1 in relatively small samples due to the shorter length of the probability interval.

latter and m induce, the belief function $\nu_\theta(\cdot \mid x)$ on S, that is conditional on each θ and x, and is given by

$$\nu_\theta(A \mid x) = m_\theta(\{u \in U : G(u \mid \theta, x) \subset A\}), \quad A \subset S.$$

To model the infinite sequence of experiments, consider $(S^\infty, X^\infty, U^\infty, G^\infty, \Theta; q^\infty, m^\infty)$, where (x_i, u_i) are assumed to be i.i.d. and distributed according to the product of q^∞ and m^∞. The outcomes for the entire sequence of experiments are described by the correspondence, $G^\infty(\cdot \mid \theta, x^\infty) : U^\infty \rightsquigarrow S^\infty$, where, for each θ and sequence of covariates $x^\infty \equiv (x_1, ..., x_i, ...) \in X^\infty$,

$$G^\infty(u_1, ..., u_i, ... \mid \theta, x^\infty) \equiv \Pi_{i=1}^\infty G(u_i \mid \theta, x_i).$$

This correspondence induces, for each $\theta \in \Theta$ and $x^\infty \in X^\infty$, the belief function $\nu_\theta^\infty(\cdot \mid x^\infty)$ on S^∞ given by

$$\nu_\theta^\infty(B \mid x^\infty) = m_\theta^\infty(\{u^\infty \in U^\infty : G^\infty(u^\infty \mid \theta, x^\infty) \subset B\}), \, B \subset S^\infty.$$

Then, $\nu_\theta^\infty(\cdot \mid x^\infty)$ gives the lower envelope of the set $\mathcal{P}_{\theta,x^\infty}$, paralleling (2.1), of all probability laws over S^∞ that are consistent with the given theory and θ and with agnosticism about selection. Consistent with such agnosticism, the set $\mathcal{P}_{\theta,x^\infty}$ does not restrict how selection varies with the covariate.

For inference we fix $A_1, ..., A_J$, subsets of S.[28] Define, for each θ and $x \in X$,

$$\begin{aligned} cov_\theta(A_i, A_j \mid x) &= v_\theta(A_i \cap A_j \mid x) - v_\theta(A_i \mid x)\, v_\theta(A_j \mid x) && (6.1)\\ var_\theta(A_j \mid x) &= cov_\theta(A_j, A_j \mid x). && (6.2) \end{aligned}$$

Let $\Lambda_{\theta,x}$ be the covariance matrix, conditional on x: $(\Lambda_{\theta,x})_{jj'} = cov_\theta(A_j, A_{j'} \mid x)$. Let Λ_θ be the $|X|\,J$-by-$|X|\,J$ block-diagonal matrix where $\Lambda_{\theta,x_1}, ..., \Lambda_{\theta,x_{|X|}}$ are the blocks; the $(k(J-1)+j, k'(J-1)+j')$ element of Λ_θ is 0 if $k \neq k'$, and equals $cov_\theta(A_j, A_{j'} \mid x_k)$ if $k = k'$.

Define $c_\theta = \min\left\{c \in \mathbb{R}_+ : \boldsymbol{N}_{|X|J}(c\sigma_\theta; \Lambda_\theta) \geq 1-\alpha\right\}$. Another way to express c_θ is as follows. Let $Z_\theta = \left(Z_{\theta,1}, ..., Z_{\theta,|X|J}\right)$ be multivariate normal with mean 0 and covariance Λ_θ, and let $W = \max_{k=1,...,|X|J} Z_{\theta,k}/\sigma_{\theta,k}$ with the conventions $1/0 = \infty$, $0/0 = 0$ and $-1/0 = -\infty$. Then c_θ is the critical value of W: $c_\theta = \min\{c \in \mathbb{R}_+ : \Pr(W \leq c) \geq 1-\alpha\}$. It can be shown that, if $0 < \alpha < 1/2$ and $\Lambda_\theta \neq 0$, then $\Pr(W \leq c_\theta) = 1-\alpha$.

[28] Below the same collection $\{A_j\}$ of events is used for each covariate value. This is only for simplicity; we could alternatively use $\{A_j^k\}_{j=1}^{J_k}$ for covariate $x = x_k$.

For each $s^\infty \in S^\infty$, $x^\infty \in X^\infty$ and $A \subset S$, denote by $\Psi_n(s^\infty, x^\infty)(A \mid x)$ the empirical frequency of A in the first n experiments counting only those experiments where $x_i = x$:

$$\Psi_n(s^\infty, x^\infty)(A \mid x) = \left(\sum_{i=1}^{n} I(x_i = x)\right)^{-1} \sum_{i=1}^{n} I(x_i = x, s_i \in A).$$

Since q has the full support, Ψ_n is well-defined asymptotically. Define the statistic

$$T_n(\theta) = \max_{(x,j)\in X\times\{1,\ldots,J\}} \left\{ \frac{\nu_\theta(A_j \mid x) - \Psi_n(s^\infty, x^\infty)(A_j \mid x)}{\sqrt{var_\theta(A_j \mid x)/n}} \right\}, \tag{6.3}$$

where we adopt the conventions $1/0 = \infty$, $0/0 = 0$ and $-1/0 = -\infty$.

Finally, define the confidence region:

$$\mathcal{C}_n = \{\theta \in \Theta : T_n(\theta) \leq c_\theta\}.$$

It is not difficult to verify that

$$\mathcal{C}_n = \bigcap_{(x,j)\in X\times\{1,\ldots,J\}} \left\{\theta \in \Theta : \nu_\theta(A_j \mid x) - \Psi_n(s^\infty, x^\infty)(A_j \mid x) \leq c_\theta \sqrt{var_\theta(A_j \mid x)/n}\right\}.$$

Theorem 6.1. *Suppose that each $x \in X$ appears in the given sequence $x^\infty = (x_1, x_2, \ldots)$ infinitely many times. Then,*

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n \mid x^\infty) \geq 1 - \alpha.$$

Moreover, equality prevails if $0 < \alpha < \frac{1}{2}$ and $\Lambda_\theta \neq 0$ for some $\theta \in \Theta$.

The main coverage property for the model with covariates follows as a corollary. Define the unconditional belief function by

$$\nu_\theta^\infty(\cdot) = \int \nu_\theta^\infty(\cdot \mid x^\infty)\, dq^\infty(x^\infty).$$

Corollary 6.2. *We have*

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n) \geq 1 - \alpha.$$

Moreover, equality prevails if $0 < \alpha < \frac{1}{2}$ and $\Lambda_\theta \neq 0$ for some $\theta \in \Theta$.

A. Appendix: Proof of CLT

Fix θ. A particular case of the conditional structure $(U, G(\cdot \mid \theta), m_\theta)$ occurs when $U = \mathcal{K}(S)$, the set of all nonempty (and necessarily closed) subsets of S, endowed with the discrete metric because S is finite, and $G(\cdot \mid \theta) = \widehat{G}$ maps any $K \in \mathcal{K}(S)$ into $\widehat{G}(K) = K \subset S$. In fact, Choquet's Theorem (Philippe, Debs and Jaffray 1999, Molchanov 2005) shows that the latter structure is without loss of generality: a belief function ν_θ on S generated by any $(U, G(\cdot \mid \theta), m_\theta)$ can also be generated by $\left(\mathcal{K}(S), \widehat{G}, \widehat{m}_\theta\right)$ for some probability measure $\widehat{m}_\theta$ on $\mathcal{K}(S)$; and similarly for ν_θ^∞. Because $\left(\mathcal{K}(S), \widehat{G}, \widehat{m}_\theta\right)$ is typically viewed as the canonical representation of a belief function, we adopt it in the following proof of the CLT. We also denote the measure on $\mathcal{K}(S)$ by m_θ rather than $\widehat{m}_\theta$. Then, without loss of generality, suppose that ν_θ and ν_θ^∞ satisfy

$$\nu_\theta(A) = m_\theta(\{K \in \mathcal{K}(S) : K \subset A\}), \; A \subset S,$$

and

$$\nu_\theta^\infty(B) = m_\theta^\infty\left(\{K_1 \times K_2 \times ... \in (\mathcal{K}(S))^\infty : \Pi_{i=1}^\infty K_i \subset B\}\right), \; B \subset S^\infty.$$

Now we consider a sequence $\{\theta_n\}$, which induces the sequence of structures $\{(U, G(\cdot \mid \theta_n), m_{\theta_n})\}$. On the probability space $\left((\mathcal{K}(S))^\infty, m_{\theta_n}^\infty\right)$, define random variables X_{ni}^j by

$$X_{ni}^j = I(K_i \subset A_j) = \begin{cases} 1 & \text{if } K_i \subset A_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for each } i, n = 1, 2, ... \text{ and } j = 1, ..., J.$$

Then, (using $m_{\theta_n}^\infty$), $EX_{ni}^j = \nu_{\theta_n}(A_j)$,

$$\begin{aligned} cov\left(X_{ni}^j, X_{ni}^l\right) &= E\left(X_{ni}^j X_{ni}^l\right) - E\left(X_{ni}^j\right) E\left(X_{ni}^l\right) \\ &= \int I(K_i \subset A_j) I(K_i \subset A_l) \, dm_{\theta_n}(K_i) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l) \\ &= \int I(K_i \subset A_j \cap A_l) \, dm_{\theta_n}(K_i) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l) \\ &= \nu_{\theta_n}(A_j \cap A_l) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l), \text{ and} \end{aligned}$$

$$var\left(X_i^j\right) = cov\left(X_i^j, X_i^j\right) = \nu_{\theta_n}(A_j)\left(1 - \nu_{\theta_n}(A_j)\right).$$

Let X_{ni} be the $\mathbb{R}^J$-valued random variable with jth component X_i^j. Define

$$Y_{ni}^j = \left(X_{ni}^j - EX_{ni}^j\right),$$

and let Y_{ni} be the corresponding $\mathbb{R}^J$-valued random variable. Then, $EY_{ni} = 0$ and Y_{ni} has the variance-covariance matrix Λ_{θ_n}.

Compute that, for any $\beta \in \mathbb{R}^J$,

$$\begin{gathered}
K_1 \times K_2 \times ... \subset \left\{s^\infty : \beta_j \leq n\Psi_n\left(s^\infty\right)\left(A_j\right) \text{ for each } j\right\} \iff \\
K_1 \times K_2 \times ... \subset \left\{s^\infty : \beta_j \leq \sum_{i=1}^n I\left(s_i \in A_j\right) \text{ for each } j\right\} \iff \\
\beta_j \leq \min_{s^\infty \in K_1 \times K_2 \times ...} \sum_{i=1}^n I\left(s_i \in A_j\right) \text{ for each } j \iff \\
\beta_j \leq \sum_{i=1}^n \min_{s_i \in K_i} I\left(s_i \in A_j\right) \text{ for each } j \iff \\
\beta_j \leq \sum_{i=1}^n I\left(K_i \subset A_j\right) \text{ for each } j \iff \\
\beta_j \leq \sum_{i=1}^n X_{ni}^j \text{ for each } j = 1, ..., J.
\end{gathered}$$

Hence,

$$\begin{aligned}
& \nu_{\theta_n}^\infty\left(\left\{s^\infty : \beta_j \leq n\Psi_n\left(s^\infty\right)\left(A_j\right) \text{ for each } j\right\}\right) \\
= \; & m_{\theta_n}^\infty\left(\left\{K_1 \times K_2 \times ... \in \left(\mathcal{K}\left(S\right)\right)^\infty : \beta_j \leq \sum_{i=1}^n X_{ni}^j \text{ for each } j\right\}\right),
\end{aligned}$$

and consequently, for any $c_n \in \mathbb{R}^J$,

$$\begin{aligned}
& \nu_{\theta_n}^\infty\left(\sqrt{n}\left(\nu_{\theta_n}\left(A_j\right) - \Psi_n\left(s^\infty\right)\left(A_j\right)\right) \leq c_{nj} \text{ for each } j\right) \\
= \; & \nu_{\theta_n}^\infty\left(n\nu_{\theta_n}\left(A_j\right) - \sqrt{n}c_{nj} \leq n\Psi_n\left(s^\infty\right)\left(A_j\right) \text{ for each } j\right) \\
= \; & m_{\theta_n}^\infty\left(n\nu_{\theta_n}\left(A_j\right) - \sqrt{n}c_{nj} \leq \sum_{i=1}^n X_{ni}^j \text{ for each } j\right) \\
= \; & m_{\theta_n}^\infty\left(-c_{nj} \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(X_{ni}^j - \nu_{\theta_n}\left(A_j\right)\right) \text{ for each } j\right) \\
= \; & m_{\theta_n}^\infty\left(-c_{nj} \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ni}^j \text{ for each } j\right).
\end{aligned}$$

Thus the assertion to be proven has been translated into one about independent (triangular) random variables and classical results can be applied.

We prove that $\widetilde{Y}_n \equiv c_n + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{ni} \rightarrow_d Z$ where Z is J-dimensional multivariate normal with mean c and covariance matrix Λ. Apply the Cramér-Wold device: let $a \in \mathbb{R}^J$ and show that $a'\widetilde{Y}_n \rightarrow_d a'Z$. Note that $\lim_{n\rightarrow\infty} var\left(a'\widetilde{Y}_n\right) = \lim_{n\rightarrow\infty} a'\Lambda_{\theta_n} a = a'\Lambda a$. If $a'\Lambda a = 0$, then $a'\widetilde{Y}_n \rightarrow_d c = a'Z$. If $a'\Lambda a > 0$, we can apply a triangular CLT (White 2001, Theorem 5.11),[29] to prove that

$$\frac{\sum_{i=1}^{n} a'Y_{ni}}{\sqrt{n\left(a'\Lambda_{\theta_n} a\right)}} \rightarrow_d N(0,1).$$

Since $\lim_{n\rightarrow\infty} a'\Lambda_{\theta_n} a = a'\Lambda a$,

$$a'\widetilde{Y}_n = a'c_n + \frac{\sum_{i=1}^{n} a'Y_{ni}}{\sqrt{n}} \rightarrow_d N\left(a'c, a'\Lambda a\right).$$

Thus $a'\widetilde{Y}_n \rightarrow_d a'Z$ for all $a \in \mathbb{R}^J$, which implies that $\widetilde{Y}_n \rightarrow_d Z$.

The proof of (3.4) is completed by noting that

$$\begin{aligned} & \nu_{\theta_n}^{\infty}\left(\cap_{j=1}^{J}\left\{s^{\infty} : \sqrt{n}\left[\nu_{\theta_n}(A_j) - \Psi_n(s^{\infty})(A_j)\right] \leq c_{nj}\right\}\right) \\ = \; & m_{\theta_n}^{\infty}\left(0 \leq \widetilde{Y}_n\right) \rightarrow \Pr(0 \leq Z) = \Pr(-Z + c \leq c) = \boldsymbol{N}_J(c;\Lambda). \end{aligned}$$

■

B. Appendix: Proof of Theorem 3.2

A preliminary remark is that $\{s^{\infty} : \theta \in \mathcal{C}_n\}$ is measurable for each θ because it equals $\bigcap_{j=1}^{J}\left\{s^{\infty} : \nu_\theta(A_j) - \Psi_n(s^{\infty})(A_j) \leq c_\theta \sqrt{var_\theta(A_j)/n}\right\}$ and because $s^{\infty} \mapsto \Psi_n(s^{\infty})(A_j)$ is measurable for each j.

For any positive semidefinite matrix $\Lambda \in \mathbb{R}^{J \cdot J}$, let $\sigma(\Lambda) \equiv \left(\sqrt{\Lambda_{11}}, ..., \sqrt{\Lambda_{JJ}}\right)$ and define

$$c(\Lambda) = \min\left\{c \in \mathbb{R}_+ : \boldsymbol{N}_J(c\sigma(\Lambda);\Lambda) \geq 1 - \alpha\right\}.$$

We show shortly that $c(\Lambda)$ is defined even if $\Lambda \notin \{\Lambda_\theta : \theta \in \Theta\}$. It will follow that $c(\Lambda_\theta) = c_\theta$ for every θ.

[29] The condition in the theorem that $E\left|a'Y_{ni}\right|^{2+\delta}$ is bounded is satisfied here because Y_{ni} is bounded.

Step 1: $\boldsymbol{N}_J\left(\sqrt{\frac{J}{\alpha}}\sigma(\Lambda);\Lambda\right) \geq 1-\alpha$: Let X be multivariate normal with mean 0 and covariance matrix Λ. Then the Chebyshev inequality implies that, for $c>0$,

$$1-\boldsymbol{N}_J(c\sigma(\Lambda);\Lambda) = \Pr\left(\bigcup_{j=1}^{J}\{X_j > c\sigma_j(\Lambda)\}\right) \leq \sum_j \Pr(X_j > c\sigma_j(\Lambda)) \leq \frac{J}{c^2}.$$

Set $c^2 = \frac{J}{\alpha}$. (In particular, when $\sigma(\Lambda)=0$, then $\boldsymbol{N}_J\left(\sqrt{\frac{J}{\alpha}}\sigma(\Lambda);\Lambda\right) = \boldsymbol{N}_J(0;\Lambda) = 1 > 1-\alpha$.)

Step 2: $c(\Lambda)$ is well-defined for every $0<\alpha<1$: Note that $c \longmapsto \boldsymbol{N}_J(c\sigma(\Lambda);\Lambda)$ is upper semicontinuous and (weakly) increasing for all Λ, and (by Step 1) $\boldsymbol{N}_J(c\sigma(\Lambda);\Lambda) \geq 1-\alpha$ for some $c \geq 0$. It follows that $c(\Lambda)$ is well-defined as a minimum. Note also that, for $c^* \geq 0$,

$$\boldsymbol{N}_J(c^*\sigma(\Lambda);\Lambda) \geq 1-\alpha \iff c^* \geq c(\Lambda). \tag{B.1}$$

Step 3: $(c,\Lambda) \longmapsto \boldsymbol{N}_J(c\sigma(\Lambda);\Lambda)$ is upper semicontinuous: Take $(c_n,\Lambda_n) \to (c,\Lambda) \in \mathbb{R}\times\mathbb{R}^{J\cdot J}$. Let X_n and X be multivariate normal random vectors with means $-c_n\sigma(\Lambda_n)$ and $-c\sigma(\Lambda)$, and variances Λ_n and Λ, respectively. Then the characteristic functions of X_n converge pointwise to the characteristic function of X, which implies that $X_n \to_d X$ by Lévy's Continuity Theorem. Thus

$$\limsup_{n\to\infty} \boldsymbol{N}_J(c_n\sigma(\Lambda_n);\Lambda_n) = \limsup_{n\to\infty} \Pr(X_n \leq 0) \leq \Pr(X \leq 0).$$

Step 4: $[\Lambda_n \to \Lambda$ and $c(\Lambda_n) \to c^*] \implies c^* \geq c(\Lambda)$: By Step 3, $\boldsymbol{N}_J(c^*\sigma(\Lambda);\Lambda) \geq 1-\alpha$. Apply (B.1).

Step 5: Let $Bel(S)$ be the set of belief functions on S equipped with the sup-norm topology. Since S is finite, $Bel(S)$ is compact. For $\nu \in Bel(S)$, let Λ_ν be the covariance matrix as defined in (3.3). Then $\nu \longmapsto \Lambda_\nu$ is continuous and hence $\{\Lambda_\nu : \nu \in Bel(S)\}$ is compact.

Step 6: Complete the proof of (3.11). Let $\{\theta_n\}$ be a sequence such that

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n) = \liminf_{n\to\infty} \nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n).$$

Since $\nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n)$ is bounded, by taking a subsequence if necessary, we can assume that $\liminf_{n\to\infty} \nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n) = \lim_{n\to\infty} \nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n)$. Moreover, by Step 5, and by taking a further subsequence if necessary, we can assume that $\Lambda_{\theta_n} \to \Lambda \in$

$\mathbb{R}^{J \cdot J}$. By Step 1 and (B.1), $0 \leq c(\Lambda_{\theta_n}) \leq \left[\frac{J}{\alpha}\right]^{1/2}$. Therefore, a further subsequence allows us to assume that $c(\Lambda_{\theta_n}) \to c^*$. Thus, the CLT (Theorem 3.1) implies that

$$\begin{aligned} \lim_{n\to\infty} \nu^{\infty}_{\theta_n}(\theta_n \in \mathcal{C}_n) &= \boldsymbol{N}_J(c^*\sigma(\Lambda);\Lambda) \\ \text{(by Step 4)} \quad &\geq \boldsymbol{N}_J(c(\Lambda)\sigma(\Lambda);\Lambda) \geq 1-\alpha. \end{aligned}$$

Step 7: If $\boldsymbol{N}_J(c_\theta\sigma_\theta;\Lambda_\theta) = 1-\alpha$, then $\lim_{n\to\infty} \nu^{\infty}_{\theta}(\{s^{\infty} : \theta \in \mathcal{C}_n\}) = 1-\alpha$: The CLT implies that

$$\lim_{n\to\infty} \nu^{\infty}_{\theta}(\{s^{\infty} : \theta \in \mathcal{C}_n\}) = \boldsymbol{N}_J(c_\theta\sigma_\theta;\Lambda_\theta) = 1-\alpha.$$

Step 8: If $0 < \alpha < \frac{1}{2}$ and $\Lambda_\theta \neq 0$, then $\boldsymbol{N}_J(c_\theta\sigma_\theta;\Lambda_\theta) = 1-\alpha$: $\Lambda_\theta \neq 0 \implies \sigma(\Lambda_\theta) \neq 0$. Wlog let $\sigma_1(\Lambda_\theta) > 0$. Then $c \longmapsto \boldsymbol{N}_J(c\sigma_\theta;\Lambda_\theta)$ is continuous and strictly increasing on $c \geq 0$.

Argue that $\boldsymbol{N}_J(0;\Lambda_\theta) < 1-\alpha$: Let Z be multivariate normal with mean 0 and covariance matrix $\Lambda_\theta \neq 0$. Then,

$$\begin{aligned} \boldsymbol{N}_J(0;\Lambda_\theta) &= \Pr(X \leq 0) = \Pr(X_1 \leq 0)\Pr(X_2,...,X_J \leq 0 \mid X_1 \leq 0) \\ &\leq \Pr(X_1 \leq 0) = \tfrac{1}{2} < 1-\alpha. \end{aligned}$$

By Step 1, $\lim_{c\to\infty} \boldsymbol{N}_J(c\sigma_\theta;\Lambda_\theta) > 1-\alpha$. Therefore, $N_J(c\sigma_\theta;\Lambda_\theta) = 1-\alpha$ has a solution $c > 0$, and $c = c_\theta$ necessarily.

Step 9: If $\boldsymbol{N}_J(c_{\overline{\theta}}\sigma_{\overline{\theta}};\Lambda_{\overline{\theta}}) = 1-\alpha$ for some $\overline{\theta} \in \Theta$, then $\lim_{n\to\infty} \inf_{\theta\in\Theta} \nu^{\infty}_{\theta}(\{s^{\infty} : \theta \in \mathcal{C}_n\}) = 1-\alpha$: Note that

$$\begin{aligned} \limsup_{n\to\infty} \inf_{\theta\in\Theta} \nu^{\infty}_{\theta}(\{s^{\infty} : \theta \in \mathcal{C}_n\}) &\leq \limsup_{n\to\infty} \nu^{\infty}_{\overline{\theta}}\left(\left\{s^{\infty} : \overline{\theta} \in \mathcal{C}_n\right\}\right) = 1-\alpha \\ &\leq \liminf_{n\to\infty} \inf_{\theta\in\Theta} \nu^{\infty}_{\theta}(\{s^{\infty} : \theta \in \mathcal{C}_n\}) \end{aligned}$$

where the equality follows from Step 7 and the last inequality follows from (3.11).
■

C. Appendix: Details for the binary example

Proof of (4.4): For any λ in $[-1,0]$, define $\Lambda(\lambda) = \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}$, and $c(\lambda)$ by

$$\boldsymbol{N}_2\left((c(\lambda), c(\lambda)); \Lambda(\lambda)\right) = 0.95.$$

Then $\lambda \longmapsto c(\lambda)$ is (strictly) decreasing on $[-1,0]$ because $\boldsymbol{N}_2(\cdot; \Lambda(\lambda)) \overset{\lambda}{\nearrow}$.[30] It follows that $c(\lambda) \overset{\lambda}{\searrow}$. In addition, $\lambda \longmapsto c(\lambda)$ is continuous on $[-1,0]$.[31]

Fix $\alpha = .05$. For θs such that one or more of the variances $var_\theta(A_1)$ and $var_\theta(A_2)$ vanish, then, as in the Jovanovic example, the dimensionality is reduced below 2 and closed-form expressions can be derived.

For θs satisfying $0 < \theta_1 < \theta_2 < 1-$ one has $\sigma_\theta >> 0$ and

$$\boldsymbol{N}_2(c\sigma_\theta; \Lambda_\theta) = \boldsymbol{N}_2\left((c,c); \Lambda(\lambda_\theta)\right),$$

where

$$\lambda'_\theta = -\left[\frac{\theta_1}{1-\theta_1} \cdot \frac{1-\theta_2}{\theta_2}\right]^{1/2}. \tag{C.1}$$

Thus $c_\theta = c(\lambda'_\theta)$, and from the preliminary arguments above, $c_{(\theta_1,\theta_2)}$ is increasing in θ_1 and decreasing in θ_2, and $c_{(\theta_1,\theta_2)}$ varies continuously with θ in this "interior" region. In addition, because $-1 < \lambda'_\theta < 0$, infer that

$$c(0) < c_\theta < c(-1), \tag{C.2}$$

and

$$c(0) = \lim_{\theta_1 \searrow 0} c_{(\theta_1,\theta_2)}, \quad \lim_{\theta_1 \nearrow \theta_2} c_{(\theta_1,\theta_2)} = c(-1).$$

Finally, note that: (1) $c(-1)$ is defined by $\boldsymbol{N}_2\left((c(-1), c(-1)); \Lambda(-1)\right) = 1-\alpha$. Because $\Lambda(-1)$ is singular, any underlying r.v. $Z = (Z_1, Z_2)$ satisfies $Z_1 = -Z_2$ a.s. Accordingly, $c(-1)$ is such that a standard 1-dimensional normal variable Z_1 satisfies $-c(-1) \leq Z_1 \leq c(-1)$ with probability $1-\alpha$; in other words, given $\alpha = .05$, $c(-1) = 1.96$. (2) $c(0)$ is defined by $\boldsymbol{N}_2\left((c(0), c(0)); \Lambda(0)\right) = .95$ or $\boldsymbol{N}_1(c(0); 1) = [.95]^{1/2} \simeq .9747$, which gives $c(0) = 1.955$. ■

[30] The simple intuition is that the probability of both component r.v.s falling below (in a vector sense) any given $\beta \in \mathbb{R}^2$ is large when the components move together, or are less negatively correlated. See Muller and Scarsini (2000, Theorem 4.2) for a formal result.

[31] A question may arise for $\lambda = -1$ because $\Lambda(-1)$ is singular. Thus here are some details. By the noted monotonicity, $\lim_{\lambda \searrow -1} c(\lambda) \leq c(-1)$; and the opposite inequality follows from Step 4 in the proof of Theorem 3.2.

D. Appendix: Proofs for covariates

We outline the proof of Theorem 6.1, which adapts the arguments for the no-covariate case. We use two lemmas that highlight the added steps needed to accommodate covariates. The assumption that each x appears infinitely often is maintained.

Write $S^\infty = S_1 \times S_2 \times ...$, where $S_i = S$ for all i. For any $I \subset \{1, 2, ...\}$, denote by Σ_I the σ-algebra generated by (Borel measurable) cylinders of the form $\Pi_{i \in I} A_i \times \Pi_{i \notin I} S_i$, where $A_i \subset S_i = S$. Say that $B_1, B_2 \subset S^\infty$ are *orthogonal* if they depend on different experiments in the sense that $B_1 \in \Sigma_{I_1}$ and $B_2 \in \Sigma_{I_2}$ for some disjoint I_1 and I_2.

Lemma D.1. $\nu_\theta^\infty \left(\bigcap_{k=1}^{K} B_k \mid x^\infty \right) = \prod_{k=1}^{K} \nu_\theta^\infty \left(B_k \mid x^\infty \right)$ *if* $B_1, ..., B_K$ *are pairwise orthogonal.*

Proof. Let $B_k \in \Sigma_{I_k}$, $k = 1, ..., K$, where $I_1, ..., I_K$ are pairwise disjoint. Observe that

$$\begin{aligned}
\nu_\theta^\infty \left(\bigcap_{k=1}^{K} B_k \mid x^\infty \right) &= m_\theta^\infty \left(\left\{ u^\infty \in U^\infty : \prod_{i=1}^{\infty} G\left(u_i \mid \theta, x_i\right) \subset \bigcap_{k=1}^{K} B_k \right\} \right) \\
&= m_\theta^\infty \left(\bigcap_{k=1}^{K} \left\{ u^\infty \in U^\infty : \prod_{i \in I_k} G\left(u_i \mid \theta, x_i\right) \subset B_k \right\} \right) \\
&= \prod_{k=1}^{K} m_\theta^\infty \left(\left\{ u^\infty \in U^\infty : \prod_{i \in I_k} G\left(u_i \mid \theta, x_i\right) \subset B_k \right\} \right) \\
&= \prod_{k=1}^{K} \nu_\theta^\infty \left(B_k \mid x^\infty \right). \quad \blacksquare
\end{aligned}$$

Lemma D.2. *Let* $\Lambda_{\theta_n, x_k} \to \Lambda_k \in \mathbb{R}^{J \cdot J}$ *for each* $k = 1, ..., |X|$, *and let* Λ *be the* $|X|\, J$*-by-*$|X|\, J$ *block diagonal matrix where* $\Lambda_1, ..., \Lambda_{|X|}$ *are the blocks. Also assume* $c_n \to c \in \mathbb{R}^{|X| J}$. *Then*

$$\nu_{\theta_n}^\infty \left(\cap_{k=1}^{|X|} \cap_{j=1}^{J} \left\{ s^\infty : \sqrt{n} \left[\nu_{\theta_n} \left(A_j \mid x_k \right) - \Psi_n \left(s^\infty, x^\infty \right) \left(A_j \mid x_k \right) \right] \leq c_{nkj} \right\} \right) \to \boldsymbol{N}_{|X| J} \left(c; \Lambda \right).$$

Proof. The events $\cap_{j=1}^{J}\{s^\infty : \sqrt{n}\left[\nu_{\theta_n}(A_j \mid x_k) - \Psi_n(s^\infty, x^\infty)(A_j \mid x_k)\right] \le c_{nkj}\}$, $k = 1, ..., |X|$, are pairwise orthogonal. Therefore, by the preceding lemma,

$$
\begin{aligned}
&\nu_{\theta_n}^\infty\left(\cap_{k=1}^{|X|}\cap_{j=1}^{J}\left\{s^\infty : \sqrt{n}\left[\nu_{\theta_n}(A_j \mid x_k) - \Psi_n(s^\infty, x^\infty)(A_j \mid x_k)\right] \le c_{nkj}\right\}\right) \\
= &\prod_{k=1}^{|X|}\nu_{\theta_n}^\infty\left(\cap_{j=1}^{J}\left\{s^\infty : \sqrt{n}\left[\nu_{\theta_n}(A_j \mid x_k) - \Psi_n(s^\infty, x^\infty)(A_j \mid x_k)\right] \le c_{nkj}\right\}\right) \\
\to &\prod_{k=1}^{|X|}\boldsymbol{N}_J(c_k; \Lambda_k) = \boldsymbol{N}_{|X|J}(c; \Lambda).
\end{aligned}
$$

Here, $c_{nkj} \in \mathbb{R}$, $c_n = (c_{nkj})_{k,j} \in \mathbb{R}^{|X|J}$, $c_k \in \mathbb{R}^J$ and $c = (c_k)_k \in \mathbb{R}^{|X|J}$. ■

The rest of the proof of Theorem 6.1 is similar to that for the no-covariate case.

Proof of Corollary 6.2: Let $X_{\inf}^\infty$ be the set of all $x^\infty \in X^\infty$ for which each value in X appears infinitely often. Then,

$$
\nu_\theta^\infty(\cdot) = \int_{X_{\inf}^\infty}\nu_\theta^\infty(\cdot \mid x^\infty)\, dq^\infty(x^\infty), \text{ and}
$$

$$
\liminf_{n\to\infty}\inf_{\theta\in\Theta}\nu_\theta^\infty(\theta \in \mathcal{C}_n) \ge \int_{X_{\inf}^\infty}\liminf_{n\to\infty}\inf_{\theta\in\Theta}\nu_\theta^\infty(\theta \in \mathcal{C}_n \mid x^\infty)\, dq^\infty(x^\infty) \ge 1 - \alpha.
$$

To show the equality assertion, let $\overline{\theta} \in \Theta$ satisfy $\Lambda_{\overline{\theta}} \ne 0$. Then,

$$
\begin{aligned}
\limsup_{n\to\infty}\inf_{\theta\in\Theta}\nu_\theta^\infty(\theta \in \mathcal{C}_n) &\le \limsup_{n\to\infty}\nu_{\overline{\theta}}^\infty\left(\overline{\theta} \in \mathcal{C}_n\right) \\
&= \limsup_{n\to\infty}\int_{X_{\inf}^\infty}\nu_{\overline{\theta}}^\infty\left(\overline{\theta} \in \mathcal{C}_n \mid x^\infty\right) dq^\infty(x^\infty) \\
&\le \int_{X_{\inf}^\infty}\limsup_{n\to\infty}\nu_{\overline{\theta}}^\infty\left(\overline{\theta} \in \mathcal{C}_n \mid x^\infty\right) dq^\infty(x^\infty) = 1 - \alpha \\
&\le \liminf_{n\to\infty}\inf_{\theta\in\Theta}\nu_\theta^\infty(\theta \in \mathcal{C}_n). \quad ■
\end{aligned}
$$

E. Appendix: Latent variables robustified

Currently, we define models via primitives (S, U, G, Θ, m), including, in particular, the probability measures m_θ on U for every θ. Model incompleteness arises only because of the multiplicity of equilibria and ignorance of selection. Here we follow up on the remarks at the end of Section 2 and consider another source of incompleteness–limited understanding of the latent variables, which seems intuitive for variables that are not observed by the analyst. Formally, we suggest that this situation can be modeled as above except that every m_θ is a belief function rather than a measure. Also in this case we obtain belief functions ν_θ on S that satisfy a CLT which in turn can be used to construct robust confidence regions. Note that in the present context, robustness with regard to (limited) ignorance about latent variables is desirable even if selection is well-understood, for example, if equilibria are unique.

Let S, U, G and Θ be as before. Instead of adopting m as another primitive, we derive it from more basic primitives. Thus let the tuple $\left(\widehat{U}, \Gamma, \widehat{m}\right)$ describe the (limited) understanding of latent variables, where $\widehat{U}$ is Polish, $\widehat{m} = (\widehat{m}_\theta)_{\theta\in\Theta}$, each $\widehat{m}_\theta$ is a Borel probability measure on $\widehat{U}$, and $\Gamma(\cdot \mid \theta) : \widehat{U} \rightsquigarrow U$ is weakly measurable. (The assumption that the same parameters θ enter here is without loss of generality since one could expand the parameter space Θ as needed.) Thus probabilistic knowledge is assumed on $\widehat{U}$ which, via the correspondence Γ, provides only coarse information about the latent variables $u \in U$. Paralleling (2.4), the elements $\widehat{U}$, Γ and $\widehat{m}$ induce (for each θ) a belief function on U, denoted by m_θ and given by

$$m_\theta(Y) = \widehat{m}_\theta(\{\widehat{u} : \Gamma(\widehat{u} \mid \theta) \subset Y\}), \; Y \subset U. \tag{E.1}$$

Consider now the model (S, U, G, Θ, m) where $m = (m_\theta)_{\theta\in\Theta}$ and each m_θ is a belief function on U. Define ν_θ on (subsets of) S exactly as in (2.4), that is,

$$\nu_\theta(A) = m_\theta(\{u : G(u \mid \theta) \subset A\}), \;\; A \subset S.$$

Then ν_θ is a belief function: To see this, take $Y = \{u : G(u \mid \theta) \subset A\}$ in (E.1) to derive

$$\begin{aligned} \nu_\theta(A) &= \widehat{m}_\theta(\{\widehat{u} : \Gamma(\widehat{u} \mid \theta) \subset \{u : G(u \mid \theta) \subset A\}\}) \\ &= \widehat{m}_\theta\left(\{\widehat{u} : \cup_{u\in\Gamma(\widehat{u}\mid\theta)} G(u \mid \theta) \subset A\}\right) \\ &= \widehat{m}_\theta\left(\left\{\widehat{u} : \widehat{G}(\widehat{u} \mid \theta) \subset A\right\}\right), \end{aligned}$$

where $\widehat{G}(\cdot \mid \theta) : \widehat{U} \rightsquigarrow S$ is the "composition" of G and Γ defined by

$$\widehat{G}(\widehat{u} \mid \theta) = \cup_{u \in \Gamma(\widehat{u}|\theta)} G(u \mid \theta). \tag{E.2}$$

Thus $\left(\widehat{U}, \widehat{G}, \widehat{m}\right)$ generates ν_θ exactly as in (2.4), which proves that ν_θ is a belief function.

Because it depends only on having a belief function ν_θ on S for each parameter θ, the inference method described in Section 3 applies without modification. Only the interpretation must be modified slightly to reflect the fact that there are now two sources of model incompleteness or areas of ignorance: in addition to ignorance of how outcomes are selected from $G(u \mid \theta)$, there is also the coarse information about u due to $\Gamma(\cdot \mid \theta)$ being set-valued. The (extended) inference method is robust to heterogeneity and dependence across experiments in both selection and in the unknown fine details regarding latent variables in U.

In a sense there is nothing new above since one could take $\left(S, \widehat{U}, \widehat{G}, \Theta, \widehat{m}\right)$ as the model. However, in applications the identity of $\widehat{U}$, Γ and $\widehat{m}$ underlying the modeling of latent variables in U may not be clear. In those cases, the analyst might begin with the reduced form model (S, U, G, Θ, m) where each m_θ is a belief function. One can view the preceding as providing a rationale for doing so when the underlying primitives are not clear. Specification of m_θ may involve some arbitrariness but this is the case also when probability distributions are adopted for latent variables.

F. Appendix: Implementation

Construction of our confidence region requires computing the belief function ν_θ and the critical value c_θ. For simple examples, one may compute ν_θ analytically. In general, it can be computed using a simulation procedure. Once ν_θ is obtained, the critical value c_θ can be computed using another simulation procedure, as demonstrated by the Monte Carlo experiments in Section 5. Below, we illustrate the simulation procedures using the entry game example studied by Bresnahan and Reiss (1990), Berry (1992), and Ciliberto and Tamer (2009); the latter is CT henceforth.

Suppose there are K firms that are potential entrants into markets $i = 1, 2, \cdots$. For each i, we let $s_i = (s_{i1}, \cdots, s_{iK}) \in \{0,1\}^K$ denote the vector of entry decisions made by the firms. For firm k in market i, CT consider the following profit function specification:

$$\pi_k(s_i, x_i, u_i; \theta) = \left(v_i'\alpha_k + z_{ik}'\beta_k + w_{ik}'\gamma_k + \sum_{j\neq k} \delta_j^k s_{ij} + \sum_{j \neq k} z_{ij}'^k{}_j s_{ij} + u_{ik} \right) s_{ik},$$

where v_i is a vector of market characteristics, $z_i = (z_{i1}, \cdots, z_{iK})$ is a matrix of firm characteristics that enter the profits of all firms in the market, while $w_i = (w_{i1}, \cdots, w_{iK})$ is a matrix of firm characteristics such that w_{ik} enters firm k's profit but not other firms' profits. We let x_i collect v_i, z_i, and w_i and stack them as a vector. The unobservable payoff shifters $u_i = (u_{i1}, \cdots, u_{iK})$ follow a multivariate normal distribution $N(0, \Sigma)$ and vary across markets in an i.i.d. way.[32] The structural parameter θ includes Σ and the parameters associated with the profit functions: $\{\beta_k, \gamma_k, \{\delta_j^k, \phi_j^k\}_{j\neq k}\}_{k=1}^K$.

In this example, firm k's profit from not entering the market is 0. Hence, the set of pure-strategy Nash equilibria is given by

$$G(u_i|\theta, x_i) = \{s_i \in S : \pi_k(s_i, x_i, u_i; \theta) \geq 0, \forall k = 1, \cdots, K\}. \tag{F.1}$$

Suppose that a sample $\{(s_i, x_i), i = 1, \cdots, n\}$ of size n is available. Let A be a subset of $S = \{0, 1\}^K$. CT only use singleton events $A = \{s\}, s \in S$ and provide a simulation procedure to calculate $\nu_\theta(A|x)$ and its conjugate (called $\mathbf{H}_1$ and $\mathbf{H}_2$ in their paper). In general, one can use any event $A \subset S$ for inference, and we describe a simulation procedure for this general setting below.

Recall that the belief function of event A conditional on x was given by

$$\nu_\theta(A|x) = m_\theta(\{u \in U : G(u|\theta, x) \subset A\}). \tag{F.2}$$

Hence, a natural way to approximate $\nu_\theta(A|x)$ for any $A \subset S$ is to simulate u from the parametric distribution m_θ and calculate the frequency of the event $G(u|\theta, x) \subset A$. We summarize the procedure below.

Simulation procedure 1

Step 1 Fix the number of draws R. Given Σ, draw random vectors $u^r = (u_1^r, \cdots, u_K^r)$, $r = 1, \cdots, R$, from $N(0, \Sigma)$.

[32] In the context of entry games played by airlines, CT model u_{ik} as a sum of independent normal random variables: firm-specific unobserved heterogeneity, market-specific unobserved heterogeneity, and airport-specific unobserved heterogeneity. This can also be handled by relaxing the i.i.d. assumption on m_θ^∞.

Step 2 For each $(s, x, u^r) \in S \times X \times U$, calculate

$$I(s, x, u^r; \theta) = \begin{cases} 1 & \pi_k(s, x, u^r; \theta) \geq 0, \ \forall k, \\ 0 & \text{otherwise.} \end{cases}$$

That is, $I(s, x, u^r) = 1$ if s is a pure strategy Nash equilibrium under (x, u^r) and θ.

Step 3 Compute the frequency of event $G(u^r|\theta, x) \subseteq A$ across simulation draws by computing that of $A^c \subseteq G^c(u^r|\theta, x)$:

$$\nu_\theta^R(A|x) = \frac{1}{R} \sum_{r=1}^{R} \prod_{s \in A^c} (1 - I(s, x, u^r; \theta)). \tag{F.3}$$

After implementing the simulation procedure above, one can evaluate the test statistic $T_n(\theta)$ in (6.3). The remaining task is to compute the critical value c_θ, which can be done by feeding Λ_θ into a commonly-used simulator for multivariate normal random vectors.

Simulation procedure 2

Step 1 Compute the covariance matrix Λ_θ, which is a $|X| J$-by-$|X| J$ block-diagonal matrix where $\Lambda_{\theta,x_1}, ..., \Lambda_{\theta,x_{|X|}}$ are the blocks.:

The (j, j')-th entry of each block $\Lambda_{\theta,x}$ is the covariance matrix, conditional on x: $(\Lambda_{\theta,x})_{jj'} = cov_\theta (A_j, A_{j'} \mid x)$, where $cov_\theta (A_j, A_{j'} \mid x)$ is calculated as in (6.1) while using the approximated belief function ν_θ^R obtained in simulation procedure 1.

Step 2 Decompose Λ_θ as LDL' for a lower triangular matrix L and a diagonal matrix D.

Step 3 Generate $w^r \overset{i.i.d.}{\sim} N(0, I_{|X|J})$ for $r = 1, \cdots, R$. Generate $z^r = LD^{1/2}w^r$, $r = 1, \cdots, R$.

Step 4 Calculate c_θ as the $1 - \alpha$ quantile of $\max_{k=1,\cdots,|X|J} z_k/\sigma_{\theta,k}$:

$$c_\theta = \min \left(c \geq 0 : \frac{1}{R} \sum_{r=1}^{R} I(\max_{k=1,\cdots,|X|J} z_k^r/\sigma_{\theta,k} \leq c) \geq 1 - \alpha \right).$$

Steps 2-3 in simulation procedure 2 are based on the Geweke-Hajivassiliou-Keane (GHK) simulator. The GHK simulator is widely used in econometrics (see, for example, Hajivassiliou, McFadden, and Ruud (1996) for details). The only difference from the standard GHK-simulator is Step 2, in which we recommend to use the LDL decomposition instead of Cholesky decomposition. This is because Λ_θ may only be positive semidefinite.

Simulation procedure 2 yields a critical value c_θ. Hence, one can determine whether or not a value of the structural parameter should be included in the confidence region by checking if $T_n(\theta) \leq c_\theta$ holds. For constructing a confidence region, one needs to repeat the procedures above for different values of $\theta \in \Theta$. To save computational costs, one can draw $\{(u_1^r, \cdots, u_K^r)\}_{r=1}^R$ and $\{w^r\}_{r=1}^R$ only once and use them repeatedly across all values of θ.

A final remark is that the procedures described above extend to other settings. In other models, the researcher may use a different solution concept (e.g. pairwise stability of networks) that defines the correspondence $G(\cdot|\theta, x)$, or a different parametric specification for the latent variables in the payoff function (e.g. random coefficients following a mixed logit specification). In such cases, one need modify only Steps 1 and 2 in simulation procedure 1.

References

[1] C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis*, 3rd ed. Springer, 2006.

[2] D.W.K. Andrews and P.J. Barwick, Inference for parameters defined by moment inequalities: a recommended moment selection procedure, *Econometrica* 80 (2012), 2805–2826.

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

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

[5] A. Aradillas-Lopez and E. Tamer, The identification power of equilibrium in simple games. *J. Bus. Econ. Statist.* 26 (2008), 261–283.

[6] P. Bajari, H. Hong, and S.P. Ryan, Identification and estimation of a discrete game of complete information, *Econometrica* 78 (2010), 1529-68.

[7] A. Beresteanu, I. Molchanov, and F. Molinari, Sharp identification regions in models with convex moment predictions, *Econometrica* 79 (2011), 1785–1821.

[8] A. Beresteanu and F. Molinari, Asymptotic properties for a class of partially identified models, *Econometrica* 76 (2008), 763-84.

[9] S. Berry, Estimation of a model of entry in the airline industry, *Econometrica* 60 (1992), 889-917.

[10] S. Berry and E. Tamer, Identification in models of oligopoly entry, *Advances in Economics and Econometrics. Ninth World Congress of the Econometric Society* (2006) 2, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press, 46-85.

[11] T.F. Bresnahan and P.C. Reis, Empirical models of discrete games, *J. Econometrics* 48 (1991), 57-81.

[12] T.F. Bresnahan and P.C. Reis, Entry in monopoly markets, *Rev. Econ. Stud.* 57 (1990), 531-53.

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

[14] I.A. Canay, EL inference for partially identified models: large deviations optimality and bootstrap validity, *J. Econometrics* 156 (2010), 408-425.

[15] V. Chernozhukov, H. Hong and E. Tamer, Estimation and confidence regions for parameter sets in econometric models, *Econometrica* 75 (2007), 1243-84.

[16] F. Ciliberto and E. Tamer, Market structure and multiple equilibria in airline markets, *Econometrica* 77 (2009), 1791-1828.

[17] A. Dasgupta, *Asymptotic Theory of Statistics and Probability*, Springer, 2008.

[18] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Statist.* 38 (1967), 325-39.

[19] A. P. Dempster, A generalization of Bayesian inference, *J. Royal Stat. Soc.* Series B (1968) 30, 205-32.

[20] L.G. Epstein and K. Seo, Exchangeable capacities, parameters and incomplete theories, *J. Econ. Theory* 157 (2015), 879-917.

[21] K. Ferentinos, On Tchebycheff's type inequalities, *Trabajos de Estadística e Investigación Operativa* 33 (1982), 125-32.

[22] A. Galichon and M. Henry, A test of non-identifying restrictions and confidence regions for partially identified parameters, *J. Econometrics* 152 (2009), 186-96.

[23] A. Galichon and M. Henry, Set identification in models with multiple equilibria, *Rev. Econ. Stud.* 78 (2011), 1261-98.

[24] A. Galichon and M. Henry, Dilation bootstrap, *J. Econometrics* 177 (2013), 109-15.

[25] A. Genz, Numerical computation of rectangular bivariate and trivariate normal and t probabilities, *Stat. Comput.* 14 (2004), 251-260.

[26] P.A. Haile and E. Tamer, Inference with an incomplete model of English auctions, *J. Pol. Econ.* 111 (2003), 1-51.

[27] V. Hajivassiliou, D. McFadden, and P. Ruud, Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results, *J. Econometrics* 72 (1996), 85–134.

[28] G.W. Imbens, and C.F. Manski, Confidence intervals for partially identified parameters, *Econometrica* 72 (2004), 1845-57.

[29] N. Jenish and I.R. Prucha, Central limit theorems and uniform laws of large numbers for arrays of random fields, *J. Econometrics* 150 (2009), 86-98.

[30] B. Jovanovic, Observable Implications of models with multiple equilibria, *Econometrica* 57 (1989), 1431-37.

[31] C.F. Manski, *Partial Identification of Probability Distributions*, Springer, 2003.

[32] F. Maccheroni and M. Marinacci, A strong law of large numbers for capacities, *Ann. Prob.* 33 (2005), 1171-1178.

[33] K. Menzel, Robust decisions for incomplete structural models of social interactions, NYU, 2011.

[34] Y. Miyauchi, Structural estimation of a pairwise stable network with nonnegative externality, MIT, 2014.

[35] I. Molchanov, *Theory of Random Sets*, Springer, 2005.

[36] A. Muller and M. Scarsini, Some remarks on the supermodular order, *J. Multivariate Anal.* 73 (2000), 107-19.

[37] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, 1994.

[38] R. Phelps, *Lectures on Choquet's Theorem*, Springer, 2001.

[39] F. Philippe, G. Debs and J.Y. Jaffray, Decision making with monotone lower probabilities of infinite order, *Math. Oper. Res.* 24 (1999), 767-84.

[40] G. Shafer, *A Mathematical Theory of Evidence*, Princeton U. Press, Princeton, 1976.

[41] G. Shafer, Belief functions and parametric models (with commentary), *J. Royal Stat. Soc. Series B* 44 (1982), 322-52.

[42] S. Sheng, A structural econometric analysis of network formation games, UCLA, 2014.

[43] A. Soetevent, and P. Kooreman, A discrete choice model with social interactions: with an application to high school teen behavior, *J. App. Econ.* 22 (2007), 599–624.

[44] J. Stoye, More on confidence regions for partially identified parameters, *Econometrica* 77 (2009), 1299-1315.

[45] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010), 167–95.

[46] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003), 147-65.

[47] K. Uetake and Y. Wanatabe, Entry by merger: estimates from a two-sided matching model with externalities, 2012.

[48] H. White, *Asymptotic Theory for Econometricians,* Academic Press, 2001.