

# 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 CWP65/15

# Robust Confidence Regions for Incomplete Models\*

Larry G. Epstein      Hiroaki Kaido      Kyoungwon Seo

September 27, 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. Yi Zhang provided excellent research assistance. 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, Kei Hirano, Boyan Jovanovic, Konrad Menzel, Yuhei Miyauchi, Francesca Molinari, Whitney Newey, Joris Pinkse, Jack Porter, Zhongjun Qu, Joerg Stoye and from audiences at Penn State, MIT, BU, Wisconsin, UNC, Korea U., Sungkyunkwan U., Osaka U., Tokyo, U. of Arizona, the Econometric Society World Congress 2015, SETA, and CEME. First version: April 2015.

# 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  $\theta \in \Theta$  and the realization  $u \in U$  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  $S$ . Importantly, the model is silent on how the realized outcome  $s$  is selected from  $G(u|\theta)$ . The object of interest is  $\theta$ .

An important example (based on Jovanovic (1989)) is a discrete normal form game where  $u$  is a latent variable,  $\theta$  is a parameter in the payoff matrix,  $S$  is the set of (pure) strategies, and  $G(u|\theta)$  is the set of Nash equilibrium (pure) strategy profiles for the given  $u$  and  $\theta$ . The multiplicity of equilibria and the absence of a theory of equilibrium selection lead to a multi-valued prediction. Alternative solution concepts can also be adopted. For example, a refinement of Nash equilibrium can be modeled via a suitable sub-correspondence; and a weaker solution concept such as rationalizability can be modeled via a suitable super-correspondence. Whichever solution concept is adopted, there is complete ignorance about selection.

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.*"

Our starting point is to observe that agnosticism about selection has implications that have been overlooked in the literature. To elaborate, think of a number of experiments, or random events, indexed by  $i = 1, 2, \dots$ , each of which may be described as above, for a common  $\Theta$ ,  $G$  and  $S$ ; for example, each experiment could correspond to a different market where an entry game is played. Then, given  $\theta$ , each infinite sequence of unobserved variables  $u^\infty \equiv (u_1, u_2, \dots)$  generates a sample  $(s_1, s_2, \dots)$  of outcomes, where  $s_i \in G(u_i|\theta)$  for all  $i$ . A prevalent assumption

in the literature is the availability of samples of outcomes that are well-behaved in the sense that sample averages converge (ergodicity) and obey classical limit theorems. This restricts heterogeneity and dependence of outcomes across experiments in an analytically convenient way because then the sampling distribution around the limiting empirical frequency of any event can be approximated in large samples by a normal distribution. A leading example is the assumption that samples are i.i.d. An alternative and more general assumption is that samples are stationary and strongly mixing (Chernozhukov, Hong, and Tamer 2004, Andrews and Soares 2010).

Though seemingly standard and innocuous, the assumption that samples admit classical limit theorems becomes subtle given incompleteness of the model and the declared agnosticism about selection. 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  $G(u_i | \theta)$  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 particular, one cannot be sure even that empirical frequencies converge, thus limiting the applicability (or robustness) of existing inference methods. Accordingly, in this paper we develop a new inference method that is robust to *heterogeneity and dependence of an unknown form*.

Before outlining our approach, we emphasize that our inference method is not only of theoretical interest—it is applicable to a number of empirical models in the literature. These models include: entry games with multiple Nash equilibria (Bresnahan and Reiss 1990 and 1991, Berry 1992, Jia 2008, Ciliberto and Tamer 2009);<sup>1</sup> first-price auctions (Haile and Tamer 2003); innovation and product variety (Eizenberg 2014); sincere versus strategic voting (Kawai and Watanabe 2013); municipal mergers (Weese 2015); discrete-choice with social interactions (Soetevent and Kooreman 2007); matching with externalities (Uetake and Watan-

---

<sup>1</sup>See Section 4.1 for a canonical example of such entry games due to Jovanovic (1989). Note also that Appendix F provides a detailed guide to implementing our method in the context of the entry games in Bresnahan and Reiss (1990,1991), Berry (1992), and Ciliberto and Tamer (2009).

abe 2012); and friendship networks (Miyachi 2014). In these models, incomplete structures arise for different reasons. For example, in discrete-game models, incompleteness arises because of the modeler’s agnosticism about the mechanism for selecting from among multiple equilibria; while in Haile and Tamer’s auction model, it is due to the modeler’s agnosticism about the precise game form underlying the auction data in her sample, which leads her to adopt only weak assumptions about bidders’ behavior. Solution concepts also vary—for example, in matching and network formation models, pairwise stability or similar concepts are used.

Here is a sketch of how we proceed (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  $\theta$ , robustness to an unknown form of dependence implies that if for each  $i$ ,  $s_i$  is a conceivable outcome in the  $i$ th experiment (in isolation) given  $u_i$ , then  $(s_1, s_2, \dots)$  is a conceivable sequence given  $(u_1, u_2, \dots)$ . Thus, without further assumptions, the model predicts that the selected outcomes  $(s_1, s_2, \dots)$  take their values in the Cartesian product of  $G(u_i|\theta)$ ,  $i = 1, 2, \dots$ , and we define:

$$G^\infty(u_1, \dots, u_i, \dots | \theta) \equiv \prod_{i=1}^\infty G(u_i | \theta). \quad (1.1)$$

Note that experiments are indistinguishable in the sense that the same correspondence  $G(\cdot | \theta)$  applies to each experiment. However, even if  $G(u_i | \theta) = G(u_j | \theta)$ , as when  $u_i = u_j$ , any outcome in  $G(u_i | \theta)$  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  $\theta$ . Assume that  $u^\infty$  jointly follows a parametric distribution  $m_\theta^\infty$ , the i.i.d. product of the measure  $m_\theta$  on  $U$ . (In fact, as indicated below, more general probability laws on  $U^\infty$  can be adopted.) For each given  $u^\infty$ , any probability distribution  $P_{u^\infty}$  supported on  $G^\infty(u^\infty | \theta)$  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

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

Note that because  $\Delta(G^\infty(u^\infty | \theta))$  equals the entire simplex of distributions on  $\prod_{i=1}^\infty G(u_i | \theta)$ , including both nonidentical product measures and nonproduct

measures, the set  $\mathcal{P}_\theta$  accommodates many forms of heterogeneity and dependence across experiments even given  $u^\infty$ .

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  $\nu_\theta^\infty$ , defined, for every measurable  $B \subset S^\infty$ , by

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

is a *belief function* on  $S^\infty$ .<sup>2</sup> We exploit this and prove a (new) central limit theorem (CLT) for each belief function  $\nu_\theta^\infty$  and thus indirectly also for each set  $\mathcal{P}_\theta$ . Then we show how, paralleling the classical analysis for complete models, the CLT can be used to construct suitably robust confidence regions for the unknown parameter  $\theta$ .

A confidence region  $\mathcal{C}_n$  is a set of parameter values constructed from a finite number of observations  $s_1, \dots, s_n$  such that, for each  $\theta$ , 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  $\nu_\theta^\infty(\{\theta \in \mathcal{C}_n\}) \rightarrow 1 - \alpha$ , 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

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

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 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.

---

<sup>2</sup>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).

## 1.2. Relation to the literature<sup>3</sup>

The analysis of incomplete economic models dates back at least to Wald (1950) who studies inference on parameters based on incomplete systems of equations where the number of endogenous variables exceeds the number of structural equations. He considers the special case where the inclusion  $s_i \in G(u_i | \theta)$  can be inverted to solve for  $u_i$  in the form  $u_i = g(s_i | \theta)$  for some function  $g$ . This structure does not intersect any of the applied models (including those based on multiple Nash equilibria) studied in the more recent literature. Jovanovic (1989) highlights the potential difficulty for identification raised by model incompleteness and provides a theoretical framework for studying the predictive content of such models. Bresnahan and Reiss (1990, 1991) and Berry (1992) 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. Tamer (2003) shows that one can (point or partially) identify structural parameters and mitigate the loss of information by using inequality restrictions on the probabilities of outcomes that are not predicted uniquely.

More recently, the theory of random sets (and induced capacities) has been applied to address identification and inference. 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, Chesher and Rosen 2014). 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 use the conjugate capacity whereby the capacity of  $A$  is measured by  $1 - \mu_\theta(S \setminus A)$ ; this defines a belief function which gives the lower envelope, and hence is directly relevant for studying the robust control of the asymptotic coverage probability. Another difference is that we focus explicitly on the entire sequence of experiments jointly. Our approach to inference is related to Beresteanu and Molinari (2008) in the sense that we both use generalized (albeit much different) limit theorems. An important difference is that they assume that the entire set of outcomes is observed for each experiment

---

<sup>3</sup>Those unfamiliar with the literature on partial identification may wish to skim or even skip this section on first reading.

rather than merely the selected outcome (for example, outcomes are interval-valued). Galichon and Henry (2006, 2009, 2013) study inference using a statistic based on capacities, but they also maintain the assumption that samples obey classical limit theorems. None of these papers addresses robustness with regard to unknown forms of heterogeneity and dependence.

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 do inference in incomplete models (Chernozhukov, Hong, and Tamer 2007, Andrews and Soares 2010, Bugni 2009, Andrews and Shi 2013). As noted above, these methods commonly assume that data are generated across experiments in such a way that classical limit theorems are applicable, which precludes robustness against unknown forms of 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.

The only paper of which we are aware that explicitly allows outcome sequences that do not necessarily obey classical limit theorems is Menzel (2011). He develops a computational tool to conduct robust Bayes inference assuming 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 unknown forms of heterogeneity and dependence.

A notable and desirable feature of our confidence region, besides robustness, 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  $\theta$ . 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  $\theta$ . Under the assumptions adopted in existing methods, the empirical frequency converges to a unique probability  $p(A)$ . The pointwise limiting distribution of the test statistic used to construct confidence regions changes depending on whether  $p(A)$  equals  $\nu_\theta(A)$ , or  $\nu_\theta^*(A)$ , or is in the interior of the interval, with the result that this limiting distribution depends discontinuously on the underlying data generating process.<sup>4</sup>

---

<sup>4</sup>For example, a commonly used test statistic is  $T_n(\theta) = \sqrt{n} \max\{\nu_\theta(A) - n^{-1} \sum_{i=1}^n I(s_i \in A), n^{-1} \sum_{i=1}^n I(s_i \in A) - \nu_\theta^*(A)\}$ . Let  $Z$  have the limiting distribution of  $\sqrt{n}(n^{-1} \sum_{i=1}^n I(s_i \in A) - p(A))$  under the assumption that a classical CLT applies. Then  $T_n(\theta)$  converges in distribution to  $-Z$ ,  $-\infty$ ,  $Z$ , or  $\max\{-Z, Z\}$  according as

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, it is also an arbitrary component of existing inference methods. We attribute this arbitrariness to the assumption that the empirical frequency converges to a unique limit. In contrast, we do not presume such convergence. Even so, inference on the structural parameter is possible because if  $\theta$  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  $\theta$ . Thus discontinuity of the limiting distribution does not arise.

Another reflection of the difference in the approach to inference adopted here is that while the notion of the "identified set" receives a great deal of attention in the literature, it does not play a role here. A brief remark may be helpful for readers familiar with the literature on partial identification. 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 assumptions that the outcome in each market is distributed according to some measure  $p \in \Delta(S)$  and that the outcome sequence is ergodic, then empirical frequencies converge almost surely to  $p$ , rendering  $p$  observable. The identified set, denoted  $\Theta^I(p)$ , consists of all  $\theta$  such that there exists a (suitably measurable) selection rule  $u \mapsto p_u \in \Delta(G(u | \theta))$  satisfying

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

which equates true and predicted empirical frequencies.<sup>5</sup> A number of papers describe (finite sample) estimators for  $\Theta^I$ ; 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 ergodic 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

---

$\nu_\theta(A) = p(A) < \nu_\theta^*(A)$ ,  $\nu_\theta(A) < p(A) < \nu_\theta^*(A)$ ,  $\nu_\theta(A) < p(A) = \nu_\theta^*(A)$ , or  $\nu_\theta(A) = p(A) = \nu_\theta^*(A)$ , respectively.

<sup>5</sup>See Beresteanu, Molchanov and Molinari (2011) and Galichon and Henry (2011), for example.

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.

Finally, 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.

## 2. The framework

Consider a setting with an infinite sequence of experiments (or random events), where  $S_i = S$  denotes the (finite) set of possible outcomes for the  $i$ th experiment. The economic model of each single experiment is described by  $(S, U, G, \Theta; m)$  with the following interpretation and restrictions.  $\Theta$  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  $m_\theta$ , which is known up to the parameter  $\theta$ ; let  $m = (m_\theta)_{\theta \in \Theta}$ . Finally, for each  $\theta \in \Theta$ ,  $G(\cdot | \theta) : U \rightsquigarrow S$  is a correspondence that describes the set of outcomes for each given  $u$  and parameter  $\theta$ . 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  $\theta$ ,  $G(\cdot | \theta)$  is weakly measurable.<sup>6</sup>

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.<sup>7</sup> Accordingly, consider the tuple  $(S^\infty, U^\infty, G^\infty, \Theta; m^\infty)$ .

---

<sup>6</sup>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  $\Gamma$  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).

<sup>7</sup>Our formal model of a single experiment is adapted from Koopmans and Reiersol (1950) and Jovanovic (1989); a similar structure is employed by Galichon and Henry (2009, 2011).

The meaning of and rationale for  $S^\infty$  and  $U^\infty$  are clear;<sup>8</sup> they have generic elements  $s^\infty = (s_1, s_2, \dots)$  and  $u^\infty = (u_1, u_2, \dots)$  respectively. By  $m^\infty$ , an abbreviation for  $(m_\theta^\infty)_{\theta \in \Theta}$ , we mean that, conditional on  $\theta$ , unobserved variables are distributed i.i.d. across experiments according to  $m_\theta$ . The parameter set  $\Theta$  remains unchanged and parameters are assumed to be constant across experiments. The remaining component  $G^\infty$ , a key to our approach, is, for each  $\theta$ , a correspondence  $G^\infty(\cdot | \theta) : U^\infty \rightsquigarrow S^\infty$  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  $G^\infty(\cdot | \theta)$  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  $\theta$ , she considers the set  $\mathcal{P}_\theta$  defined in the introduction and repeated here for convenience:

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

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

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  $\pi$  of the indices  $1, 2, \dots$ , and any probability measure  $P$  on  $S^\infty$ , denote by  $\pi P$  the unique probability measure satisfying (for all rectangles)  $(\pi P)(A_1 \times A_2 \times \dots) = P(A_{\pi^{-1}(1)} \times A_{\pi^{-1}(2)} \times \dots)$ . Then it is easy to see that  $\mathcal{P}_\theta$  is symmetric, or "exchangeable," in the sense that

$$P \in \mathcal{P}_\theta \iff \pi P \in \mathcal{P}_\theta. \quad (2.2)$$

---

However, we deviate from existing literature in adopting also the following formal model of the sequence of experiments.

<sup>8</sup>For any metric space  $X$ , we endow  $X^\infty$  with the product metric and corresponding Borel  $\sigma$ -algebra. (Then, given that  $S$  is finite,  $S^\infty$  is separable compact metric, and hence Polish). We denote by  $\Delta(X)$  the set of Borel (countably additive) probability measures on  $X$ .

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

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  $\nu_\theta^\infty$  can also be expressed in the form: For every measurable  $B \subset S^\infty$ ,

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

Thus  $\nu_\theta^\infty$  is the capacity on measurable subsets of  $S^\infty$  induced by the correspondence  $G^\infty(\cdot | \theta)$  and the probability measure  $m_\theta^\infty$  on  $U^\infty$ , 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  $\{u^\infty \in U^\infty : G^\infty(u^\infty | \theta) \subset B\}$  in (2.3) is in general not measurable for every Borel measurable  $B$ . However, it is universally measurable, and moreover, each Borel measure  $m_\theta^\infty$  has a unique extension to a probability measure (also denoted  $m_\theta^\infty$ ) on the collection of all universally measurable subsets of  $S^\infty$ . 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  $\nu_\theta^\infty$ .*

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

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

If we define  $\nu_\theta$  on  $S$  as the lower envelope of this set, then

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

from which we can conclude that  $\nu_\theta$  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  $\nu_\theta$ , denoted  $\nu_\theta^*$ , by

$$\nu_\theta^*(A) = 1 - \nu_\theta(S \setminus A). \quad (2.5)$$

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

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

There is a relation between the belief function  $\nu_\theta$  on  $S$  and the belief function  $\nu_\theta^\infty$  on  $S^\infty$  that is suggested by our notation and that is important below. Specifically, one can view  $\nu_\theta^\infty$  as an "i.i.d. product" of  $\nu_\theta$  because they satisfy:<sup>9</sup> for every finite  $n$ , and for all subsets  $A_i$  of  $S_i = S$ ,  $i = 1, \dots, n$ ,

$$\nu_\theta^\infty(A_1 \times \dots \times A_n \times \prod_{i=n+1}^\infty S_i) = \prod_{i=1}^n \nu_\theta(A_i). \quad (2.6)$$

Given the central role played by i.i.d. probability measures in the classical CLT, it should not be surprising that a corresponding notion for belief functions is important for the CLT derived in the next section. We caution, however, that the parallel with the case of probability measures should not be carried too far. The relation (2.6) does not express stochastic independence of selection across markets as such independence is usually understood. Rather, as described when interpreting  $G^\infty(\cdot | \theta)$  defined in (1.1), which underlies (2.6), it reflects agnosticism about how selection is related across markets in that ignorance about selection in one experiment is "independent" of, or unrelated to, ignorance about selection in any other experiment.

We conclude this section with three remarks about the scope of our framework. First, we emphasize that  $\mathcal{P}_\theta$  is not a primitive—if one starts with an arbitrary set  $\mathcal{P}_\theta$  and defines the lower envelope by (2.3), then  $\nu_\theta^\infty$  will typically not be a belief function and the inference procedure that follows does not apply.<sup>10</sup> It is important to keep in mind that the primitive in our approach is the tuple  $(S^\infty, U^\infty, G^\infty, \Theta; m^\infty)$ , the elements of which are used to define  $\mathcal{P}_\theta$  by (2.1). The

---

<sup>9</sup>The following is readily derived from the Cartesian product definition of  $G^\infty(\cdot | \theta)$  and the i.i.d. nature of each measure  $m_\theta$ . See Epstein and Seo (2015) and the references therein for more on i.i.d. products of belief functions.

<sup>10</sup>Also demonstrated in Philippe, Debs and Jaffray (1999) is that the set  $\mathcal{P}_\theta$  defined by (2.1) coincides with the *core* of  $\nu_\theta^\infty$ ,  $\text{core}(\nu_\theta^\infty) = \{P \in \Delta(S^\infty) : P(\cdot) \geq \nu_\theta^\infty(\cdot)\}$ , which shows again that  $\mathcal{P}_\theta$  is special—it must be the core of a belief function.

analyst must determine if a tuple of this form captures the problem at hand, and if so, as is the case for the empirical studies cited in the introduction, then our procedure can be applied.

A related point of clarification is that our method provides robustness to complete agnosticism about how selection operates, but is presumably overly conservative if the analyst is confident in restricting the possible patterns of selection across experiments. Our view is that such "maximal robustness" is what has often been called for in the literature, as illustrated by the above quote from Ciliberto and Tamer; and given also the argument that if ignorance about selection in any single market is due to missing variables, whose identity is not even known by the analyst, then ignorance about the cross-sectional variation in selection follows logically. (It may be helpful to add a few words about the formal reasons that less than maximal robustness is excluded by our analysis. One way to restrict patterns of selection across experiments is to replace  $G^\infty(\cdot | \theta)$  in (2.1) by a correspondence  $H(\cdot | \theta) : U^\infty \rightsquigarrow S^\infty$  such that  $H(\cdot | \theta) \subset G^\infty(\cdot | \theta)$ . Denote by  $\mathcal{P}'_\theta$  the set of probability laws defined in this way. Then the lower envelope of  $\mathcal{P}'_\theta$  is a belief function, but not an i.i.d. product belief function, that is, it does not satisfy (2.6). Therefore, our CLT does not apply.)

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 i.i.d. probability measure (for each  $\theta$ ) 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, we can do better in this regard. As explained in the discussion of our CLT (Theorem 3.1), the assumption that each  $m_\theta^\infty$  is i.i.d. can be relaxed. Further and more fundamentally, our framework also permits the analyst to have an incomplete model of latent variables in that one can take each  $m_\theta$  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  $\nu_\theta^\infty$ . 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.

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

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

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

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

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  $\Lambda \in \mathbb{R}^{J \times J}$ ,  $\mathbf{N}_J(\cdot; \Lambda)$  denotes the  $J$ -dimensional normal cdf with zero mean and covariance matrix  $\Lambda$ —for any  $c = (c_1, \dots, c_J) \in \mathbb{R}^J$ ,  $\mathbf{N}_J(c; \Lambda)$  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,  $A_1, \dots, A_J$ , subsets of  $S$ , and for any  $\theta$ , let

$$\text{cov}_\theta(A_i, A_j) = \nu_\theta(A_i \cap A_j) - \nu_\theta(A_i)\nu_\theta(A_j), \quad (3.3)$$

$$\text{var}_\theta(A_i) = \nu_\theta(A_i)(1 - \nu_\theta(A_i)) = \text{cov}_\theta(A_i, A_i).$$

Denote by  $\Lambda_\theta$  the  $J \times J$  symmetric and positive semidefinite matrix  $(\text{cov}_\theta(A_i, A_j))$ .<sup>11</sup>

**Theorem 3.1.** *Suppose  $\theta_n \in \Theta$  and  $c_n \in \mathbb{R}^J$ ,  $n = 1, 2, \dots$ , and let  $\Lambda_{\theta_n} \rightarrow \Lambda \in \mathbb{R}^{J \cdot J}$  and  $c_n \rightarrow c \in \mathbb{R}^J$ . Then*

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

See Appendix A for a proof.<sup>12</sup>

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  $A_{I+i} = S \setminus A_i$  for each  $i = 1, \dots, I$ , 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

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

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  $[\nu_\theta(A_i), \nu_\theta^*(A_i)]$ , 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  $\nu_\theta(A_i)$  or larger than  $\nu_\theta^*(A_i)$ , and the distribution of such deviations according to  $\nu_\theta^\infty$  is approximately normal.

When each  $\nu_{\theta_n}$  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

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

---

<sup>11</sup>Positive semidefiniteness is proven in the theorem.

<sup>12</sup>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.

But it can be shown that for this event the minimum in (1.2) is achieved at an i.i.d. measure  $P_n^*$ .<sup>13</sup> 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  $G^\infty$  given in (1.1). Together they permit, for each  $\theta_n$ , transforming our assertion about belief functions into one about i.i.d. probability measures  $m_{\theta_n}^\infty$  as follows:

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

where for each  $j$ ,  $X_{ni}^j = I(G(u_i|\theta) \subset A)$ ,  $i = 1, \dots, n$ , 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  $P_{u^\infty}$  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  $G(\cdot|\theta_n)$ . 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  $\theta_n$ ) 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  $X_{ni} = (X_{ni}^1, \dots, X_{ni}^J)'$ ,  $i = 1, \dots, n$ , obey a suitable central limit theorem under  $m_{\theta_n}^\infty$ , such an extended result becomes available.<sup>14</sup>

---

<sup>13</sup> $P_n^*$  is the i.i.d. product of  $p_n^* \in \Delta(S)$  such that  $p_n^*(A_1) = \nu_{\theta_n}(A_1)$ . When a minimizer exists in (1.2) for an event, refer to it as a minimizing or worst-case measure for that event.

<sup>14</sup>For example, Jenish and Prucha's (2009) central limit theorem for arrays of random fields allows variables to have spatial correlations.

### 3.2. Confidence regions

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

$$\sigma_\theta \equiv \left( \sqrt{\text{var}_\theta(A_1)}, \dots, \sqrt{\text{var}_\theta(A_J)} \right). \quad (3.8)$$

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

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

where<sup>15</sup>

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

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  $T_{j,n}(\theta) \equiv (\nu_\theta(A_j) - \Psi_n(s^\infty)(A_j)) / \sqrt{\text{var}_\theta(A_j)/n}$ ,  $j = 1, \dots, J$ , where we take  $1/0 = \infty$ ,  $0/0 = 0$  and  $-1/0 = -\infty$ . Here,  $\text{var}_\theta(A_j)$  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  $\theta$ . If  $\nu_\theta(A_j) = 1$ , then  $\theta$  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 next.

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

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

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

Given  $\theta$ ,  $\mathcal{P}_\theta$  is the set of all probability laws consistent with the model and  $\nu_\theta^\infty$  gives its lower envelope. Therefore, the theorem establishes that if  $\theta$  is the ‘‘true value’’ of the parameter, then, in the limit for large samples,  $\mathcal{C}_n$  contains  $\theta$  with probability at least  $1 - \alpha$  according to every probability law that is consistent with

---

<sup>15</sup>The proof of the next theorem shows that  $c_\theta$  is well-defined. If  $\sigma_\theta = 0$ , then  $\mathbf{N}_J(0; \Lambda_\theta)$  refers to a degenerate distribution at the mean, which is 0, and thus  $\mathbf{N}_J(c\sigma_\theta; \Lambda_\theta) = 1$  for all  $c \geq 0$ , and  $c_\theta = 0$ .

the model and  $\theta$ . Moreover, (3.11) can also be stated as  $\liminf_{n \rightarrow \infty} \inf_{(\theta, P) \in \mathcal{F}} P(\theta \in \mathcal{C}_n) \geq 1 - \alpha$ , 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  $\theta$ .<sup>16</sup>

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 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  $\theta$  lies in  $[0, 1]$  and  $u = (u_1, u_2)$  is observed by players but not by the analyst. She views  $\theta$  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  $\theta$  and uniform on  $[0, 1]^2$ ; and  $G$  equal to the (pure strategy) Nash equilibrium correspondence given by

$$G(u_1, u_2 | \theta) = \begin{cases} \{0, 1\} & \text{if } 0 \leq u_1, u_2 \leq \theta^{1/2} \\ \{0\} & \text{otherwise.} \end{cases} \quad (4.1)$$

---

<sup>16</sup>Note that because  $\Lambda_\theta$  is positive semidefinite,  $\sigma_\theta = 0$  if and only if  $\Lambda_\theta = 0$ .

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  $\nu_\theta$ , where

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

Turn to inference about  $\theta$ . Suppose first that  $J = 1$  and  $A_1 = \{1\}$ . Then, for all  $\theta$ ,  $\nu_\theta(1) = 0$  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  $\theta$ .

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

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

where  $c_\theta = 0$  if  $\theta = 0$  or  $1$ , and otherwise  $c_\theta$  is the critical value for the standard normal variable and satisfies  $\mathbf{N}_1(c_\theta; 1) \geq 1 - \alpha$ .<sup>17</sup> Thus the interval constraint imposed by the LLN (see the appropriate form of (3.1)), whereby asymptotically the empirical frequency of  $s = 1$  is bounded above by  $\theta$ , is relaxed here to the degree expressed by  $c_\theta [\theta(1 - \theta)]^{1/2} / \sqrt{n}$ . 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.<sup>18</sup> Then  $s_i$  is a Bernoulli random variable with parameter  $\theta$  which is the largest (unconditional) probability consistent with the incomplete model. The MLE for  $\theta$  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  $\theta$ ,

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

<sup>17</sup>The reduction to a univariate distribution is a consequence of the fact that  $\text{var}_\theta(\{1\}) = 0$  for all  $\theta$ .

<sup>18</sup>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.

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  $\theta$ ) 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  $G(\cdot | \theta)$  is nonempty-valued for every parameter  $\theta$ . 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  $G(\cdot | \theta)$  can be defined accordingly and inference can proceed as described above.<sup>19</sup> However, the restriction to pure strategies is important. If we allowed mixed strategies, then the equilibrium correspondence  $G(\cdot | \theta)$  would map into subsets of the probability simplex  $\Delta(S)$  and  $\nu_\theta$  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  $\theta_1$  and  $\theta_2$  are interpreted as the minimal and maximal probabilities for

---

<sup>19</sup>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).

the outcome  $s = 1$ . For  $(U, m)$ , take any nonatomic probability space (with  $U$  Polish and  $m_\theta = m$  for all  $\theta$ ). Finally, define  $G(\cdot | \theta) : U \rightsquigarrow S$  by

$$G(u | \theta) = \begin{cases} \{1\} & \text{if } u \in U_{\theta_1} \\ \{0\} & \text{if } u \in U_{\theta_2} \\ \{1, 0\} & \text{otherwise,} \end{cases}$$

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  $\theta$  induces the belief function  $\nu_\theta$  on  $S$ , where  $\nu_\theta(1) = \theta_1$  and  $\nu_\theta(0) = 1 - \theta_2$ .

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

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

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  $\theta$ s for which  $\theta_1$  is "too large," but all sufficiently small  $\theta_1$  satisfy the first indicated inequality. This is because  $\theta_1$  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  $\theta_2$  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  $c_\theta$  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  $\theta_1$  can the empirical frequency be and still be seen as consistent with  $\theta_1$ ? 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  $\theta_1$ . 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  $\theta_2$ , 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  $\{s^\infty : \theta \in \mathcal{C}_n\} =$

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

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.<sup>20</sup> However, our confidence regions provide the desired coverage no matter what form that dependence might take.

Fix  $\alpha = .05$ . The critical value  $c_\theta$  depends on  $\theta$  according to (3.10). Though closed-forms are not available for all  $\theta$ , the following can be shown by elementary arguments applied to the bivariate normal distribution (Appendix C):

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

In addition,  $c_{(\theta_1,\theta_2)}$  is (strictly) increasing in  $\theta_1$  and decreasing in  $\theta_2$  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

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

---

<sup>20</sup>Dependence in a cross-sectional context is often modeled by various parametric copulas.

The critical value  $\tilde{c}_{n,\theta}$  is given by:<sup>21</sup>

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

where  $\{\kappa_n\}$  is a sequence of positive constants (tuning parameters) such that  $\kappa_n \rightarrow \infty$  and  $\kappa_n/\sqrt{n} \rightarrow 0$  and

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

$\mathcal{C}_n^{MI}$  is a confidence region based on moment inequalities.<sup>22</sup> The studentized moments  $\hat{l}_{j,n}$  are used to select those constraints to enter into calculation of the critical value. For example, when  $\hat{l}_{1,n}(\theta) \leq \kappa_n$ , the MI approach interprets this as indicating that the corresponding population constraint  $p(1) - \theta_1 \geq 0$  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,  $\hat{l}_{1n}(\theta) \leq \kappa_n$  and  $\hat{l}_{2n}(\theta) \leq \kappa_n$ . 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

<sup>21</sup>For comparison purposes, we use the critical value based on an asymptotic normal approximation instead of bootstrap approximations commonly used in the literature.

<sup>22</sup>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).

either side of the interval, which motivates the two-sided critical value.<sup>23</sup> 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  $\theta_1$  but not to  $\theta_2$ . The MI approach interprets this as the length of the interval being large relative to the sampling variation and  $p(1)$  being close to  $\theta_1$  but not to  $\theta_2$ . Hence, if the empirical frequency is convergent to  $p(1)$ , then asymptotically it may fall to the left of  $\theta_1$  but not to the right of  $\theta_2$ . 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  $\theta$  through the parameters in the bivariate normal distribution according to (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  $\theta_1$  or  $\theta_2$  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  $\theta_1$ ; hence, it suffices to consider  $\Psi_n(s^\infty)(1)$ 's variation around  $\theta_2$ , which motivates the one-sided critical value. Finally,  $c_{(\theta_1, \theta_2)} = 0$  when both  $\theta_1$  and  $\theta_2$  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  $\theta_1$  or on the right of  $\theta_2$ , 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.<sup>24</sup>

<sup>23</sup>This was pointed out previously by Imbens and Manski (2004) and Stoye (2009).

<sup>24</sup>The MATLAB code for our simulations is available at: <http://sites.google.com/site/seo8240>.

Simulations are based on the binary experiment, slightly specialized so that  $U = [0, 1]$ ,  $m$  is uniform on  $[0, 1]$ ,  $\Theta = \{(\theta_1, \theta_2) \in [0, 1]^2 : \theta_1 \leq \theta_2\}$ , 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  $\theta$  induces the belief function  $\nu_\theta$  on  $\{0, 1\}$  given by

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

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  $\text{prob}(v_i = 1) = \tau$  for some  $\tau \in [0, 1]$ .

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

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

The non-i.i.d. specification selects  $s_i = 1$  from  $\{0, 1\}$  when  $\Psi_n^G(u^\infty)$ , 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  $\Lambda_\theta$  is straightforward. To compute the critical value  $c_\theta$ , one needs to evaluate a CDF of a multivariate normal distribution with covariance matrix  $\Lambda_\theta$ . We do so by using simulated draws from the Geweke-Hajivassiliou-Keane (GHK) simulator and approximating the CDF  $\mathbf{N}_J(\cdot; \Lambda_\theta)$  by Monte Carlo integration.<sup>25</sup>

---

<sup>25</sup>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 critical value is then computed according to (3.10) replacing  $\mathbf{N}_J(\cdot; \Lambda_\theta)$  by its approximation. Throughout this section, we denote our confidence region by  $\mathcal{C}_n^{\text{EKS}}$ .

We compare the performance of the robust confidence region with that of existing methods. For each  $\theta$ , 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  $\Gamma : \mathbb{R}^J \times \mathbb{R}^{J \cdot J} \rightarrow \mathbb{R}$  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  $\kappa_n$ .

We consider two confidence regions that belong to this class. The first, denoted  $\mathcal{C}_n^{\text{MI}}$ , based on Galichon and Henry (2009) and Andrews and Soares (2010), uses the following criterion function and estimator of the asymptotic variance:

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

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  $\kappa_n$  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  $(\bar{M}_{n,\theta}, \hat{\Sigma}_{n,\theta})$ .<sup>26</sup>

<sup>26</sup>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  $\kappa_n$  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 $n$	Robust $\mathcal{C}_n^{\text{EKS}}$	MI		Robust $\mathcal{C}_n^{\text{EKS}}$	MI	
			$\mathcal{C}_n^{\text{MI}}$	$\mathcal{C}_n^{\text{AB}}$		$\mathcal{C}_n^{\text{MI}}$	$\mathcal{C}_n^{\text{AB}}$
			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	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,  $N_k = 2^{2^k} \in \{4, 16, 256, 65536\}$ .  $\mathcal{C}_n^{\text{MI}}$  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).

The second confidence region, denoted  $\mathcal{C}_n^{\text{AB}}$ , 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:

$$\Gamma(M, \Sigma) = \inf_{t \in \mathbb{R}_+^J} (M - t)'^{-1} (M - t)$$

$$\tilde{\Sigma}_{n,\theta} \equiv \hat{\Sigma}_{n,\theta} + \max(0.012 - \det(\hat{\Omega}_{n,\theta}), 0) \hat{D}_{n,\theta},$$

where  $\hat{D}_{n,\theta} = \text{diag}(\hat{\Sigma}_{n,\theta})$  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  $\kappa_n$ , which we set following their recommendations.

Table 5.1 reports the coverage probabilities of the three confidence regions  $\mathcal{C}_n^{\text{EKS}}$ ,  $\mathcal{C}_n^{\text{MI}}$ , and  $\mathcal{C}_n^{\text{AB}}$  under alternative values of  $(\theta_1, \theta_2)$  for a nominal level of 0.95. We set  $\tau = 0.5$  and 1 for the i.i.d. selection mechanism, and  $N_k =$

Table 5.2: Volume of Confidence Regions

Eq. Sel.	Sample Size $n$	Robust $\mathcal{C}_n^{\text{EKS}}$	MI		Robust $\mathcal{C}_n^{\text{EKS}}$	MI	
			$\mathcal{C}_n^{\text{MI}}$	$\mathcal{C}_n^{\text{AB}}$		$\mathcal{C}_n^{\text{MI}}$	$\mathcal{C}_n^{\text{AB}}$
			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).

$2^{2^k} \in \{4, 16, 256, 65536\}$  for the non-i.i.d. selection mechanism. We report simulation results based on samples of size  $n \in \{100, 256, 400, 10000, 65536\}$ .  $\mathcal{C}_n^{\text{MI}}$  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).<sup>27</sup>

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

<sup>27</sup>The moment selection tuning parameter  $\kappa_n$  and size correction factors  $(\eta_{1n}, \eta_{2n})$  are selected from Table I in Andrews and Barwick (2012) based on the smallest off-diagonal element of  $\hat{\Omega}_{n,\theta}$ .

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  $\mathcal{C}_n^{\text{MI}}$  and  $\mathcal{C}_n^{\text{AB}}$  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  $\mathcal{C}_n^{\text{MI}}$  is slightly below the nominal level in small samples.<sup>28</sup> 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  $\mathcal{C}_n^{\text{MI}}$  and  $\mathcal{C}_n^{\text{AB}}$  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

---

<sup>28</sup>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.

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:  $S, U, \Theta, m$  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  $G(\cdot | \theta, x) : U \rightsquigarrow S$ , for each  $\theta \in \Theta$  and  $x \in X$ . The latter and  $m$  induce, the belief function  $\nu_\theta(\cdot | x)$  on  $S$ , that is conditional on each  $\theta$  and  $x$ , and is given by

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

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  $q^\infty$  and  $m^\infty$ . The outcomes for the entire sequence of experiments are described by the correspondence,  $G^\infty(\cdot | \theta, x^\infty) : U^\infty \rightsquigarrow S^\infty$ , where, for each  $\theta$  and sequence of covariates  $x^\infty \equiv (x_1, \dots, x_i, \dots) \in X^\infty$ ,

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

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

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

Then,  $\nu_\theta^\infty(\cdot | x^\infty)$  gives the lower envelope of the set  $\mathcal{P}_{\theta, x^\infty}$ , paralleling (2.1), of all probability laws over  $S^\infty$  that are consistent with the given theory and  $\theta$  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  $A_1, \dots, A_J$ , subsets of  $S$ .<sup>29</sup> Define, for each  $\theta$  and  $x \in X$ ,

$$\text{cov}_\theta(A_i, A_j | x) = \nu_\theta(A_i \cap A_j | x) - \nu_\theta(A_i | x) \nu_\theta(A_j | x) \quad (6.1)$$

$$\text{var}_\theta(A_j | x) = \text{cov}_\theta(A_j, A_j | x). \quad (6.2)$$

---

<sup>29</sup>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  $x = x_k$ .

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

Define  $c_\theta = \min \{c \in \mathbb{R}_+ : \mathbf{N}_{|X|J}(c\sigma_\theta; \Lambda_\theta) \geq 1 - \alpha\}$ . Another way to express  $c_\theta$  is as follows. Let  $Z_\theta = (Z_{\theta,1}, \dots, Z_{\theta,|X|J})$  be multivariate normal with mean 0 and covariance  $\Lambda_\theta$ , and let  $W = \max_{k=1, \dots, |X|J} Z_{\theta,k} / \sigma_{\theta,k}$  with the conventions  $1/0 = \infty$ ,  $0/0 = 0$  and  $-1/0 = -\infty$ . Then  $c_\theta$  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  $\Pr(W \leq c_\theta) = 1 - \alpha$ .

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

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

Since  $g$  has the full support,  $\Psi_n$  is well-defined asymptotically. Define the statistic

$$T_n(\theta) = \max_{(x,j) \in X \times \{1, \dots, J\}} \left\{ \frac{\nu_\theta(A_j | x) - \Psi_n(s^\infty, x^\infty)(A_j | x)}{\sqrt{\text{var}_\theta(A_j | x) / n}} \right\}, \quad (6.3)$$

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

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

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

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

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

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

**Corollary 6.2.** *We have*

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

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  $\theta$ . A particular case of the conditional structure  $(U, G(\cdot | \theta), m_\theta)$  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  $G(\cdot | \theta) = \widehat{G}$  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  $\nu_\theta$  on  $S$  generated by any  $(U, G(\cdot | \theta), m_\theta)$  can also be generated by  $(\mathcal{K}(S), \widehat{G}, \widehat{m}_\theta)$  for some probability measure  $\widehat{m}_\theta$  on  $\mathcal{K}(S)$ ; and similarly for  $\nu_\theta^\infty$ . Because  $(\mathcal{K}(S), \widehat{G}, \widehat{m}_\theta)$  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  $m_\theta$  rather than  $\widehat{m}_\theta$ . Then, without loss of generality, suppose that  $\nu_\theta$  and  $\nu_\theta^\infty$  satisfy

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

and

$$\nu_\theta^\infty(B) = m_\theta^\infty(\{K_1 \times K_2 \times \dots \in (\mathcal{K}(S))^\infty : \prod_{i=1}^\infty K_i \subset B\}), \quad B \subset S^\infty.$$

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

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

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

$$\begin{aligned}
\text{cov}(X_{ni}^j, X_{ni}^l) &= E(X_{ni}^j X_{ni}^l) - E(X_{ni}^j) E(X_{ni}^l) \\
&= \int I(K_i \subset A_j) I(K_i \subset A_l) dm_{\theta_n}(K_i) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l) \\
&= \int I(K_i \subset A_j \cap A_l) dm_{\theta_n}(K_i) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l) \\
&= \nu_{\theta_n}(A_j \cap A_l) - \nu_{\theta_n}(A_j) \nu_{\theta_n}(A_l), \text{ and} \\
\text{var}(X_i^j) &= \text{cov}(X_i^j, X_i^j) = \nu_{\theta_n}(A_j) (1 - \nu_{\theta_n}(A_j)).
\end{aligned}$$

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

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

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  $\Lambda_{\theta_n}$ .

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

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

Hence,

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

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

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

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  $\tilde{Y}_n \equiv c_n + \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ni} \rightarrow_d Z$  where  $Z$  is  $J$ -dimensional multivariate normal with mean  $c$  and covariance matrix  $\Lambda$ . Apply the Cramér-Wold device: let  $a \in \mathbb{R}^J$  and show that  $a'\tilde{Y}_n \rightarrow_d a'Z$ . Note that  $\lim_{n \rightarrow \infty} \text{var} (a'\tilde{Y}_n) = \lim_{n \rightarrow \infty} a'\Lambda_{\theta_n}a = a'\Lambda a$ . If  $a'\Lambda a = 0$ , then  $a'\tilde{Y}_n \rightarrow_d c = a'Z$ . If  $a'\Lambda a > 0$ , we can apply a triangular CLT (White 2001, Theorem 5.11),<sup>30</sup> to prove that

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

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

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

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

The proof of (3.4) is completed by noting that

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

■

---

<sup>30</sup>The condition in the theorem that  $E|a'Y_{ni}|^{2+\delta}$  is bounded is satisfied here because  $Y_{ni}$  is bounded.

## B. Appendix: Proof of Theorem 3.2

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

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

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

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  $\theta$ .

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

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

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

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

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

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

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

Step 4:  $[\Lambda_n \rightarrow \Lambda \text{ and } c(\Lambda_n) \rightarrow c^*] \implies c^* \geq c(\Lambda)$ : By Step 3,  $\mathbf{N}_J(c^*\sigma(\Lambda); \Lambda) \geq 1 - \alpha$ . 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  $\Lambda_\nu$  be the covariance matrix as defined in (3.3). Then  $\nu \mapsto \Lambda_\nu$  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

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

Since  $\nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n)$  is bounded, by taking a subsequence if necessary, we can assume that  $\liminf_{n \rightarrow \infty} \nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n) = \lim_{n \rightarrow \infty} \nu_{\theta_n}^\infty(\theta_n \in \mathcal{C}_n)$ . Moreover, by Step 5, and by taking a further subsequence if necessary, we can assume that  $\Lambda_{\theta_n} \rightarrow \Lambda \in \mathbb{R}^{J \times J}$ . 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  $c(\Lambda_{\theta_n}) \rightarrow c^*$ . Thus, the CLT (Theorem 3.1) implies that

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

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

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

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

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

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

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

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

$$\begin{aligned} \limsup_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\{s^\infty : \theta \in \mathcal{C}_n\}) &\leq \limsup_{n \rightarrow \infty} \nu_{\bar{\theta}}^\infty(\{s^\infty : \bar{\theta} \in \mathcal{C}_n\}) = 1 - \alpha \\ &\leq \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\{s^\infty : \theta \in \mathcal{C}_n\}) \end{aligned}$$

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  $\lambda$  in  $[-1, 0]$ , define  $\Lambda(\lambda) = \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}$ , and  $c(\lambda)$  by

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

Then  $\lambda \mapsto c(\lambda)$  is (strictly) decreasing on  $[-1, 0]$  because  $\mathbf{N}_2(\cdot; \Lambda(\lambda)) \nearrow^\lambda$ .<sup>31</sup> It follows that  $c(\lambda) \searrow^\lambda$ . In addition,  $\lambda \mapsto c(\lambda)$  is continuous on  $[-1, 0]$ .<sup>32</sup>

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

For  $\theta$ s satisfying  $0 < \theta_1 < \theta_2 < 1$ —one has  $\sigma_\theta \gg 0$  and

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

where

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

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

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

and

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

<sup>31</sup>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.

<sup>32</sup>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.

Finally, note that: (1)  $c(-1)$  is defined by  $\mathbf{N}_2((c(-1), c(-1)); \Lambda(-1)) = 1 - \alpha$ . 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  $\mathbf{N}_2((c(0), c(0)); \Lambda(0)) = .95$  or  $\mathbf{N}_1(c(0); 1) = [.95]^{1/2} \simeq .9747$ , which gives  $c(0) = 1.955$ . ■

## 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  $S^\infty = S_1 \times S_2 \times \dots$ , where  $S_i = S$  for all  $i$ . For any  $I \subset \{1, 2, \dots\}$ , denote by  $\Sigma_I$  the  $\sigma$ -algebra generated by (Borel measurable) cylinders of the form  $\prod_{i \in I} A_i \times \prod_{i \notin I} S_i$ , 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.**  $\nu_\theta^\infty \left( \bigcap_{k=1}^K B_k \mid x^\infty \right) = \prod_{k=1}^K \nu_\theta^\infty (B_k \mid x^\infty)$  if  $B_1, \dots, B_K$  are pairwise orthogonal.

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

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

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

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

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

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

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

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

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n) \geq \int_{X_{\text{inf}}^\infty} \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n | x^\infty) dq^\infty(x^\infty) \geq 1 - \alpha.$$

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

$$\begin{aligned} \limsup_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n) &\leq \limsup_{n \rightarrow \infty} \nu_{\bar{\theta}}^\infty(\bar{\theta} \in \mathcal{C}_n) \\ &= \limsup_{n \rightarrow \infty} \int_{X_{\text{inf}}^\infty} \nu_{\bar{\theta}}^\infty(\bar{\theta} \in \mathcal{C}_n | x^\infty) dq^\infty(x^\infty) \\ &\leq \int_{X_{\text{inf}}^\infty} \limsup_{n \rightarrow \infty} \nu_{\bar{\theta}}^\infty(\bar{\theta} \in \mathcal{C}_n | x^\infty) dq^\infty(x^\infty) = 1 - \alpha \\ &\leq \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \nu_\theta^\infty(\theta \in \mathcal{C}_n). \quad \blacksquare \end{aligned}$$

## E. Appendix: Latent variables robustified

Currently, we define models via primitives  $(S, U, G, \Theta, m)$ , including, in particular, the probability measures  $m_\theta$  on  $U$  for every  $\theta$ . 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  $m_\theta$  is a belief function rather than a measure. Also in this case we obtain belief functions  $\nu_\theta$  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  $\Theta$  be as before. Instead of adopting  $m$  as another primitive, we derive it from more basic primitives. Thus let the tuple  $(\widehat{U}, \Gamma, \widehat{m})$  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  $\Gamma(\cdot | \theta) : \widehat{U} \rightsquigarrow U$  is weakly measurable. (The assumption that the same parameters  $\theta$  enter here is without loss of generality since one could expand the parameter space  $\Theta$  as needed.) Thus probabilistic knowledge is assumed on  $\widehat{U}$  which, via the correspondence  $\Gamma$ , provides only coarse information about the latent variables  $u \in U$ . Paralleling (2.4), the elements  $\widehat{U}, \Gamma$  and  $\widehat{m}$  induce (for each  $\theta$ ) a belief function on  $U$ , denoted by  $m_\theta$  and given by

$$m_\theta(Y) = \widehat{m}_\theta(\{\widehat{u} : \Gamma(\widehat{u} | \theta) \subset Y\}), \quad Y \subset U. \quad (\text{E.1})$$

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

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

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

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

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

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

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

Because it depends only on having a belief function  $\nu_\theta$  on  $S$  for each parameter  $\theta$ , 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  $G(u | \theta)$ , there is also the coarse information about  $u$  due to  $\Gamma(\cdot | \theta)$  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  $(S, \widehat{U}, \widehat{G}, \Theta, \widehat{m})$  as the model. However, in applications the identity of  $\widehat{U}$ ,  $\Gamma$  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  $(S, U, G, \Theta, m)$  where each  $m_\theta$  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  $m_\theta$  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  $\nu_\theta$  and the critical value  $c_\theta$ . For simple examples, one may compute  $\nu_\theta$  analytically. In general, it can be computed using a simulation procedure. Once  $\nu_\theta$  is obtained, the critical value  $c_\theta$  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  $i = 1, 2, \dots$ . For each  $i$ , we let  $s_i = (s_{i1}, \dots, s_{iK}) \in \{0, 1\}^K$  denote the vector of entry decisions made by the firms. For firm  $k$  in market  $i$ , CT consider the following profit function

specification:

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

where  $v_i$  is a vector of market characteristics,  $z_i = (z_{i1}, \dots, z_{iK})$  is a matrix of firm characteristics that enter the profits of all firms in the market, while  $w_i = (w_{i1}, \dots, w_{iK})$  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  $u_i = (u_{i1}, \dots, u_{iK})$  follow a multivariate normal distribution  $N(0, \Sigma)$  and vary across markets in an i.i.d. way.<sup>33</sup> The structural parameter  $\theta$  includes  $\Sigma$  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

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

Suppose that a sample  $\{(s_i, x_i), i = 1, \dots, n\}$  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

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

Hence, a natural way to approximate  $\nu_\theta(A|x)$  for any  $A \subset S$  is to simulate  $u$  from the parametric distribution  $m_\theta$  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  $\Sigma$ , draw random vectors  $u^r = (u_1^r, \dots, u_K^r), r = 1, \dots, R$ , from  $N(0, \Sigma)$ .

<sup>33</sup>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  $m_\theta^\infty$ .

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  $\theta$ .

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

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

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  $c_\theta$ , which can be done by feeding  $\Lambda_\theta$  into a commonly-used simulator for multivariate normal random vectors.

### Simulation procedure 2

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

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

Step 2 Decompose  $\Lambda_\theta$  as  $LDL'$  for a lower triangular matrix  $L$  and a diagonal matrix  $D$ .

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

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

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

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  $\Lambda_\theta$  may only be positive semidefinite.

Simulation procedure 2 yields a critical value  $c_\theta$ . 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  $\{(u_1^r, \dots, u_K^r)\}_{r=1}^R$  and  $\{w^r\}_{r=1}^R$  only once and use them repeatedly across all values of  $\theta$ .

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

- [6] A. Beresteanu and F. Molinari, Asymptotic properties for a class of partially identified models, *Econometrica* 76 (2008), 763-84.
- [7] S. Berry, Estimation of a model of entry in the airline industry, *Econometrica* 60 (1992), 889-917.
- [8] T.F. Bresnahan and P.C. Reiss, Empirical models of discrete games, *J. Econometrics* 48 (1991), 57-81.
- [9] T.F. Bresnahan and P.C. Reiss, Entry in monopoly markets, *Rev. Econ. Stud.* 57 (1990), 531-53.
- [10] F. A. Bugni, Bootstrap inference in partially identified models defined by moment inequalities: coverage of the identified set, *Econometrica* 78 (2010), 735-753.
- [11] I.A. Canay, EL inference for partially identified models: large deviations optimality and bootstrap validity, *J. Econometrics* 156 (2010), 408-425.
- [12] A. Chesher and A. Rosen, Generalized instrumental variable models, CEMMAP Working Paper, 2014.
- [13] V. Chernozhukov, H. Hong and E. Tamer, Inference on parameter sets in econometric models, MIT, 2004.
- [14] V. Chernozhukov, H. Hong and E. Tamer, Estimation and confidence regions for parameter sets in econometric models, *Econometrica* 75 (2007), 1243-84.
- [15] F. Ciliberto and E. Tamer, Market structure and multiple equilibria in airline markets, *Econometrica* 77 (2009), 1791-1828.
- [16] A. Dasgupta, *Asymptotic Theory of Statistics and Probability*, Springer, 2008.
- [17] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Statist.* 38 (1967), 325-39.
- [18] A. P. Dempster, A generalization of Bayesian inference, *J. Royal Stat. Soc. Series B* 30 (1968), 205-32.
- [19] A. Eizenberg, Upstream innovation and product variety in the U.S. home PC market, *Rev. Econ. Stud.* 81 (2014), 1003-45.

- [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] P. Jia, What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry, *Econometrica* 76 (2008), 1263-1316.
- [30] 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.
- [31] B. Jovanovic, Observable Implications of models with multiple equilibria, *Econometrica* 57 (1989), 1431-37.
- [32] K. Kawai and Y. Watanabe, Inferring strategic voting, *Amer. Econ. Rev.* 103 (2013), 624-62.

- [33] C.F. Manski, *Partial Identification of Probability Distributions*, Springer, 2003.
- [34] F. Maccheroni and M. Marinacci, A strong law of large numbers for capacities, *Ann. Prob.* 33 (2005), 1171-1178.
- [35] M. Marinacci, Limit laws for non-additive probabilities and their frequentist interpretation, *J. Econ. Theory* 84 (1999), 145-95.
- [36] K. Menzel, Robust decisions for incomplete structural models of social interactions, NYU, 2011.
- [37] Y. Miyauchi, Structural estimation of a pairwise stable network with non-negative externality, MIT, 2014.
- [38] I. Molchanov, *Theory of Random Sets*, Springer, 2005.
- [39] A. Muller and M. Scarsini, Some remarks on the supermodular order, *J. Multivariate Anal.* 73 (2000), 107-19.
- [40] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, 1994.
- [41] R. Phelps, *Lectures on Choquet's Theorem*, Springer, 2001.
- [42] F. Philippe, G. Debs and J.Y. Jaffray, Decision making with monotone lower probabilities of infinite order, *Math. Oper. Res.* 24 (1999), 767-84.
- [43] G. Shafer, *A Mathematical Theory of Evidence*, Princeton U. Press, Princeton, 1976.
- [44] G. Shafer, Belief functions and parametric models (with commentary), *J. Royal Stat. Soc. Series B* 44 (1982), 322-52.
- [45] 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.
- [46] J. Stoye, More on confidence regions for partially identified parameters, *Econometrica* 77 (2009), 1299-1315.

- [47] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010), 167–95.
- [48] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003), 147-65.
- [49] K. Uetake and Y. Wanatabe, Entry by merger: estimates from a two-sided matching model with externalities, 2012.
- [50] A. Wald, Remarks on the estimation of unknown parameters in incomplete systems of equations, pp. 305-310 in *Statistical Inference in Dynamic Economic Models*, ed. T.C. Koopmans. John Wiley, New York, 1950.
- [51] E. Weese, Political mergers as coalition formation: an analysis of the Heisei municipal amalgamations, *Quant. Econ.* 6 (2015), 257–307.
- [52] H. White, *Asymptotic Theory for Econometricians*, Academic Press, 2001.