CIRJE-F-820

Set Inference in Latent Variables Models

Ismael Mourifié
Université de Montréal

Marc Henry
Université de Montréal

October 2011

SET INFERENCE IN LATENT VARIABLES MODELS

ISMAEL MOURIFIÉ AND MARC HENRY

ABSTRACT. We propose a methodology for constructing valid confidence regions in incomplete models with latent variables satisfying moment equality restrictions. These include moment equality and inequality models with latent variables. The confidence regions are obtained by inverting tests based on the characterization of the identified set derived in Ekeland, Galichon, and Henry (2010). A valid bootstrap approximation of the distribution of the test statistic is derived under mild conditions and the confidence regions are shown to have correct asymptotic size.

Keywords: incomplete models, latent variables, partial identification, confidence regions.

JEL subject classification: C13, C72

INTRODUCTION

In areas of economic investigation with structural data insufficiencies or incompletely specified economic mechanisms, the hypothesized structure fails to identify a unique possible data generating mechanism for the data that is actually observed. In such cases, many traditional estimation and testing techniques become inapplicable and a framework for inference in incomplete models is developing, with an initial focus on estimation of the set of structural parameters compatible with true data distribution (hereafter *identified set*). A question of particular relevance in applied work is how to construct valid confidence regions for the identified set. Formal methodological proposals abound since the seminal work of Chernozhukov, Hong, and Tamer (2007), but they concentrate on models defined by moment inequality restrictions (or criterion minimization) on the observables.

Most structural economic models, however, involve latent variables and take the form $f(Y, U; \theta) \leq 0$, with f a function of an observable vector of variables Y, a latent vector of variables U and an unknown vector of deterministic parameters θ (or equivalently $Y \in G(U; \theta)$, where G is a many-to-many mapping, or *correspondence*). Inference on such structures was considered by Andrews, Berry, and Jia (2003), Ciliberto and Tamer (2009) and Bajari, Hong, and Ryan (2010) in the special case, where $f(Y, U; \theta) \leq 0$ defines the set of equilibria of a game, and more generally by Galichon and Henry (2009) and Henry, Méango, and Queyranne (2010), but always with a parametric restriction on the distribution of latent variables. Matzkin (1994) (and references therein) considers the case, where the implicit function theorem can be applied and Y is a function of U and θ, but does not consider the case of interest here, where the resulting relation between Y and U is many-to-many. Finally, Chesher (2010) considers set inference in the special case of single equation instrumental variable models for discrete outcomes, and the frameworks of Chernozhukov, Lee, and Rosen (2009), Menzel (2009) and Andrews and Shi (2010) encompass cases where the latent variable model can be transformed into intersection bounds or a continuum of moment inequality restrictions on the observables.

Here we consider the general case, where only moment restrictions are entertained for the latent variables, and thereby include in particular all models defined by moment equalities and inequalities with latent variables and models defined by the equilibrium correspondence of a games, without parametric assumptions on player types or other unobserved heterogeneity. The confidence regions we propose are based on the characterization of the identified set in general incomplete models with latent variables given in Ekeland, Galichon, and Henry (2010). The confidence region is obtained by inverting a test of the null hypothesis $H_0(\theta)$ characterizing the identified set Θ_I in the sense that $\Theta_I = \{\theta : H_0(\theta) \text{ holds}\}$. The limit distribution of the test statistic is achieved with the construction of a local empirical process. In that sense, the method of proof is related to that of Chernozhukov, Hong, and Tamer (2007) and particularly Galichon and Henry (2009). However, replacing parametric restrictions on latent variables by moment equality restrictions requires the construction of a completely different local empirical process and also raises new issues in showing weak convergence, as the test statistic involves a supremum over a random class of functions.

The limiting distribution obtained is not distribution free, but a version of the test statistic, where the empirical process is replaced by a bootstrapped version is shown to provide a valid approximation, and the confidence region is shown to have asymptotic correct size (in a uniform sense). The procedure is explained and illustrated on a revealed preference example.

The rest of the paper is organized as follows. The next section describes the econometric framework and defines the identified set. The following section explains the construction and shows validity of the confidence region. The last section concludes and proofs are collected in the appendix.

1. Econometric framework and identified set

We consider the problem of inference on the structural parameters of an economic model, when the latter are (possibly) only partially identified. The economic structure is defined as in Jovanovic (1989), which generalizes Koopmans and Reiersol (1950). Variables under consideration are divided into two groups. Latent variables U capture unobserved heterogeneity in the model. They are typically not observed by the analyst, but some of their components may be observed by the economic actors. Observable variables Y include outcome variables and other observable heterogeneity. They are observed by the analyst and the economic actors. We call *observable distribution P* the true probability distribution generating the observable variables. The econometric structure under consideration is given by a binary relation between observable and latent variables, i.e. a subset of $\mathcal{Y} \times \mathcal{U}$, which can be written without loss of generality as a many-to-many mapping (or *correspondence*) G from $\mathcal{U}$ to $\mathcal{Y}$.

Assumption 1 (Econometric specification)**.** *Observable variables Y, with realizations $y \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$ and latent variables U, with realizations $u \in \mathcal{U} \subseteq \mathbb{R}^{d_u}$, are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and satisfy the relation: $Y \in G(U) \subseteq \mathcal{Y}$ almost surely.*

Example 1 (Games)**.** *A family of examples of our framework arises with parametric games. Let N players with observable characteristics $X = (X_1, \ldots, X_N)$ and unobservable characteristics $U = (U_1, \ldots, U_N)$ have strategies $Z = (Z_1, \ldots, Z_N)$ and payoffs parameterized by X, U, Z and θ. For a given choice of equilibrium concept in pure strategies, call $\mathcal{C}(X, U, \theta)$ the equilibrium correspondence, i.e. the set of pure strategy equilibrium profiles. Then the empirical content of the game is characterized by $Z \in \mathcal{C}(X, U, \theta)$, which can be equivalently rewritten $Y \in G(U; \theta)$ with $Y = (Z, X)$.*

We assume a parametric structure for the model linking unobserved heterogeneity variables to observable ones and a set of moment equality restrictions for unobserved heterogeneity.

Assumption 2 (Correspondence)**.** *The correspondence G is known by the analyst up to a finite dimensional vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$. It is denoted $G(\cdot; \theta)$. For all $\theta \in \Theta$, $G(\cdot; \theta)$ is measurable (i.e. the set $\{G(U; \theta) \cap A \neq \varnothing\}$ is measurable for each open subset A of $\mathcal{Y}$) and has non empty values.*

Note that the measurability and closed values assumptions are very mild conditions. The assumption that the correspondence is non-empty, however, may be restrictive. In the revealed preferences example, we require that the demand correspondence be non empty. In the games example, we require existence of equilibrium.

Assumption 3 (Latent variables). *The unobservable variables U is assumed to satisfy $\mathbb{E}m(U;\theta) = 0$, $\theta \in \Theta$ and $m(u;\theta) \in \mathbb{R}^{d_m}$. The same notation is used for the parameters of ν and G to highlight the fact that they may have components in common.*

Example 2 (Moment inequalities). *Moment inequality models are a special case of assumptions 1, 2 and 3. Call G the correspondence defined for all $u \in \mathcal{U}$ by $G(u;\theta) = \{y \in \mathcal{Y} : \phi(y;\theta) \leq u\}$. Note that G satisfies assumption 2. Suppose further that unobserved heterogeneity U satisfies assumption 3, i.e. $\mathbb{E}m(U;\theta) = 0$ where m is the identity. Then, under assumption 1, the model satisfies $\mathbb{E}\phi(Y;\theta) \leq 0$. Extension to conditional moment inequality models is also possible with an extension of lemma 2.*

Example 3 (Moment equalities and inequalities with latent variables). *Moment equality models with latent variables is also a special case of assumptions 1, 2 and 3. Indeed, if the relevant model takes the form $\mathbb{E}\phi(Y,U_1;\theta) = 0$, set $\phi(Y,U_1;\theta) = U_2$ with $\mathbb{E}U_2 = 0$ (U_1 and U_2 are both unobservable random vectors). Call $u = (u_1, u_2)$. Define the correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ such that $\phi(y,u_1;\theta) = u_2$ is equivalent to $y \in G(u;\theta)$ (without loss of generality). Finally define $m(u) = u_2$ (the function returns the last components of the vector u). We see therefore that the moment equality model with latent variables satisfies assumptions 1 and 3. An easy sufficient condition for assumption 2 is continuity of ϕ. Note that the same construction applies for moment inequality models with latent variables, setting $\phi(y,u_1;\theta) \leq u_2$ instead.*

The pair of random vectors (Y, U) involved in the model is generated by a probability distribution, that we denote π. Since the vector U is unobservable, the probability distribution π is not directly identifiable from the data. However, the econometric model imposes restrictions on π. The distribution of its component Y is the observable distribution P. The distribution of its component U satisfies assumption 3. Finally, the joint distribution is further restricted by the fact that it gives mass 0 to the event that the relation $Y \in G(U;\theta)$ is violated. For any given value of the structural parameter vector θ, a joint distribution satisfying all these restrictions may or may not exist. If it does, it is generally non unique. The identified set Θ_I is the collection of values of the structural parameter vector θ for which such a joint probability distribution does indeed exist.

- If $\Theta_I = \varnothing$, the model is rejected.

- If Θ_I is a singleton, the parameter vector θ is point identified.
- Otherwise, the parameter θ is set identified.

The set Θ_I, first formalized in this way in Galichon and Henry (2008) is sometimes called "sharp identification region" to emphasize the fact that it exhausts all the information on the parameter available in the model. No value $\theta \in \Theta_I$ could be rejected on the basis of the knowledge of the model and the observable distribution P only. Take a parameter value $\theta \in \Theta$. It belongs to the identified set Θ_I if and only if there exists a joint distribution satisfying the required restrictions, in other words, if and only if there exists a random vector $\tilde{U}$ satisfying assumption 3, such that $Y \in G(\tilde{U}; \theta)$ with probability 1. Hence, denoting by $X \sim \mu$ the statement "the random vector X has probability distribution μ," we can characterize the identified set in the following way, which we take as our formal definition.

Definition 1 (Identified set).

$$\Theta_I = \left\{ \theta \in \Theta \mid \exists \tilde{Y} \sim P, \ \tilde{U} \text{ satisfying assumption 3}: \ \mathbb{P}(\tilde{Y} \notin G(\tilde{U}; \theta)) = 0 \right\}$$

The identified set is therefore the set of parameter values such that the minimum over all joint distributions for $(\tilde{Y}, \tilde{U})$ satisfying 3 of the quantity $\mathbb{P}(\tilde{Y} \notin G(\tilde{U}; \theta))$ is zero. This optimization problem is shown in Ekeland, Galichon, and Henry (2010) to be equal to its dual under the following regularity conditions for each θ.

Assumption 4 (Duality conditions). *(Uniform integrability) The family $\{\|m(U; \theta)\| : U \text{ satisfies assumption 3}\}$ is uniformly integrable. (Tightness) For each $K \geq 0$, $\{u \in \mathcal{U} : \|m(u; \theta)\| \leq K\}$ is included in a compact set (in all that follows, we assume the simple sufficient condition that m is continuous and $\mathcal{U}$ compact). (Closed graph) The graph $\{(y, u) \in \mathcal{Y} \times \mathcal{U} : y \in G(u; \theta)\}$ of the correspondence G is closed.*

Ekeland, Galichon, and Henry (2010) show that the characterization of parameter values θ in the identified set Θ_I is equivalent to the dual version

$$H_0(\theta): \ \sup_{\lambda \in \mathbb{R}^{dm}} \mathbb{E} f_\lambda(Y; \theta) \leq 0,$$

where $f_\lambda(y; \theta) = \sup_{u \in \mathcal{U}} [1_{\{y \notin G(u; \theta)\}} - \lambda' m(u)]$ with 1_A denoting the indicator function of a set A.

Proposition 1 (Dual characterization). *Under assumption 4, the identified set Θ_I of definition 1 is equal to the set of $\theta \in \Theta$ such that $H_0(\theta)$ holds.*

The dual characterization, proved in Ekeland, Galichon, and Henry (2010) will be the basis for a test of assumption $H_0(\theta)$, which we invert to obtain confidence regions for the identified set.

2. Confidence region for the identified set

2.1. **Inference strategy.** We consider now the construction of valid confidence regions for partially identified models, based on a sample of realizations of the observable variables.

Assumption 5 (Sampling). *Let $(Y_1, \ldots, Y_n)$ be a sample of independent and identically distributed random vectors with distribution P and let $P_n = \sum_{j=1}^n \delta_{Y_j}$ be the empirical distribution associated with the sample.*

We propose to construct confidence regions for each element of the identified set Θ_I of definition 1.

Definition 2 (Confidence region). *A valid α-confidence region for the identified set Θ_I is a sequence of random regions Θ_n^α satisfying for each θ, $\mathbb{P}(\theta \in \Theta_n^\alpha) \to 1 - \alpha$.*

The confidence regions derived here will be obtained by inverting tests of $H_0(\theta)$. For simplicity, we drop θ from the notation. The test statistic will be based on an empirical counterpart of H_0, i.e. $\breve{T}_n := \sqrt{n} \sup_{\lambda \in \mathbb{R}^{d_m}} \mathbb{E}_n f_\lambda(Y)$, where $\mathbb{E}_n$ is the expectation relative to the empirical distribution P_n. Calling $\mathbb{G}_n := \sqrt{n}(\mathbb{E}_n - \mathbb{E})$ the empirical process relative to the sample, we can write: $\breve{T}_n = \sup_{\lambda \in \mathbb{R}^{d_m}} [\mathbb{G}_n f_\lambda(Y) + \sqrt{n}\mathbb{E}f_\lambda(Y)]$. Heuristically, under H_0, which rules out $\mathbb{E}f_\lambda(Y) > 0$, the second term in the brackets will either be equal to 0 if $\mathbb{E}f_\lambda(Y) = 0$ or diverge to $-\infty$ if $\mathbb{E}f_\lambda(Y) < 0$. Hence, we would expect $\breve{T}_n$ to converge weakly to the supremum of a Gaussian process over the restricted set Λ_0 of λ such that $\mathbb{E}f_\lambda(Y) = 0$. Similar decompositions arise in Chernozhukov, Hong, and Tamer (2007), Galichon and Henry (2009) and Andrews and Soares (2010). However, the class of functions f_λ, $\lambda \in \mathbb{R}^{d_m}$ may not be P-Donsker, so that the weak convergence result would fail to hold. Instead, we consider the test statistic

$$T_n := \sup_{\lambda \in \Lambda(h_n)} \sqrt{n}\mathbb{E}_n f_\lambda(Y),$$

where $h_n \downarrow 0$ and for $l \geq 0$, $\Lambda(l) = \{\lambda \in \mathbb{R}^{d_m} : \mathbb{E}f_\lambda(Y) \geq -l\}$ and show that the class of functions $\mathcal{F}^l := \{f_\lambda(\cdot) : \lambda \in \Lambda(l)\}$ is a P-Donsker class for $l \geq 0$ (lemma 2 in the appendix). The idea of replacing $\breve{T}_n$ by T_n to obtain a Donsker class is similar to the strategy employed in Bugni (2010) to apply the bootstrap in moment inequality models and it allows us to show the desired weak convergence result of the test statistic T_n to the supremum of a Gaussian process $\mathbb{G}_P$ over the class

of functions $\mathcal{F}^0 := \{f_\lambda(\cdot) : \lambda \in \Lambda_0\}$ (proposition 2 in the appendix). However, basing a test of H_0 on this convergence result is infeasible for two reasons.

First: The test statistic T_n is infeasible, since the definition of $\Lambda(h_n)$ involves the population expectation $\mathbb{E}f_\lambda(Y)$. This leads to replacing T_n by $\tilde{T}_n := \sup_{\lambda \in \Lambda_n(h_n)} \sqrt{n}\mathbb{E}_n f_\lambda(Y)$, where $\Lambda_n(h_n) = \{\lambda \in \mathbb{R}^{d_m} : \mathbb{E}_n f_\lambda(Y) \geq -h_n\}$ and showing that $\tilde{T}_n$ has the same limit as T_n. The latter is more involved since the class of functions $\mathcal{F}_n^{h_n} := \{f_\lambda(\cdot) : \lambda \in \Lambda_n(h_n)\}$ is now random.

Second: The limit $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$ is not distribution free, as it depends on the population distribution P of Y through the limiting process $\mathbb{G}_P$ and the definition of Λ_0. This leads to considering a bootstrapped version of the empirical process, $\mathbb{G}_n^* := \sqrt{n}(\mathbb{E}_n^* - \mathbb{E}_n)$, where $\mathbb{E}_n^*$ is the expectation relative to the empirical distribution of a bootstrapped sample, and showing that $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n^* f_\lambda(Y)$ has the same limit as T_n conditionally almost surely.

The next section summarizes the proposed inference procedure and the theoretical results on its validity.

2.2. Description and validity of the procedure.

2.2.1. *Summary of the procedure:* The proposed confidence regions for the identified set Θ_I are constructed as follows.

1. For a given value of the parameter vector θ and a sequence $h_n \downarrow 0$ (e.g. $h_n = \ln n$) compute the test statistic
$$\tilde{T}_n(\theta) = \sup_{\lambda \in \Lambda_n(h_n)} (1/\sqrt{n}) \sum_{j=1}^n f_\lambda(Y_j),$$
with $\Lambda_n(l) = \{\lambda \in \mathbb{R}^{d_m} : \sum_{j=1}^n f_\lambda(Y_j) \geq -nl\}$ and $f_\lambda(y) = \inf_{u \in \mathcal{U}}[1_{\{y \notin G(u;\theta)\}} - \lambda' m(u)]$.

2. Compute the $1 - \alpha$ quantile c_α^* of the distribution of
$$T_n^*(\theta) := \sup_{\lambda \in \Lambda_n(h_n)} (1/\sqrt{n}) \sum_{j=1}^n [f_\lambda(Y_j^*) - f_\lambda(Y_j)],$$
where $(Y_1^*, \ldots, Y_n^*)$ is a bootstrapped sampled.

3. Include θ in Θ_{CR} if and only if $\tilde{T}_n(\theta) \leq c_\alpha^*$.

Steps 1 and 2 are saddle point optimizations. As is customary in inference for partially identified models, search in the parameter space is the most computationally costly step in the construction of the confidence region.

2.2.2. *Validity of the confidence regions:* To show that the confidence region Θ_{CR} has the correct asymptotic size, we need further regularity assumptions on the correspondence G, the moment functions m and a rate assumption for the approximation of Λ_0 by $\Lambda_n(h_n)$.

Assumption 6 (LIL). h_n *satisfies* $h_n \ln \ln n + h_n^{-1}\sqrt{\ln \ln n/n} \to 0$ *as* $n \to \infty$.

Assumption 6 ensures that the terms $\sqrt{n}\mathbb{E}f_\lambda(Y)$ (which are negative under H_0) are dominated by the oscillations of the empirical process $\mathbb{G}_n f_\lambda(Y)$ on the class $\mathcal{F}_n^{h_n} := \{f_\lambda : \lambda \in \Lambda_n(h_n)\}$, so that $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{E}_n f_\lambda(Y)$ is indeed close to $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)$. The law of iterated logarithm for the oscillations of the empirical process was first invoked in a related context in Hansen (2005) and appears also in Galichon and Henry (2009) and Andrews and Soares (2010).

Assumption 7 (Regularity). *(i) The domain of $m(U;\theta)$ contains a closed ball $B(0,\eta)$ in $\mathbb{R}^{d_m}$ for some $\eta > 0$. (ii) The support function $\lambda \mapsto \sup\{\lambda'z : z \in m \circ G^{-1}(y;\theta)\}$ of $m \circ G^{-1}(y;\theta)$ is continuous for almost all y, where $m \circ G^{-1}(y;\theta) = \{z \in \mathbb{R}^{d_m} : \exists u \in \mathcal{U},\ z = m(u;\theta)\ and\ y \in G(u;\theta)\}$.*

Assumption 7(i) can always be achieved by renormalization (i.e. an observationally equivalent definition of m and G) unless the domain of $m(U;\theta)$ is degenerate. If the set $m \circ G^{-1}(y;\theta) = \{z \in \mathbb{R}^{d_m} : \exists u \in \mathcal{U},\ z = m(u;\theta)\ and\ y \in G(u;\theta)\}$ is convex, it is closed in $m(\mathcal{U};\theta)$ hence compact, by assumption 4, so that its support function $\sup_{z \in m \circ G^{-1}(y;\theta)} \lambda'z$ is continuous and assumption 7(ii) holds.

Theorem 1 (Size). *Under assumptions 2 to 7, $\mathbb{P}^*(\tilde{T}_n \leq c_\alpha^* | H_0) \to 1 - \alpha$ almost surely.*

Defining the alternative hypothesis H_a as the violation of H_0, we have consistency of the proposed test.

Theorem 2 (Power). *Under the assumptions of theorem 1, $\mathbb{P}^*(\tilde{T}_n > c_\alpha^* | H_a) \to 1$ almost surely.*

By Theorem 1 and 2, we therefore immediately see that for all $\theta \in \Theta_I$, the probability that $\theta \in \Theta_{CR}$ tends to $1 - \alpha$, whereas for all $\theta \notin \Theta_I$, that same probability tends to 0. In addition, the confidence region provides uniform coverage in the sense that $\liminf_n \inf_{\theta \in \Theta_I} \mathbb{P}(\tilde{T}_n(\theta) \leq c_\alpha(\theta) \mid H_0) \geq 1 - \alpha$, where $c_\alpha(\theta)$ is the quantile of $\hat{T}_n(\theta)$, for which $c_\alpha^*(\theta)$ is a valid approximation. This follows immediately from the proof of theorem 1, which shows the validity of c_α^* as a quantile for $\hat{T}_n$ and inequality $\tilde{T}_n(\theta) \leq \hat{T}_n(\theta)$ shown for all $\theta \in \Theta_I$ in step 1 of the proof of proposition 2.

2.2.3. *Illustration with a revealed preferences example:* Consider for instance revealed preferences in voting behaviour. The spatial model of voting (Downs (1957) and Black (1958)) postulates a common ideological space. Voters face simultaneous elections indexed by e (here $e \in \{1, 2, 3\}$) and each voter chooses exactly one candidate j^e among the candidates competing in election e (here two candidates j^e and k^e for each election e). All candidates j^e are characterized by their position x^{j^e} in the ideological space, which is observed by the voters and the econometrician. Voters are said to "vote ideologically" if their preferences are satiated at an unobservable bliss point x^i for voter i in the ideological space and if they maximize in each election utility function $U(j^e) = -d(x, x^{j^e}; \theta)^2 + u^{j^e}$ where $d(x, y; \theta)^2 = (x - y)^t \theta (x - y)$ and θ is a conformable matrix. The spatial model captures the effect of ideological dimensions through $-d(y, y^j; \theta)^2$, where u^{j^e} is a random utility term and the vector u of $(u^{j^e} - u^{k^e})$ for $e \in \{1, 2, 3\}$ is bounded in norm by $\bar{u}$. The voting profiles Y are observed for a sample of voters, but the ideological position of voters is unobserved, so that the empirical content of the model is $Y \in G(U; \theta)$, where G is the set of profiles compatible with spatial utility maximization. By lemma 1 of Henry and Mourifié (2010), there exist six voting profiles among the eight possible voting profiles which belong to the correspondence equilibrium $G(U; \theta)$ for all realizations of U and for all values of $\theta \in \Theta$ and exactly one of the remaining two voting profiles $\{\bar{v}, \underline{v}\}$ is excluded from $G(U; \theta)$. The function f_λ takes the form $f_\lambda(y) = (1 - \bar{u}\|\lambda\|)1_{\{\bar{v}, \underline{v}\}^c} + \inf_{\|u\| \leq \bar{u}}[1_{\{\bar{v} \notin G(U; \theta)\}} - \lambda' u]1_{\{\bar{v}\}} + \inf_{\|u\| \leq \bar{u}}[1_{\{\underline{v} \notin G(U; \theta)\}} - \lambda' u]1_{\{\underline{v}\}}$. The identified set is the set of all values θ such that $\sup_{\lambda \in \mathbb{R}^3}\{(1 - \bar{u}\|\lambda\|)(1 - p(\bar{v}) - p(\underline{v})) + \inf_{\|u\| \leq \bar{u}}[1_{\{\bar{v} \notin G(U; \theta)\}} - \lambda' u]p(\bar{v}) + \inf_{\|u\| \leq \bar{u}}[1_{\{\underline{v} \notin G(U; \theta)\}} - \lambda' u]p(\underline{v})\} \leq 0$ where $p(\bar{v})$ and $p(\underline{v})$ are the population probabilities that the observable voting profile takes respective values $\bar{v}$ and $\underline{v}$. The test statistic therefore takes the form $\tilde{T}_n = \sup_{\lambda \in \Lambda_n(h_n)}(1/\sqrt{n})\sum_{j=1}^n[(1 - \bar{u}\|\lambda\|)1_{Y_j \notin \{\bar{v}, \underline{v}\}} + \inf_{\|u\| \leq \bar{u}}[1_{\{\bar{v} \notin G(U; \theta)\}} - \lambda' u]1_{Y_j = \bar{v}} + \inf_{\|u\| \leq \bar{u}}[1_{\{\underline{v} \notin G(U; \theta)\}} - \lambda' u]1_{Y_j = \underline{v}}]$ with $\Lambda_n(h_n)$ the set of λ such that the sum in the previous expression is larger than $-h_n$. The procedure delivers a confidence region for θ, which defines voters' utility, and allows to test the pure spatial model of voting (when the identity matrix belongs to the confidence region) or hypotheses of the form *voters value dislike distance in the liberal-conservative dimension of the ideological space more than distance in the dimension of social issues* (see Henry and Mourifié (2010) for details).

CONCLUSION

This paper provides the only methodology for inference in incomplete models with latent variables, where observable variables Y are related to latent variables U by the many-to-many mapping $Y \in$

$G(U)$ and no parametric assumption is entertained on U. This includes moment inequalities (and equalities) with latent variables. Feasible valid confidence regions are derived under continuity of the support function of $m \circ G^{-1}(y; \theta)$, which is satisfied in particular when $m \circ G^{-1}$ has convex values. More work is needed to fully understand the scope of this regularity assumption in specific models beyond the convex case. Confidence regions based on equivalent characterizations of the identified set should also be explored for greater computational efficiency.

Appendix A. Additional results

Proposition 2 (Weak convergence of test statistics). *Under the assumptions of theorem 1, the statistics T_n, $\tilde{T}_n$ and $\hat{T}_n = \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y)$ converge weakly (i.e. in distribution) to the same limit $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$, where $\mathbb{G}_P$ is a Gaussian process.*

Proof of proposition 2. Call ζ_n the indicator function of the event $\{\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n \le (l_n - h_n)\sqrt{n}\}$, with l_n and h_n satisfying assumption 2 and $l_n > h_n$ and $(l_n - h_n)^{-1}\sqrt{\ln \ln n / n} \to 0$. We will prove the theorem in 3 steps:

- First show that $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y) \le T_n \le \tilde{T}_n \le \hat{T}_n$ and $\zeta_n \hat{T}_n \le \zeta_n \sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)$.
- Show that $\zeta_n \to_p 1$.
- Finally show that $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)$ and $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)$ converge weakly to the same limit $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$.

First step: We have

$$
\begin{aligned}
\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y) &= \sup_{\lambda \in \Lambda_0} \left(\mathbb{G}_n f_\lambda(Y) + \sqrt{n} \mathbb{E} f_\lambda(Y) \right) \\
&\le \sup_{\lambda \in \Lambda(h_n)} \sqrt{n} \mathbb{E}_n f_\lambda(Y) = T_n \\
&\le \sup_{\lambda \in \Lambda_n(h_n)} \sqrt{n} \mathbb{E}_n f_\lambda(Y) = \tilde{T}_n \\
&\le \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y) = \hat{T}_n,
\end{aligned}
$$

where the first inequality above holds because $\Lambda_0 \subseteq \Lambda(h_n)$, the second holds because $\mathbb{E}_n f_\lambda(Y) \ge -h_n$ on $\Lambda_n(h_n)$ and $\mathbb{E}_n f_\lambda(Y) < -h_n$ on $\Lambda(h_n) \setminus \Lambda_n(h_n)$ and the third inequality holds under the maintained null hypothesis $\mathbb{E} f_\lambda(Y) \le 0$. Finally, we have

$$
\zeta_n \hat{T}_n \le \zeta_n \sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y).
$$

Indeed, when $\zeta_n = 1$ we have the following implications: $\lambda \in \Lambda_n(h_n) \Rightarrow \mathbb{E}_n f_\lambda(Y) \geq -h_n \Rightarrow$ $\mathbb{G}_n f_\lambda(Y)/\sqrt{n} + \mathbb{E} f_\lambda(Y) \geq -h_n \Rightarrow \mathbb{E} f_\lambda(Y) \geq -h_n - \mathbb{G}_n f_\lambda(Y)/\sqrt{n}$. Since $-h_n - \mathbb{G}_n f_\lambda(Y)/\sqrt{n} > -l_n$ when $\zeta_n = 1$, we have $\lambda \in \Lambda(l_n)$. So $\Lambda_n(h_n) \subseteq \Lambda(l_n)$ if $\zeta_n = 1$ and the result follows.

Second step: Let $\mathcal{F}^{l_n} = \{f_\lambda(y) = \inf_{u \in \mathcal{U}}[1_{\{y \notin G(u)\}} - \lambda' m(u)], \lambda \in \Lambda(l_n)\}$. By lemma 2, $\mathcal{F}^{l_n}$ is a P-Donsker class and the law of iterated logarithm holds. Hence $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)/\sqrt{\ln \ln n} \to_{a.s.}$ $C > 0$. Since $\sqrt{n}(l_n - h_n)/\sqrt{\ln \ln n} \to +\infty$, we have, for any $\varepsilon \in (0, 1)$,

$$\mathbb{P}(|\zeta_n - 1| > \varepsilon) = \mathbb{P}(\zeta_n = 0) = \mathbb{P}\left(\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)/\sqrt{\ln \ln n} > \sqrt{n}(l_n - h_n)/\sqrt{\ln \ln n}\right) \to 0.$$

Third step: By lemma 2, $\mathcal{F}^0 \subseteq \mathcal{F}^{l_n}$ is a P-Donsker class. Hence, $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)$ converges weakly to $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$, where $\mathbb{G}_P$ is a Gaussian process. Taking $\lambda_0 \in \Lambda_0$ (the latter is non-empty since $0 \in \Lambda_0$) , we have: $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y) = \mathbb{G}_n f_{\lambda_0}(Y) + \sup_{\lambda \in \Lambda(l_n)}[\mathbb{G}_n f_\lambda(Y) - \mathbb{G}_n f_{\lambda_0}(Y)] \leq$ $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y) + \sup_{\lambda \in \Lambda(l_n)}[\mathbb{G}_n f_\lambda(Y) - \mathbb{G}_n f_{\lambda_0}(Y)]$. $\Lambda_0 \subseteq \Lambda(l_n)$ implies $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y) \leq$ $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)$. Hence the desired result will follow from $\sup_{\lambda \in \Lambda(l_n)}[\mathbb{G}_n f_\lambda(Y) - \mathbb{G}_n f_{\lambda_0}(Y)] \to_P$ 0. Again, by lemma 2, $\mathcal{F}^{l_n}$ is a P-Donsker class. Now, in view of theorem 1.5.7 of van der Vaart and Wellner (1996), a class $\mathcal{F}$ is P-Donsker if and only $\mathcal{F}$ is totally bounded and asymptotically equicontinuous. Recall that an empirical process is asymptotically equicontinuous if for every sequence $\delta_n \to 0$, $\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \to_P 0$ where $\mathcal{F}_{\delta_n} = \{f - g : f, g \in \mathcal{F}, \rho(f - g) < \delta_n\}$ and $\rho(f) = (\mathbb{E}(f - \mathbb{E}f)^2)^{1/2}$. By lemma 3, for each $\lambda \in \Lambda(l_n)$, there exists a $\lambda_0 \in \Lambda_0$ such that $\rho(f_\lambda - f_{\lambda_0}) < K l_n$ for some $K > 0$ independent of λ, hence by asymptotic equicontinuity, $\sup_{\lambda \in \Lambda(l_n)}[\mathbb{G}_n f_\lambda(Y) - \mathbb{G}_n f_{\lambda_0}(Y)] \to_P 0$, which completes our proof. $\qquad \square$

Lemma 1. $\Lambda(l) = \{\lambda \in \mathbb{R}^l : \mathbb{E} f_\lambda(Y) \geq -l\} \subseteq \{\lambda \in \mathbb{R}^l : \|\lambda\| \leq (1 + l)/\eta\}$.

Proof of Lemma 1. Define $u^*(\lambda)$ such that $\sup_{u \in \mathcal{U}} \lambda' m(u) = \lambda' m(u^*(\lambda))$ (existence is guaranteed by assumption 4).

- If $y \notin G(u^*(\lambda))$ then $f_\lambda(y) = \min(1 - \lambda' m(u^*(\lambda)), -\sup_{y \in G(u)} \lambda' m(u))$.
- If $y \in G(u^*(\lambda))$, then $f_\lambda(y) = \min(-\lambda' m(u^*(\lambda)), -\sup_{y \notin G(u)} \lambda' m(u)) = -\lambda' m(u^*(\lambda))$.

By assumption 7(i), there exists $\eta > 0$ such that $\mathrm{dom}(m(u)) \supseteq B(0, \eta)$, so that $\lambda' m(u^*(\lambda)) \geq \|\lambda\|\eta$. Hence we have $-\lambda' m(u^*(\lambda)) \leq f_\lambda(y) \leq 1 - \|\lambda\|\eta < -l$ for all $\lambda \in \mathbb{R}^l$ such that $\|\lambda\| > (1 + l)/\eta$. We therefore have $\mathbb{E} f_\lambda(Y) < -l$ for all λ such that $\|\lambda\| > (1 + l)/\eta$, and the result follows. $\qquad \square$

Lemma 2. *The class* $\mathcal{F}^l = \{f_\lambda(\cdot), \lambda \in \Lambda(l)\}$ *is P-Donsker.*

Proof of Lemma 2. By lemma 1, for all $\lambda \in \Lambda(l)$ and $y \in \mathcal{Y}$, we have $-(1+l)M/\eta \leq f_\lambda(y) \leq 1$, where $M = \sup_{u \in \mathcal{U}} \|m(u)\|$. Fix $\epsilon > 0$. Consider the sequence $x_0 = -(1+l)M/\eta, \ldots, x_k = 1$. Consider a bracket of the form $[x_{i-1}, x_i]$ with the property $(x_i - x_{i-1}) < \epsilon$ for each i. Hence we have $\mathcal{F}^l \subseteq \cup_{i=1}^k ([x_{i-1}, x_i])$. This bracket has L^1 and L^2-size ϵ since $[(x_i - x_{i-1})^2] < \epsilon^2$. The total number k of such brackets can be chosen smaller than $2(1 + (1+l)M/\eta)/\epsilon$. Then we have $N_{[]}(\epsilon, \mathcal{F}, L^2) \leq 2(1 + (1+l)M/\eta)/\epsilon \Rightarrow J_{[]}(1, \mathcal{F}, L_2) \leq \int_0^1 \sqrt{log(2(1 + (1+l)M/\eta)/\epsilon)}d\epsilon < \infty$. By theorem 19.5 page 270 of van der Vaart (1998), we have $\mathcal{F}^l \in CLT(P)$, which is the desired result. $\square$

Lemma 3. *There exist a constant $K > 0$ such that for all $\lambda \in \Lambda(l)$ and $0 < l < 1$ sufficiently small, there exists $\lambda_0 \in \Lambda_0$ such that $\rho(f_\lambda - f_{\lambda_0}) < Kl$, where $\rho(f) = (\mathbb{E}(f - \mathbb{E}f)^2)^{1/2}$.*

Proof of Lemma 3. As before, set $u^*(\lambda) = \text{argmax } \lambda'm(u)$. Consider the partition $\mathcal{Y}_{(1)} \cup \mathcal{Y}_{(2)} \cup \mathcal{Y}_{(3)}$ of $\mathcal{Y}$ where $\mathcal{Y}_{(1)}(\lambda) = \{y \in \mathcal{Y} : y \in G(u^*(\lambda))\}$, $\mathcal{Y}_{(2)}(\lambda) = \{y \in \mathcal{Y} : y \notin G(u^*(\lambda))$ and $1 - \lambda'm(u^*(\lambda) < -\sup_{y \in G(u)} \lambda'm(u)\}$ and $\mathcal{Y}_{(3)}(\lambda) = \{y \in \mathcal{Y} : y \notin G(u^*(\lambda))$ and $1 - \lambda'm(u^*(\lambda) \geq -\sup_{y \in G(u)} \lambda'm(u)\}$. For each $\lambda \in \Lambda(l)$, there exists under H_0, $0 \leq c \leq 1$ such that $\mathbb{E}f_\lambda(Y) = -\lambda'm(u^*(\lambda))\mathbb{P}(\mathcal{Y}_1(\lambda)) + (1 - \lambda'm(u^*(\lambda)))\mathbb{P}(\mathcal{Y}_2(\lambda)) + \int_{y \in \mathcal{Y}_3(\lambda)} -\sup_{y \in G(u)} \lambda'm(u)d\mathbb{P}(y) = -cl$. If λ_0 can be chosen equal to $(1-h)\lambda$ with $h > 0$ small enough, we show in claim 1 that $u^*(\lambda_0) = u^*(\lambda)$ and $\mathcal{Y}_i(\lambda) = \mathcal{Y}_i(\lambda_0)$ for $i = 1, 2, 3$. Hence,

$$
\begin{aligned}
f_{\lambda_0}(y) &= -\lambda_0'm(u^*(\lambda_0))1_{Y_1(\lambda_0)} + (1 - \lambda_0'm(u^*(\lambda_0)))1_{Y_2(\lambda_0)} - \sup_{y \in G(u)} \lambda_0'm(u_0)1_{Y_3(\lambda_0)} \\
&= -(1-h)\lambda'm(u^*(\lambda))1_{\mathcal{Y}_1(\lambda)} + (1 - (1-h)\lambda'm(u^*(\lambda)))1_{\mathcal{Y}_2(\lambda)} \\
&\qquad\qquad\qquad\qquad\qquad - \sup_{y \in G(u)} (1-h)\lambda'm(u)1_{\mathcal{Y}_3(\lambda)}.
\end{aligned}
$$

So we have

$$
\begin{aligned}
f_\lambda(y) - f_{\lambda_0}(y) &= -h\lambda'm(u^*(\lambda))1_{\mathcal{Y}_{(1)}(\lambda)} - h\lambda'm(u^*(\lambda))1_{\mathcal{Y}_{(2)}(\lambda)} - h\sup_{y \in G(u)} \lambda'm(u)1_{\mathcal{Y}_3(\lambda)} \\
&= h[f_\lambda(y) - 1_{\mathcal{Y}_{(2)}(\lambda)}] \\
&= \frac{h}{1-h}[f_{\lambda_0}(y) - 1_{\{\mathcal{Y}_{(2)}(\lambda)\}}]
\end{aligned}
$$

for $h < 1$ and

$$
\mathbb{E}f_{\lambda_0}(Y) = \mathbb{E}f_\lambda(Y) - h[\mathbb{E}f_\lambda(Y) - \mathbb{P}(\mathcal{Y}_2(\lambda))] = -cl + h[cl + \mathbb{P}(\mathcal{Y}_2(\lambda))],
$$

where the last equality holds because we set $\mathbb{E}f_\lambda(Y) = -cl$. For $h < 1$, we also have

$$
\begin{aligned}
\mathbb{E}(f_{\lambda_0}(Y) - f_\lambda(Y))^2 &= \frac{h^2}{(1-h)^2}\mathbb{E}(f_{\lambda_0}(Y) - 1_{\{\mathcal{Y}_{(2)}(\lambda)\}})^2 \\
&= \frac{h^2}{(1-h)^2}(\mathbb{E}f_{\lambda_0}^2(Y) + \mathbb{P}(\mathcal{Y}_2(\lambda)) - 2\mathbb{E}[f_{\lambda_0}(Y)1_{\{\mathcal{Y}_{(2)}(\lambda)\}}]) \\
&= \frac{h^2}{(1-h)^2}(\mathbb{E}f_{\lambda_0}^2(Y) + \mathbb{P}(\mathcal{Y}_2(\lambda)) - 2\mathbb{E}[(1 - \lambda_0'm(u^*(\lambda)))1_{\{\mathcal{Y}_{(2)}(\lambda)\}}]) \\
&= \frac{h^2}{(1-h)^2}(\mathbb{E}f_{\lambda_0}^2(Y) - \mathbb{P}(\mathcal{Y}_2(\lambda)) + 2\lambda_0'm(u^*(\lambda))\mathbb{P}(\mathcal{Y}_2(\lambda))) \\
&\leq \frac{h^2}{(1-h)^2}\left(1 + \frac{(1+l)M}{\eta}\right)\frac{(1+l)M}{\eta},
\end{aligned}
$$

where the last inequality holds because $-(1+l)M/\eta \leq f_\lambda(y) \leq 1$ for all y, $\|\lambda\| \leq (1+l)/\eta$ for all $\lambda \in \Lambda(l)$ with $M = \sup_{u\in\mathcal{U}}\|m(u)\|$. If $\mathbb{P}(\mathcal{Y}_2(\lambda)) \neq 0$ and we choose $h = cl/[(cl + \mathbb{P}(\mathcal{Y}_2(\lambda)))]$, then $h \to 0$ as $l \to 0$ and we have $\mathbb{E}f_{\lambda_0}(Y) = 0$ hence $\lambda_0 \in \Lambda_0$ and $\mathbb{E}(f_{\lambda_0}(Y) - f_\lambda(Y))^2 \to 0$ when $l \to 0$. The proof is complete in case $\mathbb{P}(\mathcal{Y}_2(\lambda)) \neq 0$.

In the case when $\mathbb{P}(\mathcal{Y}_2(\lambda)) = 0$, define $\tilde{f}_\lambda(y) = -\lambda'm(u^*(\lambda))1_{\{Y_1(\lambda)\}} - \sup_{y\in G(u)}\lambda'm(u)1_{\{Y_3(\lambda)\}} = -\sup_{y\in G(u)}\lambda'm(u)$ and $\lambda_0 = H\lambda$ with $H = I_d - \text{diag}(h)$ and $h = (h_1, ..., h_{d_m})' \longrightarrow 0_d$ when $l \to 0$. Since G^{-1} has closed values under the closed graph condition of assumption 4 and m is continuous, also by assumption 4, $\lambda \mapsto \mathbb{E}\sup_{u\in G^{-1}(Y)}\lambda'm(u)$ is continuous. So for l small enough, we can find a vector $(h_1, ..., h_{d_m})$ such that $\lambda_0 = H\lambda$ and $\lambda_0 \in \Lambda_0$. There remains to show that $\mathbb{E}(\tilde{f}_{\lambda_0}(Y) - \tilde{f}_\lambda(Y))^2 \to 0$ when $l \to 0$. Define now $u^*(y, \lambda) = \text{argmax}_{y\in G(u)}\lambda'm(u)$. The latter exists because G^{-1} has closed values and m is continuous. Now, $\mathbb{E}(\tilde{f}_{\lambda_0}(Y) - \tilde{f}_\lambda(Y))^2 = \mathbb{E}(\lambda_0'm(u^*(y, \lambda_0)) - \lambda'm(u^*(y, \lambda)))^2$ which tends to 0 when $l \to 0$ by assumption 7(ii) and the dominated convergence theorem. This completes the proof. $\square$

Claim 1. *For $h > 0$ sufficiently small, we have $\mathcal{Y}_i(\lambda) = \mathcal{Y}_i(\lambda_0)$ for $i = 1, 2, 3$.*

Proof of claim 1.

First step: Since $u^*(\lambda_0) = \text{argmax } \lambda_0'm(u) = \text{argmax } \lambda'm(u) = u^*(\lambda)$, we have $\mathcal{Y}_1(\lambda) = \mathcal{Y}_1(\lambda_0)$.

Second step:

$$
\begin{aligned}
y \in \mathcal{Y}_2(\lambda) &\Leftrightarrow 1 - \lambda' m(u^*(\lambda)) < - \sup_{y \in G(u)} \lambda' m(u) \\
&\Leftrightarrow 1 - h - \lambda' m(u^*(\lambda)) < - \sup_{y \in G(u)} \lambda_0' m(u) \\
&\Leftrightarrow 1 - \lambda' m(u^*(\lambda_0)) < - \sup_{y \in G(u)} \lambda_0' m(u) \\
&\Leftrightarrow y \in \mathcal{Y}_2(\lambda_0).
\end{aligned}
$$

The third equivalence holds for $h > 0$ sufficiently small and we have $\mathcal{Y}_2(\lambda) = \mathcal{Y}_2(\lambda_0)$.

Third step:

$$
\begin{aligned}
y \in \mathcal{Y}_3(\lambda) &\Rightarrow 1 - \lambda' m(u^*(\lambda)) \geq - \sup_{y \in G(u)} \lambda' m(u) \\
&\Rightarrow 1 - h - \lambda' m(u^*(\lambda)) \geq - \sup_{y \in G(u)} \lambda_0' m(u) \\
&\Rightarrow 1 - \lambda' m(u^*(\lambda_0)) \geq - \sup_{y \in G(u)} \lambda_0' m(u) \\
&\Rightarrow y \in \mathcal{Y}_3(\lambda_0)
\end{aligned}
$$

The third implication holds because $h > 0$. Steps one and two imply that $\mathcal{Y}_3(\lambda) = \mathcal{Y}_3(\lambda_0)$. This completes the proof of claim 1. $\qquad\square$

Appendix B. Proof of results in the main text

Proof of Theorem 1. The goal of this proof is to show that $\tilde{T}_n = \sup_{\lambda \in \Lambda_n(h_n)} \sqrt{n} \mathbb{E}_n f_\lambda(Y)$ and $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n^* f_\lambda(Y)$ have the same limit in distribution, where $\mathbb{G}_n^*$ is the bootstrapped empirical process. As in the proof of proposition 2, Call ζ_n the indicator function of the event $\{\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n \leq (l_n - h_n)\sqrt{n}\}$, with l_n and h_n satisfying assumption 2 and $l_n > h_n$ and $(l_n - h_n)^{-1} \sqrt{\ln \ln n / n} \to 0$. In addition, define $\zeta_n^0 = \{\sup_{\lambda \in \Lambda_0} \mathbb{G}_n \geq -h_n \sqrt{n}\}$. We will prove the theorem in 5 steps.

First step: As shown in the proof of proposition 2, $\tilde{T}_n$ converges in distribution to $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$.

Second step: By construction, $\Lambda_0 \subseteq \Lambda(l_n)$. Hence $\mathcal{F}^0 \subseteq \mathcal{F}^{l_n}$ which is a P-Donsker class by lemma 2. Since $\Lambda(l_n)$ is bounded by lemma 1, $\mathcal{F}^0$ and $\mathcal{F}^{l_n}$ admit a square integrable envelope, so by Theorem 2.4 page 857 of Giné and Zinn (1990), the bootstrapped empirical process $\mathbb{G}_n^*$ converges uniformly (over $\mathcal{F}^{l_n}$, hence also $\mathcal{F}^0$) to $\mathbb{G}_P$ conditionally almost surely and $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n^* f_\lambda(Y)$ and $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n^* f_\lambda(Y)$ have the same respective limits (conditionally almost surely) as $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)$ and $\sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n f_\lambda(Y)$. The latter both converge to $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$ by the proof of proposition 2.

Third step: We have

$$
\begin{aligned}
\lambda \in \Lambda_0 \quad &\Rightarrow \quad \mathbb{E}f_\lambda(Y) = 0 \\
&\Rightarrow \quad \mathbb{G}_n f_\lambda(Y)/\sqrt{n} - \mathbb{E}_n f_\lambda(Y) = 0 \\
&\Rightarrow \quad \mathbb{E}_n f_\lambda(Y) = \mathbb{G}_n f_\lambda(Y)/\sqrt{n} \\
&\Rightarrow \quad \mathbb{E}_n f_\lambda(Y) \geq -h_n \text{ if } \zeta_n^0 = 1 \\
&\Rightarrow \quad \lambda \in \Lambda_n(h_n) \text{ if } \zeta_n^0 = 1.
\end{aligned}
$$

Hence $\Lambda_0 \subseteq \Lambda_n(h_n)$ if $\zeta_n^0 = 1$. So $\zeta_n^0 \sup_{\lambda \in \Lambda_0} \mathbb{G}_n^* f_\lambda(Y) \leq \zeta_n^0 \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n^* f_\lambda(Y)$.

Fourth step: In the first step of proposition 2, we showed that $\Lambda_n(h_n) \subseteq \Lambda(l_n)$ if $\zeta_n = 1$. Hence,
$\zeta_n \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n^* f_\lambda(Y) \leq \zeta_n \sup_{\lambda \in \Lambda(l_n)} \mathbb{G}_n^* f_\lambda(Y)$.

Fifth step: As seen in step 2, $\mathcal{F}^0$ is a P-Donsker class, so the law of iterated logarithm holds and $\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)/\sqrt{\ln \ln n} \to_{a.s}$ constant, since $-\sqrt{n}h_n/\sqrt{\ln \ln n} \to -\infty$. We have for any $\epsilon \in (0, 1)$,
$\mathbb{P}(|\zeta_n^0 - 1| > \epsilon) = \mathbb{P}(\zeta_n^0 = 0) = \mathbb{P}(\sup_{\lambda \in \Lambda_0} \mathbb{G}_n f_\lambda(Y)/\sqrt{\ln \ln n} < -\sqrt{n}h_n/\sqrt{\ln \ln n}) \to 0$. So we have $\zeta_n^0 \to_p 1$.

By steps 1 to 5, $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n^* f_\lambda(Y)$ converges to $\sup_{\lambda \in \Lambda_0} \mathbb{G}_P f_\lambda(Y)$ conditionally almost surely, which completes the proof. $\qquad \square$

Proof of Theorem 2. Under H_a, there is a $\lambda^a \in \Lambda(h_n)$ such that $\mathbb{E}f_{\lambda^a}(Y) > 0$. $\tilde{T}_n \geq T_n = \sup_{\lambda \in \Lambda(h_n)} \sqrt{n}\mathbb{E}_n f_\lambda(Y) = \sup_{\lambda \in \Lambda(h_n)} \mathbb{G}_n f_\lambda(Y) + \sqrt{n}\mathbb{E}f_\lambda(Y) \geq \mathbb{G}_n f_{\lambda^a}(Y) + \sqrt{n}\mathbb{E}f_{\lambda^a}(Y)$. The first inequality is shown in the first step of the proof of proposition 2. Hence, $\tilde{T}_n - \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}^* f_\lambda(Y) \geq \mathbb{G}_n f_{\lambda^a}(Y) + \sqrt{n}\mathbb{E}f_{\lambda^a}(Y) - \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y) + (\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y) - \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}^* f_\lambda(Y))$. Since $\mathbb{E}f_{\lambda^a}(Y) > 0$ we have $\sqrt{n}\mathbb{E}f_{\lambda^a}(Y) \to +\infty$.

We have also $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y) - \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}^* f_\lambda(Y) = O_p(1)$, since we showed in the proof of theorem 1 that $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y)$ and $\sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}^* f_\lambda(Y)$ have a same weak limit $\sup_{\lambda \in \Lambda_0} \mathbb{G}(f_\lambda(Y))$. Hence, since $\mathbb{G}_n f_{\lambda^a}(Y) - \sup_{\lambda \in \Lambda_n(h_n)} \mathbb{G}_n f_\lambda(Y)$ is a tight sequence (this is can be derived from exponential bounds in 2.14.9 page 246 of van der Vaart and Wellner (1996)), we have the desired result. $\qquad \square$

References

ANDREWS, D., S. BERRY, and P. JIA (2003): "On placing bounds on parameters of entry games in the presence of multiple equilibria," unpublished manuscript.

ANDREWS, D., and X. SHI (2010): "Inference for papameters defined by conditional moment inequalities," unpublished manuscript.

ANDREWS, D., and G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

BAJARI, P., H. HONG, and S. RYAN (2010): "Identification and estimation of discrete games of complete information," forthcoming in *Econometrica*.

BLACK, D. (1958): *The theory of committees and elections*. Cambridge University Press.

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

CHERNOZHUKOV, V., H. HONG, and E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, 75, 1243–1285.

CHERNOZHUKOV, V., S. LEE, and A. ROSEN (2009): "Interaction bounds: estimation and inference," unpublished manuscript.

CHESHER, A. (2010): "Instrumental variable models for discrete outcomes," *Econometrica*, 78, 575–601.

CILIBERTO, F., and E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 70, 1791–1828.

DOWNS, A. (1957): *An economic theory of democracy*. New York: Harper and Row.

EKELAND, I., A. GALICHON, and M. HENRY (2010): "Optimal transportation and the falsifiability of incompletely specified economic models.," *Economic Theory*, 42, 355–374.

GALICHON, A., and M. HENRY (2008): "Set identification in models with multiple equilibria," forthcoming in the *Review of Economic Studies*.

GALICHON, A., and M. HENRY (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152, 186–196.

GINÉ, E., and S. ZINN (1990): "Bootstrapping general empirical measures," *Annals of Probability*, 18, 851–859.

HANSEN, P. (2005): "A test of superior predictive ability," *Journal of Business and Economics Statistics*, 23, 365–380.

HENRY, M., R. MÉANGO, and M. QUEYRANNE (2010): "Bootstrap inference in normal form games," Available at SSRN: http://ssrn.com/abstract=1758743.

HENRY, M., and I. MOURIFIÉ (2010): "Euclidean revealed preferences," forthcoming in the *Journal of Applied Econometrics*.

JOVANOVIC, B. (1989): "Observable implications of models with multiple equilibria," *Econometrica*, 57, 1431–1437.

KOOPMANS, T., and O. REIERSOL (1950): "The identification of structural characteristics," *Annals of Mathematical Statistics*, 21, 165–181.

MATZKIN, R. (1994): "Restrictions of economic theory in nonparametric methods," in *Handbook of Econometrics, vol 4, R. F. Engel and D. L. McFadden, eds.*, pp. 1–16. North Holland.

MENZEL, K. (2009): "Estimation and inference with many moment inequalities," unpublished manuscript.

VAN DER VAART, A. (1998): *Asymptotic Statistics*. Cambridge University Press.

VAN DER VAART, A., and J. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer.