FINITE SAMPLE INFERENCE IN INCOMPLETE MODELS

LIXIONG LI AND MARC HENRY

Abstract. We propose confidence regions for the parameters of incomplete models with exact coverage of the true parameter in finite samples. Our confidence region inverts a test, which generalizes Monte Carlo tests to incomplete models. The test statistic is a discrete analogue of a new optimal transport characterization of the sharp identified region. Both test statistic and critical values rely on simulation draws from the distribution of latent variables and are computed using solutions to discrete optimal transport, hence linear programming problems. We also propose a fast preliminary search in the parameter space with an alternative, more conservative yet consistent test, based on a parameter free critical value.

Keywords: Incomplete models, sharp identification region, simulation-based testing, finite sample inference, optimal transport.

JEL codes: C15, C57, C61

Introduction

In this paper, we study a class of incomplete econometric models that combines (i) a restriction on the support of the random variables involved in the model specification, and (ii) a restriction on the distribution of those variables in the model, that the analyst cannot observe. The support restriction is implied by economic theory, and usually involves the implications of behavioral assumptions, equilibrium concepts and structural features of the economic environment. A game of perfect information with a pure strategy equilibrium concept, as in Jovanovic [1989] and Tamer [2003] is a prime example. Other examples include models of choice with limited attention, as in Barseghyan et al. [2021], discrete choice with endogeneity, as in Chesher et al. [2013],

The first version is of February 14, 2022. This version is of April 4, 2022. Corresponding author: Lixiong Li: lixiong.li@jhu.edu Department of Economics, Johns Hopkins University, Wyman Park Building 5th Floor, 3100 Wyman Park Drive, Baltimore, MD 21211.

auction models, as in Haile and Tamer [2003], network formation, as in de Paula et al. [2018], and structural vector autoregressions, as in Giacomini and Kitagawa [2021] and Giacomini et al. [2021]. Molinari [2020] and Chesher and Rosen [2020] provide comprehensive surveys of the literature on incomplete structural models.

Incomplete structural models are called *incomplete* because the model structure does not predict a single data generating process for the observed variables for all values of the model parameter. Incompleteness arises because of multiple equilibria in games, unobserved heterogeneity in choice sets in limited attention models, interval predictions in auctions, and unknown sample selection mechanisms. Model incompleteness generally leads to partial identification, where more than one value of the model parameter could have given rise to the true data generating process for the observed variables. However, model incompleteness and partial identification are distinct concepts.

The current state of the art in deriving confidence regions for the parameters of incomplete structural models involves the Beresteanu et al. [2011]-Galichon and Henry [2011] characterization of the sharp identified region as a collection of conditional moment inequality restrictions, and the application of one of the existing inference methods with conditional moment inequality models, surveyed in Canay and Shaikh [2018] and Molinari [2020]. This method, however, results in a very large, possibly infinite, number of conditional moment inequalities. Even in cases, where the endogenous variable is discrete, such as discrete games, the cardinality of the number of moment inequalities increases exponentially in the number of strategy profiles.

The challenge is both computational and statistical, as the number of inequalities may be much larger than the sample size, requiring new methods, such as Chernozhukov et al. [2019]. Basing inference on a non sharp reduced collection of inequalities leads to low power and loss of robustness to misspecification. See Kédagni et al. [2020] for a discussion. Methods to reduce the number of conditional moment inequalities without losing sharpness exist. They are based on core determining classes, as proposed in Galichon and Henry [2011] and further developed in Chesher et al. [2013], Chesher and Rosen [2017], and Luo and Wang [2017]. However, these methods are complex, model specific, and only partially alleviate the problem. In addition, when the conditional moment inequalities are transformed into unconditional ones, as in Andrews and Shi [2013], sharpness is preserved only when the number of moment inequalities increases with sample size, which induces an extra layer of computational burden.[1] Moreover, inference methods in moment inequalities rely on asymptotic arguments[2] and some user chosen tuning parameter to preselect inequalities that are close to binding in the sample and thereby avoid overly conservative inference.

We propose an alternative method to construct confidence regions for the parameters of incomplete structural models that circumvents the many moments and conditioning issues, and avoids tuning parameters and asymptotic arguments. As is customary with moment inequality models, we construct our confidence region by inverting a test. However, the test statistic is based on a different characterization of the sharp identified region, and we show that it controls size in finite samples. Our testing procedure relies on two key ingredients. First, the test statistic is based on an optimal transport characterization of the sharp identified region, inspired by formulations in Galichon and Henry [2006] and Ekeland et al. [2010]. As a result, the test statistic is the solution of a discrete optimal transport problem, which is a special kind of linear programming problem, the computation of which has a long history. Second, the test generalizes Monte Carlo tests of Dwass [1957] and Barnard [1963][3] to incomplete models to control size in finite samples. The test statistic and critical values are based on simulation draws from the conditional distribution of latent variables.

Our test controls size, hence coverage probability of the confidence region for any finite sample size. Finite sample validity has several advantages, beyond the obvious benefit of avoiding reliance on often questionable asymptotic approximations. First, the support constraint and the dimension of the vector of unobservables may change with sample size, as would arise in applications to games on networks and network formation games[4]. Second, our finite sample validity result requires no restriction on

[1]The dimensionality of the conditioning set in such models generally precludes the alternative approach to conditional moment inequalities, which involves estimating them, as in Chernozhukov et al. [2013].

[2]Chernozhukov et al. [2013] and Chernozhukov et al. [2019] derive non asymptotic bounds on the rejection probabilities of their confidence regions. These bounds are useful to derive asymptotic rates of convergence, not for finite sample inference.

[3]See also Dufour [2006] and Dufour and Khalaf [2001].

[4]See Example 1.2.3.

the dependence between observations in the sample. This property is particularly desirable with incomplete models. As discussed in Epstein et al. [2016], it is hard to reconcile the customary independence or mixing assumptions across units of observation with total ignorance of the mechanism that selected each realization from the model prediction set. The degree of dependence between observations does not affect size control, but it does affect the power of the test, hence informativeness of the confidence region. However, a simple ergodicity condition is sufficient to ensure that parameter sequences that violate the optimal transport characterization of the sharp identified region ultimately lie outside the confidence region.

Our method requires a search in the space of parameters. At each value of the parameter in the search, we must compute a test statistic and a critical value. This computational burden is shared by inference methods in partially identified models, where the objective is coverage of the true value of the parameter. In order to accelerate the search, we propose a conservative superset of our confidence region. The conservative superset is based on a parameter free critical value, and hence covers the sharp identified region. Once this conservative confidence region is computed, all values of the parameter that lie outside of it can be excluded a priori from the exact confidence region in our main proposal.

Notation and preliminaries. All random vectors are defined on the same complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. All vectors are written as row vectors throughout. Throughout the paper, (Y, X, U) will denote a random vector on $\mathcal{Y} \times \mathcal{X} \times \mathcal{U}$, and $\theta \in \Theta$ a fixed parameter vector, where $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, $\mathcal{X} \subseteq \mathbb{R}^{d_X}$, $\mathcal{U} \subseteq \mathbb{R}^{d_U}$, and $\Theta \subseteq \mathbb{R}^{d_\theta}$. We will denote $\mathcal{Q}$ and $\mathcal{P}$ the collections of Borel probability measures on $\mathcal{U} \times \mathcal{X}$ and $\mathcal{Y} \times \mathcal{X}$ respectively. $\mathcal{M}(Q, P)$ is the set of probability measures on $(\mathcal{U} \times \mathcal{X}) \times (\mathcal{Y} \times \mathcal{X})$ with marginals Q on $\mathcal{U} \times \mathcal{X}$ and P on $\mathcal{Y} \times \mathcal{X}$. We denote by d a lower semi-continuous metric on $\mathcal{U} \times \mathcal{X}$, and the distance $d(a, A)$ between a point a and a set A is be defined as $d(a, A) = \inf_{a' \in A} d(a, a')$. The convex hull of a set A is denoted co$A$. We denote $\mathcal{M}_n^+$ the set of $n \times n$ non negative matrices, and Π_n the subset of $\mathcal{M}_n^+$ containing matrices π such that $n\pi$ is doubly stochastic, i.e., such that $\Sigma_i \pi_{ij} = \Sigma_j \pi_{ij} = 1/n$, for all $i, j \leq n$. Finally, $\mathcal{S}_n$ is the set of permulations σ on $\{1, \ldots, n\}$, and δ_x denotes the Dirac mass concentrated at x.

Overview. Section 1 defines the model and characterizes the sharp identified region. We present the finite sample inference procedure in Section 2. Section 3 proposes a procedure to reduce the computational burden of the search in the parameter space. Section 4 shows consistency of the specification test, and Section 5 is a simulation analysis of the informativeness and computational intensiveness of the proposed procedure.

1. Incomplete models

1.1. **Theoretical structural model.** We restrict attention to the class of parametric incomplete structural models introduced in Jovanovic [1989]. The vector of variables of interest $(Y, X, U) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{U}$ satisfies support constraint $(Y, X, U) \in \Gamma(\theta_1) \subseteq \mathcal{Y} \times \mathcal{X} \times \mathcal{U}$, and the probability distribution of U conditional on possible realizations x of X is given by $Q_{U|x;\theta_2}$. The object of inference is the finite dimensional parameter $\theta := (\theta_1, \theta_2) \in \Theta$. Both vectors or variables Y and X are observed, in the sense that available data consists in a sample $((Y_1, X_1), \ldots, (Y_n, X_n)))$. Variables in vector U are unobserved. Variables in vector X are exogenous, in the sense that the conditional distribution $Q_{U|x;\theta_2}$ is fixed a priori. This includes $U \perp\!\!\!\perp X$ as special case. All endogenous (i.e., non-exogenous) variables are subsumed in vector Y.

The incompleteness of the model is reflected in two ways. First, multiple values of the endogenous variables may be consistent with a single value of the exogenous and unobserved variables. This can be seen in the fact that the set $\{y \in \mathcal{Y} : (y, x, u) \in \Gamma(\theta_1)\}$ may not be a singleton for all $(u, x) \in \mathcal{U} \times \mathcal{X}$. This corresponds to the fact that the model fails to produce a unique prediction. Second, multiple values of the unobservable variable may be consistent with a single value of the observable variables. This can similarly be seen in the fact that the set $\{u \in \mathcal{U} : (y, x, u) \in \Gamma(\theta_1)\}$ may not be a singleton for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Multiple unobservables could have given rise to the same observations.

1.2. **Examples.** Incomplete models as described above encompass examples as diverse as static simultaneous move games with complete information and pure strategy equilibrium concepts, choice models with limited attention or partially observed consideration sets, auctions with independent private values. Section 3 in Molinari [2020] gives a detailed account of such incomplete structural models with extensive references. In what follows, we concentrate on three recent examples to illustrate precisely how they fit within the framework.

1.2.1. *Discrete choice with unobserved heterogeneity in consideration sets.* We set out the structural model in Barseghyan et al. [2021] in their notation, before translating it into our framework. Consider a finite set of alternatives $\mathcal{D}$ for a decision maker to choose from. A decision maker is characterized by observed covariates X on $\mathcal{X}$ and unobserved random vector $\nu \in \mathcal{V}$ with distribution $P_{\nu;\delta_2}$, where δ_2 is a fixed unknown parameter vector (see Assumption 2.1 page 4 of Barseghyan et al. [2021]). The decision maker makes observed decision d based on the maximization of utility $d^*(G, X, \nu; \delta_1) := \arg\max_{c \in G} W(c, X, \nu; \delta_1)$, over a subset G of the full set $\mathcal{D}$ of alternatives, where δ_1 is a fixed unknown parameter vector. Unovserved heterogeneity in choice sets is the driver of incompleteness in this model. It is disciplined by the assumption that the realized choice set $C \subseteq \mathcal{D}$ under consideration satisfies $\mathbb{P}(|C| \geq \kappa) = 1$, for some $\kappa \geq 2$, fixed and known. The model therefore stipulates that the observed choice d must be in

$$D^*_\kappa := \bigcup_{G \subseteq \mathcal{D}: |G| \geq \kappa} \{d^*(G, X, \nu; \delta_1)\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \{d^*(G, X, \nu; \delta_1)\},$$

where the equality follows from Sen's property α, as shown in Barseghyan et al. [2021].[5]

This example fits into the current framework, with $Y := d$, $U := \nu$, $\theta = (\theta_1, \theta_2) := (\delta_1, \delta_2)$, $\Gamma(\theta_1) := \{(y, x, u) : y \in D^*_\kappa\}$, and $Q_{U|X;\theta_2} := P_{\nu;\delta_2}$.

1.2.2. *Market Structure and Competition in Airline Markets.* Once again, we set out the structural model in the notation of Ciliberto et al. [2021], before translating it into our framework. Two firms, indexed $j \in \{1, 2\}$ decide whether to enter a market based on the profit they expect under optimal pricing. If Firm 1 enters, it faces demand $\tilde{s}_j(p, X, y, \xi; \beta)$, which is a function of the vector of endogenous prices $p = (p_1, p_2)$, the vector of exogenous demand relevant firm characteristics $X = (X_1, X_2)$, the binary entry decisions $y = (y_1, y_2)$ of both firms, unobservable demand shocks ξ

[5]Sen's property α is the independence of irrelevant alternatives of individual choice theory.

and parameter vector β. Fixed costs of entry for Firm $j \in \{1,2\}$, is $F(Z_j, \nu_j; \gamma)$, and marginal unit cost of production is $c(W_j, \eta_j; \delta)$, where $W = (W_1, W_2)$ and $Z = (Z_1, Z_2)$ are exogenous observed cost shifters, $\nu = (\nu_1, \nu_2)$, $\eta = (\eta_1, \eta_2)$ and ξ are unobserved cost shifters, and β, γ, δ, are parameters.

Structural model constraints include for each firm $j \in \{1,2\}$: equality of predicted and realized demand share

$$S_j = \tilde{s}_j(p, X, y, \xi; \beta), \tag{1.1}$$

an entry condition, namely $y_j = 1$ if and only if

$$\pi_j := (p_j - c(W_j, \eta_j; \delta))\mathcal{M}\tilde{s}_j(p, X, y, \xi; \beta) - F(Z_j, \nu_j; \gamma) \geq 0, \tag{1.2}$$

where $\mathcal{M}$ is observed market size, and an equilibrium pricing condition in case of entry

$$(p_j - c(W_j, \eta_j; \delta))\frac{\partial \tilde{s}_j}{\partial p_j}(p, X, y, \xi; \beta) + \tilde{s}_j(p, X, y, \xi; \beta) \;\; = \;\; 0. \tag{1.3}$$

This example fits into the current framework, with $Y := (p, S, y)$, $X = (\mathcal{M}, W, Z, X)$, $U := (\nu, \xi, \eta) \sim Q_{U|X;\theta_2} := N(0, \Sigma)$, $\theta_1 = (\beta, \gamma, \delta)$, $\theta_2 := \Sigma$ and $\Gamma(\theta_1) := \{(y, x, u) :$ (1.1) $-$ (1.3) hold for $j = 1, 2\}$.

1.2.3. *Network formation.* Observe a single network with adjacency matrix G, a vector of individual characteristics X, a matrix of player-pair unobservable shocks ε (ε_{ij} enters utility of player i is linked to j). The utility of individual i is $u(G, X, \varepsilon_i)$, where $\varepsilon_i = (\varepsilon_{ij})_j$. See (1) in de Paula et al. [2018] for an example. Assume all links are mutually beneficial. This is weaker than pairwise stability since some mutually beneficial links may be missing (which can be rationalized with search frictions). Call A_i the local adjacency matrix that is utility relevant to player i. This may be the whole network adjacency matrix, of it may be a restriction. For instance, in de Paula et al. [2018], players cannot form more than L links and cannot link at distance larger than D. Call $A_{i,-l}$ the local adjacency matrix after link l was removed. Then, for each i and each $l \in N(i)$ (set of direct neighbors of i),

$$u(A_i, X, \varepsilon_i) \geq u(A_{i,-l}, X, \varepsilon_i). \tag{1.4}$$

This fits into our framework, with $Y = A$, $U = \varepsilon$ distributed according to $Q_{U|X;\theta_2}$, θ_1 the parameters of the utility function, and $(A, X, \varepsilon,) \in \Gamma(\theta)$ if and only if (1.4) holds.

1.3. **Sharp identified region.** The sample $((y_1, x_1), \dots, (y_n, x_n)))$ of observed data is assumed to be a realization from the random vector $((Y_1, X_1), \dots, (Y_n, X_n)))$ with true distribution $P_0^{(n)}$. The model stipulates that the latter is an element of a subset $\mathcal{P}_\theta^{(n)}$ of the set of distributions on $(\mathcal{Y} \times \mathcal{X})^n$. The set $\mathcal{P}_\theta^{(n)}$ is defined as follows.

Definition 1 (Structural model)**.** For each $\theta = (\theta_1, \theta_2) \in \Theta$, $\mathcal{P}_\theta^{(n)}$ is the set of distributions $P^{(n)}$ on $(\mathcal{Y} \times \mathcal{X})^n$ with identical marginals $P_{0n} := P_{Y|X,0n} \times P_{X,0n}$ on $\mathcal{Y} \times \mathcal{X}$, such that for any random vector $((Y_1, X_1), \dots, (Y_n, X_n)))$ distributed according to $P^{(n)}$, there exists a random vector $(U_1, \dots, U_n)$ with support $\mathcal{X}^n$ that satisfies the following constraints:

(1) Support restriction: $(Y_i, X_i, U_i) \in \Gamma(\theta_1) \subseteq \mathcal{Y} \times \mathcal{X} \times \mathcal{U}$ for all $i \leq n$, almost surely.
(2) Latent variables generating process restriction: U_i has distribution $Q_{U|X_i;\theta_2}$ conditionally on X_i for all $i \leq n$, and the $(U_1, \dots, U_n)$ are independently distributed[6] conditionally on $X^{(n)} := (X_1, \dots, X_n)$.

Compatibility between the structural model of Definition 1 and the true data generating process is defined as the fact that $P_0^{(n)}$ is an element of $\mathcal{P}_\theta^{(n)}$. Because of the incompleteness of the model, for any given θ, the structure model may generate multiple predictions for the process generating the observed data, i.e., $\mathcal{P}_\theta^{(n)}$ may not be a singleton. Conversely, any given true data generating process $P_0^{(n)}$ may be compatible with the structural model, i.e., $P_0^{(n)} \in \mathcal{P}_\theta^{(n)}$, for more than one value of the parameter $\theta \in \Theta$. Hence the parameter vector θ is partially identified. The sharp identified region $\Theta_I^{(n)}$ is defined as the set of values of the parameter θ such that our model is compatible with the true data generating process.

Definition 2. The sharp identification region is defined for each $n \geq 1$ as

$$\Theta_I^{(n)} := \{\theta \in \Theta : P_0^{(n)} \in \mathcal{P}_\theta^{(n)}\}.$$

[6]It is important to note that the assumption of independence of the latent variable across observation units does not imply independence of outcomes. In particular, the outcome selection process may be arbitrarily correlated across observation units.

We assume the structural model specification is non trivial in the sense that for all $n \geq 1$, there exists $\theta \in \Theta$ such that $\mathcal{P}_\theta^{(n)} \neq \emptyset$. In other words, the sharp identification region is non empty for at least one true data generating process. However, the sharp identification region may be empty for some true data generating process $P_0^{(n)}$, in which case the structural model is incompatible with the data, and should be rejected.

The existing characterization of the sharp identified region, derived in Beresteanu et al. [2011] and Galichon and Henry [2011], takes the form of a collection of conditional moment inequalities of typically very large cardinality. Our inference strategy is based on a different characterization of the sharp identified region as the solution of an optimal transport problem, and as such, is related to characterization in Galichon and Henry [2006] and Ekeland et al. [2010].

The fundamental idea applied here also underlies characterizations in Galichon and Henry [2006] and Ekeland et al. [2010]: the existence of a joint distribution $\tilde{\pi}$ for (Y, X, U) that satisfies the model is equivalent to the minimum of $\tilde{\pi}((Y, X, U) \notin \Gamma(\theta_1))$ among joint distributions $\tilde{\pi}$ satisfying the marginal constraints being equal to 0.

The way we treat dependence on exogenous variables X is crucially different from those previous proposals. It relies on a reformulation of the support constraint in the model. Define the correspondences Γ_u and Γ_y between $\mathcal{Y} \times \mathcal{X}$ and $\mathcal{U} \times \mathcal{X}$ by:

$$\begin{aligned} \Gamma_u(y, x; \theta) &= \{(u, x') \in \mathcal{U} \times \mathcal{X} : x' = x \text{ and } (y, x, u) \in \Gamma(\theta_1)\}, \\ \Gamma_y(u, x; \theta) &= \{(y, x') \in \mathcal{Y} \times \mathcal{X} : x' = x \text{ and } (y, x, u) \in \Gamma(\theta_1)\}. \end{aligned} \tag{1.5}$$

Correspondence Γ_y defines the set of model predictions for the endogenous variables, whereas correspondence Γ_u defines the set of latent variables that can rationalize the data. We define the correspondences between $\mathcal{Y} \times \mathcal{X}$ and $\mathcal{U} \times \mathcal{X}$ instead of simply $\mathcal{U}$ in order to avoid conditioning on X.

With the notation of (1.5), and writing $q := (U, X)$ and $p := (Y, X)$, the distributional constraint (Constraint (2) in Definition 1) can be written $q \sim Q := Q_{U|X;\theta_2} \times P_{X,0n}$. Moreover, for this model to be consistent with the true data generating process, we need $p \sim P := P_{0n}$. Let $\mathcal{M}(Q, P)$ be defined as the set of joint distributions with marginals Q and P (see notations and preliminaries). Then, the above two restrictions imply that the joint distribution π of (q, p) must satisfy $\pi \in \mathcal{M}(Q, P)$.

Finally, the support constraint in the definition of the structural model (Definition 1) is $\pi(q \in \Gamma_u(p;\theta)) = 1$, or, equivalently, $\int d(q, \Gamma_u(p;\theta))d\pi(q,p) = 0$ for any metric d on $\mathcal{U} \times \mathcal{X}$, if $\Gamma_u(p;\theta)$ is closed.

Therefore, if the model and parameter θ are compatible with the true data generating process, the following must hold:

$$\mathcal{D}(Q, P;\theta) \ := \ \min_{\pi \in \mathcal{M}(Q,P)} \int d\left(q, \Gamma_u(p;\theta)\right) d\pi(q,p) \ = \ 0. \tag{1.6}$$

Here $\mathcal{D}(Q, P;\theta)$ can be viewed as an optimal transport problem (see Villani [2003]) with cost function $(q,p) \mapsto d(q, \Gamma_u(p;\theta))$. The following theorem shows that condition (1.6) is not only necessary, but also sufficient.

Theorem 1 (Characterization of the sharp identified region). *Assume the true data generating process $P_0^{(n)}$ has n identical and independent marginals P_{0n}, and Γ_u is closed-valued (i.e., $\Gamma_u(y,x;\theta)$ is closed for all (y,x) and all θ). Then*

$$\Theta_I^{(n)} = \left\{\theta \in \Theta : \ \mathcal{D}(Q_{U|X;\theta_2} \times P_{X,0n}, P_{0n};\theta_1) = 0\right\}.$$

An immediate benefit of characterizing the sharp identified region in Theorem 1 with the optimal transport formulation (1.6) is that a sample analogue, where P is replaced with the sample empirical distribution, readily provides a test statistic. Although Theorem 1 is shown to hold for independent observations, our inference procedure, detailed in the next section, allows for a general pattern of dependence. This avoids the need for statistical restrictions on the data generating process, such as independence or mixing, whose suitability is difficult to assess in an incomplete model framework (see Epstein et al. [2016] for a discussion).

2. Finite sample inference

The objective of this section is to provide a confidence region for the parameters of interest θ. The confidence region CR_n is obtained by test inversion, as in Anderson and Rubin [1949]. For each value of θ, we test the null hypothesis

$$H_0^{(n)}(\theta) : P_0^{(n)} \in \mathcal{P}_\theta^{(n)}.$$

The hypothesis is rejected and θ deemed outside the confidence region if and only if the test statistic $T_n(\theta)$, a function of the sample $((Y_1, X_1), \ldots, (Y_n, X_n))$, is larger than a corresponding critical value $c_{n,1-\alpha}(\theta)$. Hence

$$CR_n \;:=\; \{\theta \in \Theta : T_n(\theta) \leq c_{n,1-\alpha}(\theta)\}. \tag{2.1}$$

The rest of this section is devoted to constructing the test statistic $T_n(\theta)$ and the critical value $c_{n,1-\alpha}(\theta)$ to ensure exact coverage of the true parameter value in finite samples. Finite sample inference is achieved with an extension to incomplete models of traditional Monte Carlo tests of Dwass [1957] and Barnard [1963]. Our testing procedure relies on simulated samples of unobservables.

Definition 3 (Monte Carlo latent samples)**.** A Monte Carlo latent sample $\tilde{U}^{(n)}$ is a collection $(\tilde{U}_1, \ldots, \tilde{U}_n)$ of independent vectors conditional on $X^{(n)} := (X_1, \ldots, X_n)$ such that for each $i \leq n$, $\tilde{U}_i$ is drawn from the conditional distribution $Q_{U|X_i;\theta_2}$.

More precisely:

(1) Let $\nu^{(n)} := (\nu_1, \ldots, \nu_n)$ be an i.i.d. sample of uniform random vectors on $[0,1]^{d_U}$;
(2) Let $\preccurlyeq$ be an arbitrary ordering of $\mathcal{X}$, e.g., the lexicographic order on $\mathbb{R}^{d_X}$;
(3) Let r be a permutation[7] of $\{1, \ldots, n\}$ such that $X_{r(1)} \preccurlyeq \ldots \preccurlyeq X_{r(n)}$;
(4) Let q be a map such that $q(\nu|x;\theta_2) \sim Q_{U|x;\theta_2}$ when $\nu \sim U[0,1]^{d_U}$ (see Chernozhukov et al. [2017] for existence).

Then, for each $i \leq n$, $\tilde{U}_i := q(\nu_{r(i)}|X_i;\theta_2)$.

A Monte Carlo latent sample $\tilde{U}^{(n)}$ is designed to mimic the true sample $U^{(n)}$ of realizations of the latent variable in the sense that $\tilde{U}^{(n)}$ has the same distribution as $U^{(n)}$ conditionally on the sample of covariates $X^{(n)}$. The goal of the formal construction described in Definition 3 is to formalize exact size in finite samples. However, when actually conducting inference based on a single realization of the covariates $X^{(n)} = (X_1, \ldots, X_n)$, all the analyst needs are independent draws from $Q_{U|X_i;\theta_2}$, for each $i = 1, \ldots, n$.

2.1. **Test statistic.** Our test statistic is based on a sample analogue of the optimal transport problem, which characterizes the sharp identified region in Theorem 1. The

[7]When such a permutation is not unique, any will do.

sample analogue is based on the data sample $((Y_1, X_1), \ldots, (Y_n, X_n))$ combined with a Monte Carlo latent sample $\tilde{U}^{(n)}$. Our chosen test statistic $T_n(\theta)$ is the sample analogue of the optimal transport problem $\mathcal{D}(Q_{U|X;\theta_2} \times P_{X,0n}, P_{0n}; \theta_1)$. The latter is the discrete optimal transport problem

$$T_n(\theta) = \mathcal{D}_n(C(\theta)),$$

where:

(1) For any $n \times n$ cost matrix $C \in \mathcal{M}_n^+$, the program $\mathcal{D}$ is defined by

$$\mathcal{D}_n(C) := \min_{\pi \in \Pi_n} \sum_{i,j=1}^{n} \pi_{ij} C_{ij}; \tag{2.2}$$

where Π_n is the set of $n \times n$ non negative matrices π such that $\Sigma_i \pi_{ij} = \Sigma_j \pi_{ij} = 1/n$, for all $i, j \leq n$, as defined in the *notations and preliminaries* section.

(2) The cost matrix $C(\theta)$ has entries

$$C_{ij}(\theta) \;\; := \;\; d((\tilde{U}_i, X_i), \Gamma_u(Y_j, X_j; \theta_1)), \text{ for each } i, j \leq n, \tag{2.3}$$

where $\tilde{U}^{(n)} := (\tilde{U}_1, \ldots, \tilde{U}_n)$ is a Monte Carlo latent sample as in Definition 3.

Computation of the test statistic is discussed in Section 2.3 below. For now, note that (2.2) solves a discrete optimal transport problem, which is a special kind of linear programming problem.

2.2. **Critical values.** To achieve valid coverage of the true parameter with Confidence region CR_n, we choose as critical value $c_{n,1-\alpha}(\theta)$, the $1 - \alpha$ quantile of a distribution that first order stochastically dominates $T_n(\theta)$ for each $n \geq 1$. We then show exact coverage by exhibiting a data generating process in $\mathcal{P}_\theta^{(n)}$ such that $T_n(\theta)$ has $1 - \alpha$ quantile $c_{n,1-\alpha}(\theta)$.

Let $\tilde{U}'^{(n)}$ be a Monte Carlo latent sample independent of $\tilde{U}^{(n)}$, conditionally on $X^{(n)}$. The critical value we propose is the $1 - \alpha$ quantile $c_{n,1-\alpha}(\theta)$ of the distribution of

$$\tilde{T}_n(\theta) \;\; = \;\; \sup_{C \in \mathcal{C}_\theta(\tilde{U}'^{(n)})} \mathcal{D}_n(C), \tag{2.4}$$

where the supremum is taken over the class $\mathcal{C}_\theta(\tilde{U}'^{(n)})$ of $n \times n$ cost matrices with elements C_{ij} satisfying

$$C_{ij} = d((\tilde{U}_i, X_i), \Gamma_u(y, X_j; \theta_1)), \text{ where } (y, X_j) \in \Gamma_y(\tilde{U}'_j, X_j; \theta_1), \tag{2.5}$$

for some $y \in \mathcal{Y}$. The next theorem shows that $\tilde{T}_n(\theta)$ satisfies the desired requirements: it first order stochastically dominates the test statistic $T_n(\theta)$, and we can construct a data generating process under which both $T_n(\theta)$ and $\tilde{T}_n(\theta)$ have the same distribution. Hence, our proposed confidence region has the correct coverage probability in finite samples.

Theorem 2. *For all $\theta \in \Theta$, all $\alpha \in (0,1)$ and all $n \in \mathbb{N}$ such that $\mathcal{P}_\theta^{(n)}$ is non empty, Confidence region CR_n defined in (2.1) has correct coverage probability, i.e., for all realizations $v^{(n)}$ of $\mathcal{V}^{(n)} := (\nu_1, \ldots, \nu_n)$ in Definition 3,*

$$\inf_{P^{(n)} \in \mathcal{P}_\theta^{(n)}} P^{(n)}\left(T_n(\theta) \leq c_{n,1-\alpha}(\theta) \mid v^{(n)}\right) \quad \geq \quad 1 - \alpha, \tag{2.6}$$

with equality if the cumulative distribution function of $\tilde{T}_n(\theta)$ is continuous and increasing in a neighborhood of $c_{n,1-\alpha}(\theta)$.

The formal proof of Theorem 2 is given in the appendix. Proof heuristics are as follows. By construction, the Monte Carlo latent sample $\tilde{U}'^{(n)}$ has the same distribution as the true latent sample $U^{(n)} := (U_1, \ldots, U_n)$ conditional on the sample of covariates $X^{(n)} = (X_1, \ldots, X_n)$. Now, if the true data generating process $P_0^{(n)}$ is in $\mathcal{P}_\theta^{(n)}$, then each realization (Y_j, X_j), $j \leq n$, falls in $\Gamma_y(U_j, X_j; \theta_1)$ almost surely (according to the support restriction in the model). Hence, the cost matrix $C(\theta)$ in (2.3) belongs to $\mathcal{C}_\theta(U^{(n)})$. Hence the test statistic $T_n(\theta)$ is smaller than $\sup\left\{\mathcal{D}_n(C) : C \in \mathcal{C}_\theta(U^{(n)})\right\}$. Since the latter is identically distributed to $\sup\left\{\mathcal{D}_n(C) : C \in \mathcal{C}_\theta(\tilde{U}'^{(n)})\right\} = \tilde{T}_n(\theta)$, size control follows. To see that the inequality in (2.6) is an equality, we find $(Y^{(n)}, X^{(n)})$ that achieves the maximum of $T_n(\theta)$ under the constraint $(Y_i, X_i) \in \Gamma_y(\tilde{U}'_i, X_i; \theta_1)$. The formal construction of $\tilde{U}^{(n)}$ in Definition 3 is designed to ensure that such $Y^{(n)}$ is exchangeable, hence has identical marginals, as required by the definition of $\mathcal{P}_\theta^n$.

The following section discusses the simulation procedure we propose to approximate the critical values.

2.3. **Numerical implementation.** Computation of the test statistic requires computing cost matrix (2.3), and solving optimization problem (2.2). The former requires computing the distance between $(\tilde{U}_i, X_i)$ and region $\Gamma_u(Y_i, X_i; \theta_1)$, the computational complexity of which is model specific. The latter is a discrete optimal transport problem, which is a special kind of linear programming problem. There is a large literature on its implementation, reviewed in part in Peyré and Cuturi [2019]. Discrete optimal transport problems are equivalent to *assignment* problems, for which many efficient algorithms exist in the literature, most notably the *auction algorithm* (Bertsekas [1988]) and the *Hungarian algorithm* (see for instance Section 11.2 of Papadimitriou and Steiglitz [1998]), with $O(n^3)$ computational complexity. It can also be viewed as a network flow problem, for which efficient algorithms are available (see for instance Chapter 6 of Papadimitriou and Steiglitz [1998]). Efficient ready-to-use implementations abound. For example, The R implementation of the Hungarian algorithm from the package "transport" by Schuhmacher et al. [2020] performs optimal matching of two samples with size 1,000 (resp. 10,000) each in 0.1 (resp. 28) seconds on a standard 2020 MacBookAir.

The generic simulation procedure to compute Critical value $c_{n,1-\alpha}$ is the following.

(1) Generate S independent Monte Carlo latent samples $\tilde{U}^{(s)} := (\tilde{U}^s_j)_{j\leq n}$.
(2) For each $s \in \{1, \dots, S\}$, compute

$$\tilde{T}^s_n(\theta) = \sup_{C \in \mathcal{C}_\theta(\tilde{U}^{(s)})} \mathcal{D}_n(C),$$

and let $\tilde{T}^{(s)}_n(\theta)$, $s = 1, \dots, n$, be the order statistics.
(3) The critical value $c_{n,1-\alpha}(\theta)$ is approximated with

$$\hat{c}_{n,1-\alpha}(\theta) := \tilde{T}^{(\lceil S(1-\alpha)\rceil)}_n(\theta).$$

In practice, test statistic $\tilde{T}_n(\theta)$ may be costly to compute. We propose an alternative $\tilde{T}'_n(\theta)$ with critical value $c'_{n,1-\alpha}(\theta)$ that satisfies three desiderata. (1) It can be computed efficiently, (2) It is equal to $\tilde{T}_n(\theta)$ under suitable assumptions, (3) it still provides valid coverage (but may be conservative) if the latter assumptions fail.

On Step (2), we replace $\tilde{T}_n^s(\theta)$ with

$$\begin{aligned}\tilde{T}_n'^s(\theta) &= \min_{\pi\in\Pi_n} \max_{C\in\mathcal{C}_\theta(\tilde{U}^{(s)})} \sum_{i,j} \pi_{ij} C_{ij} \\ &= \sup_{C\in \operatorname{co}\mathcal{C}_\theta(\tilde{U}^{(s)})} \mathcal{D}_n(C), \end{aligned} \tag{2.7}$$

which is obtained from $\tilde{T}_n^s(\theta)$ by exchanging the order of the min and the max. Desideratum (3) follows immediately, since $\min\max \geq \max\min$. Desideratum (2) is fulfilled since $\tilde{T}_n'^s(\theta)$ and $\tilde{T}_n^s(\theta)$ are identical when the set $\mathcal{C}_\theta(\tilde{U}'^{(n)})$ is convex. Finally, Desideratum (1) is fulfilled since $\tilde{T}_n'^s(\theta)$ is the maximizer of a concave function, namely $\mathcal{D}_n(C)$, on a convex set, namely $\operatorname{co}\mathcal{C}_\theta(\tilde{U}^{(s)})$.

We therefore propose the following algorithm to check if a parameter value θ is in the $1-\alpha$ level confidence region CR_n.

(1) Generate S independent Monte Carlo latent samples $\tilde{U}^{(s)} := (\tilde{U}_j^s)_{j\leq n}$.
(2) For each $s \in \{1,\dots,S\}$, compute $\tau_s := 1\{T_n(\theta) \leq T_n'^s(\theta)\}$.
(3) Add θ to CR_n if and only if $\Sigma_s \tau_s/S \geq 1-\alpha$.

We perform Step (2) with a variation on Algorithm 1 in Dhouib et al. [2020]. This algorithm consists in a sequence of linear programing problems and converges from below in a finite number of steps. Since it converges from below, Step (2) does not require computation of $T_n'^s(\theta)$, because τ_s is known to be equal to 1 as soon as Algorithm 1 in Dhouib et al. [2020] returns a value larger than the test statistic $T_n(\theta)$.

3. Fast preliminary search in the parameter space

When the dimension of the parameter is large and there is no information about the geometry of the sharp identified region, a major computational hurdle is the search in the parameter space. This computational hurdle is common to all existing inference procedures for incomplete structural models, where the confidence region is based on inverting a test. To reduce the computational burden, we propose a conservative modification of our test, which relies on parameter free critical values. This allows a fast initial search in the parameter space and and what amounts to a dramatic reduction of the search area in our Monte Carlo simulations.

To construct an outer confidence region based on parameter free critical values, we need to reformulate the model in such a way that the unobserved variable U^* in the reformulation has fixed distribution Q_U^* with support $\mathcal{U}^*$. The basic ingredient in the reformulation is a transformation of the vector of unobservable variables U. We fix the distribution Q_U^* on $\mathcal{U}^*$ and make the following assumption.

Assumption 1. *There is a function h on $\mathcal{U}^* \times \mathcal{X} \times \Theta$ such that for any U^* with distribution Q_U^* on $\mathcal{U}^*$, the random vector $U := h(U^*, X; \theta_2)$ has distribution $Q_{U|X;\theta_2}$.*

Although we state it as an assumption, the function h in Assumption 1 always exists. When U is scalar, the conditional quantile transform is an example of such a function h. More generally, the vector quantile of U conditional on X, as defined in Chernozhukov et al. [2017] is an example of such a function h. It can also be computed as the solution of an optimal transport problem. However, simpler transformations often satisfy Assumption 1. For instance, in Example 1.2.2, $Q_{U|X;\theta_2}$ is a multivariate normal with mean zero and variance covariance matrix Σ. In that case, we can simply let Q_U^* be the standard multivariate normal and h be defined by $U = \Sigma^{\frac{1}{2}} U^*$.

Under Assumption 1, the incomplete model can be reformulated as the combination of the support constraint $(Y, X, U^*) \in \Gamma^*(\theta)$, where

$$\Gamma^*(\theta) \;\; := \;\; \left\{(y, x, u^*) : \, (y, x, h(u^*, x; \theta_2)) \in \Gamma(\theta_1)\right\},$$

and the marginal constraint $U^* \sim Q_U^*$ and $U^* \perp\!\!\!\perp X$. The metric d is replaced with a metric d^* on $(\mathcal{U}^* \times \mathcal{X}) \times (\mathcal{U}^* \times \mathcal{X})$. Statistics $T_n^*(\theta)$ and $\tilde{T}_n^*(\theta)$ and critical value $c_{n,1-\alpha}^*(\theta)$ are obtained with the same procedure as $T_n(\theta)$, $\tilde{T}_n(\theta)$, and $c_{n,1-\alpha}(\theta)$ respectively, with $\Gamma^*(\theta)$ replacing $\Gamma(\theta_1)$ and Q_U^* replacing $Q_{U|X;\theta_2}$. Correspondences Γ_u^* and Γ_y^* are obtained from Γ^* in the same way Γ_u and Γ_y are obtained from Γ in (1.5). Monte Carlo latent samples are generated in the same way, except that $Q_{U|X;\theta_2}$ is replaced with Q_U^*. Finally CR_n^* is the set of parameters θ such that $T_n^*(\theta)$ is smaller than or equal to $c_{n,1-\alpha}^*(\theta)$.

The outer confidence region

$$CR_n^0 \;\; := \;\; \{\theta \in \Theta : \; T_n^*(\theta) \leq c_{n,1-\alpha}^0\}$$

is defined with test statistic $T_n^*(\theta)$ and parameter free critical value $c_{n,1-\alpha}^0$. Our conservative parameter free critical value $c_{n,1-\alpha}^0$ is chosen to be the $1-\alpha$ quantile of the distribution of

$$\tilde{T}_n^0 = \mathcal{D}_n(\tilde{C}^0), \quad \text{with} \quad \tilde{C}_{ij}^0 = d((U_i^*, X_i), (\tilde{U}_j^*, X_j)),$$

where $(U_i^*)_{i\le n}$ and $(\tilde{U}_j^*)_{j\le n}$ are two independent Monte Carlo latent samples, simulated according to Q_U^*.

By construction, for any y such that $(y, X_j) \in \Gamma_y^*(\tilde{U}_j^*, X_j; \theta)$, we have $(\tilde{U}_j^*, X_j) \in \Gamma_u^*(y, X_i; \theta)$. Hence $d((U_i^*, X_i), \Gamma_u^*(y, X_j; \theta)) \le d((U_i^*, X_i), (\tilde{U}_j^*, X_j))$. It follows that by construction, for all $\theta \in \Theta$, $\tilde{T}_n^*(\theta) \le \tilde{T}_n^0$, and, therefore:

$$\sup_{\theta\in\Theta} c_{n,1-\alpha}^*(\theta) \le c_{n,1-\alpha}^0 \quad \text{and} \quad CR_n^* \subseteq CR_n^0\,. \tag{3.1}$$

From Statement (3.1), we deduce three advantages of the outer confidence region CR_n^0. First, the critical value is independent of the parameter value. Hence, it needs to be computed only once, and only the test statistic $T_n^*(\theta)$ needs to be computed for each value of the parameter θ. Second, the outer confidence region CR_n^0 covers the whole identified set as opposed to each value in the identified set[8]. Third, given that $CR_n^* \subseteq CR_n^0$, the computation of Confidence region CR_n^* can be performed with a search limited to CR_n^0 as opposed to the whole parameter space Θ.

4. Consistency

In this section, we theoretically assess informativeness of the confidence region, as sample size increases. We characterize sequences of data generating processes and parameters that violate the model, and show that such parameter sequences are outside the confidence region, eventually. We prove this consistency result for the conservative outer region CR_n^0. Since the latter includes our proposed confidence region CR_n^*, the result also holds for CR_n^*.

Let $(P_0^{(n)})_{n\ge1}$ be a sequence of data generating processes. Let $(Y_{i,n}, X_{i,n})_{i\le n}$ be a triangular array where, for any $n \ge 1$, the size n sample $(Y_{i,n}, X_{i,n})_{i\le n}$ follows

[8]See Section 4.3.1 of Molinari [2020] for a discussion of the distinction between coverage of the identified set and coverage of each of its elements.

distribution $P_0^{(n)}$. For each n, $P_0^{(n)}$ has identical marginals $P_{0n} := P_{Y|X,0n} \times P_{X,0n}$. We consider parameters θ that violate the condition that characterizes the sharp identified region in Theorem 1. Formally, the alternative is defined as follows, where $\mathcal{D}$ is defined as in (1.6).

Assumption 2 (Sequence of alternatives)**.** *Parameter θ satisfies*

$$\liminf_{n\to\infty} \ \mathcal{D}(Q_U^* \times P_{X,0n}, P_{0n}; \theta) > 0. \tag{4.1}$$

In order to detect violations defined in Assumption 2, or equivalently, to make sure such a parameter ultimately falls outside the confidence region, the data sequence must be sufficiently informative to identify the marginal distributions P_{0n}. Independence across observations is sufficient, but not necessary, as any dependence structure that allows estimation of P_{0n} from the sequence of empirical distributions $\hat{P}_n := \Sigma_{i\leq n}\delta_{(Y_i,X_i)}/n$ is suitable.

Assumption 3 (Data generating process)**.** *The sequence of data generating processes $P_0^{(n)}$ with marginals $P_{0n} := P_{Y|X,0n} \times P_{X,0n}$ is such that $\{P_{0n} : n \geq 1\}$ is tight, and $d(\hat{P}_n, P_{0n}) \to 0$ almost surely for any distance d that metrizes weak convergence.*

Detection of violations of the type (4.1) also requires continuity of the cost function in the optimal transport problem.

Assumption 4 (Regularity of the structure)**.** *Metric d^* is continuous on $(\mathcal{U}^* \times \mathcal{X}) \times (\mathcal{U}^* \times \mathcal{X})$, and Function $((u^*, x), (y, x')) \mapsto d((u^*, x), \Gamma_u^*((y, x'); \theta))$ is continuous on $(\mathcal{U}^* \times \mathcal{X}) \times (\mathcal{Y} \times \mathcal{X})$.*

The condition is stated in its most general form. However, sufficient conditions on the model structure can be derived. For instance, by Lemma 16.30 page 538 of Aliprantis and Border [1999], Assumption 4 holds if Γ_u^* is a continuous correspondence (i.e., both upper- and lower-hemicontinuous) with non empty and compact values.

Theorem 3 (Consistency)**.** *Under Assumptions 1, 2, 3, and 4, for all $\alpha \in (0, 1)$, for almost all sequences $v^{(n)}$, $n \in \mathbb{N}$ of realizations of $\mathcal{V}^{(n)} := (\nu_1, \ldots, \nu_n)$, in Definition 3,*

$$\liminf_{n\to\infty} \ P_0^{(n)} \left(\, T_n^*(\theta) > c^0_{n,1-\alpha} \mid v^{(n)} \, \right) = 1.$$

Given that, by (3.1), the exact critical value $c^*_{n,1-\alpha}$ is uniformly smaller than the conservative critical value $c^0_{n,1-\alpha}$, Theorem 3 also implies that Parameter θ defined in Assumption 2 eventually falls outside the confidence region CR^*_n.

5. Simulation evidence

We derive confidence regions and coverage probabilities in a simple entry game example. Consider the classic entry game model of Bresnahan and Reiss [1991] and Tamer [2003] with S players. Each player s can choose a binary action $Y_s \in \{0,1\}$. The payoff u_s of player s is

$$u_s = 1(Y_s = 1)\left(X_s'\beta - \sum_{s'\neq s} \delta Y_{s'} + \epsilon_s\right),$$

where X_s is a vector of player specific covariates and ϵ_s is a random shock that is known to the players but unknown to the researcher. If $\delta \geq 0$, then the game always has at least one pure-strategy Nash equilibrium. Assume the observed decision profile $Y = (Y_s)_{s\leq S}$ is a pure strategy Nash equilibrium. The vector of random shocks $U = (\epsilon_1,\dots,\epsilon_S)$ follows a standard multivariate normal distribution. The parameter vector is $\theta = (\beta,\delta)$, and the support restriction is $(Y,X,U) \in \Gamma(\theta)$, almost surely, where $(y,x,u) \in \Gamma(\theta)$ if and only if

$$\forall s \in \{1,...,S\}, \quad (-1)^{y_s}\left(x_s'\beta - \sum_{s'\neq s} \delta y_{s'} + \epsilon_s\right) \leq 0.$$

In our simulations, the number of players is $S = 6$, $X_s = (1\ X_{1s})$, with $X_{1s} \sim N(0,1)$ for each s, and $\beta = (\beta_0\ \beta_1)$, with $\beta_0 = 0.6$, $\beta_1 = 0.6$. Finally, $\delta = 0.3$.

Table 1 gives coverage probabilities for the test of $H_0 : (\beta_0,\beta_1,\delta) = (0.6,0.6,0.3)$ with confidence levels 0.9, 0.95 and 0.99, sample sizes $n \in \{50,100,150,200,500,1000\}$, based on $5,000$ replications. For each replication, the true data generating process is chosen by selecting the equilibrium such that the value of the test statistic is maximized. Column "ncx" (which stands for "non convexified") reports coverage probabilities based on quantiles of $\tilde{T}_n(\theta)$ in (2.4) as critical values. This column shows exact coverage for small samples, which conforms with Theorem 2. For sample size $1,000$, to reduce computational time, the running time for each replication is capped at 10^3s,

Table 1. Coverage probabilities: Column "ncx" reports coverage probabilities based on quantiles of $\tilde{T}_n(\theta)$ as critical values. Column "cx" reports coverage probabilities based on quantiles of $\tilde{T}'_n(\theta)$ as critical values. Columns "Time" report total testing time in seconds for all $5,000$ replications.

	Confidence level						Time	
	0.90		0.95		0.99			
Sample size	ncx	cx	ncx	cx	ncx	cx	ncx	cx
50	0.898	0.901	0.946	0.949	0.988	0.989	81s	79s
100	0.905	0.912	0.956	0.960	0.992	0.992	78s	77s
150	0.904	0.914	0.952	0.957	0.991	9.992	92s	82s
200	0.904	0.918	0.954	0.963	0.990	0.991	124s	109s
500	0.907	0.931	0.954	0.968	0.991	0.994	10^4s	548s
1,000	0.948	0.954	0.980	0.982	0.996	0.997	10^5s	10^3s

and the number of iterations to $100,000$, which produces coverage probabilities that are higher than nominal confidence level. Given that the algorithm converges from above, validity is not affected when the running time is capped. Column "cx" (which stands for "convexified") reports coverage probabilities based on quantiles of $\tilde{T}'_n(\theta)$ in (2.7) as critical values. This column shows exact coverage for very small samples. For larger samples, the difference between $\tilde{T}_n(\theta)$ and $\tilde{T}'_n(\theta)$ is detected, and coverage probability exceeds nominal level. Column "time" reports total time for all $5,000$ replications, when running on a server with 2 AMD EPYC 7702 processors with 128 CPU cores in total. Time involved in simulating samples is subtracted from the reported number. In the case of $n = 1,000$, the ratio of simulating time to testing time is approximately 10 to 1.

Figure 1 shows what confidence region CR_n looks like with one simulated sample for each sample sizes $n \in \{100, 500, 1000, 5000\}$. The orange area is the confidence region based on the exact test, and the blue area is the outer region CR_n^0. In each case, the true data generating process is chosen by selecting randomly within the set of multiple equilibrium, when they arise.

Discussion

We have proposed a procedure to compute confidence regions in incomplete models with exact coverage in finite samples. Compared to existing approaches, our procedure has many advantages, some straightforward and others more subtle. First, finite sample validity avoids reliance on asymptotic approximations, which are often suspect. It also removes the need for user-chosen tuning parameters, that inference results are often very sensitive to. Second, our procedure removes the need for transforming conditional into unconditional moment inequalities, and for reducing the very large number of moment inequalities with complex and model-specific core determining classes. Third, finite sample validity allows us to conduct inference in models, where the specification depends on the sample size. This is particularly important in applications to games on networks and network formation games, when a single network is observed. In such cases, the support constraint in the model specification depends on the sample size, and so does the dimension of the latent variable, which involves an individual's neighbors in the network. Finally, although we haven't developed it here, our method extends to specifications, where the structural support constraint is individual-specific, thereby allowing us to conduct inference with the structural vector autoregressions proposed in Giacomini and Kitagawa [2021] and Giacomini et al. [2021].

This paper has contributed to a growing literature that shows how optimal transport theory provides a rich set of tools in econometrics in general, and incomplete models in particular. We expect these tools to underlay the development of a subvector version of our inference method, as well as an extension to semiparametric incomplete models with independence constraints. The latter is well under way, the former still more speculative.

Appendix A. Proofs of results in the main text

Proof of Theorem 1. Let the true data generating process $P_0^{(n)}$ have n identical and independent marginals P_{0n}. Call $\tilde{\Theta}_I^{(n)}$ the region defined on the right-hand side of (1.6). First show that $\Theta_I^{(n)} \subseteq \tilde{\Theta}_I^{(n)}$. If $\theta \in \Theta_I^{(n)}$, then there exists a joint probability $\tilde{\pi}$ over $\mathcal{Y} \times \mathcal{X} \times \mathcal{U}$ with marginals P_{0n} and $Q_{U|X;\theta_2}$, such that $\mathbb{E}_{\tilde{\pi}} 1\{(Y,X,U) \notin$

$\Gamma(\theta_1)\} = 0$. The latter implies the existence of a probability π on $\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X}$, which is in $\mathcal{M}(Q_{U|X;\theta_2} \times P_{X,0n}, P_{0n})$, $\pi(X = X') = 1$, and such that $\mathbb{E}_\pi 1\{(U,X) \notin \Gamma_u(Y,X';\theta_1)\} = 0$. This implies $\mathbb{E}_\pi d\left((U,X), \Gamma_u(Y,X';\theta_1)\right) = 0$. Hence, $\theta \in \tilde{\Theta}_I^{(n)}$. Now show that $\tilde{\Theta}_I \subseteq \Theta_I$. Since d is a metric and Γ_u is closed-valued, if $\theta \in \tilde{\Theta}_I$, then there exists a random vector (U,X,Y,X') such that (U,X) has distribution $Q_{U|X;\theta_2} \times P_{X,0n}$, (Y,X) has distribution P_{0n}, $X' = X$ and $(Y,X,U) \in \Gamma(\theta_1)$ almost surely. Given any sample $(Y_i,X_i)_{i\leq n}$ distributed according to $P_0^{(n)}$, construct $(U_i)_{i\leq n}$ as follows: Conditional on $(Y_i,X_i)_{i\leq n}$, draw $(U_i)_{i\leq n}$ from $\pi^*_{U|(Y,X)=(Y_1,X_1)} \times \cdots \times \pi^*_{U|(Y,X)=(Y_n,X_n)}$. Then, $(Y_i,X_i,U_i)_{i\leq n}$ satisfies all the conditions of Definition 1 and $\theta \in \Theta_I^{(n)}$.

□

Proof of Theorem 2. Throughout the proof, the realization $v^{(n)}$ of $\mathcal{V}^{(n)} = (\nu_1, \ldots, \nu_n)$ from Definition 3 is fixed. We also fix and arbitrary θ such that $\mathcal{P}_\theta^{(n)}$ is non empty and an arbitrary $\alpha \in (0,1)$.

Proof of (2.6). Take an arbitrary distribution $P^{(n)}$ in $\mathcal{P}_\theta^{(n)}$, and let $(Y^{(n)}, X^{(n)})$ be a random vector distributed according to $P^{(n)}$. By Definition 3, the Monte Carlo latent sample $\tilde{U}^{(n)}$ is determined by $X^{(n)}$ and $v^{(n)}$. Let $T_n(\theta)$ be the resulting test statistic. By the definition of $\mathcal{P}_\theta^{(n)}$, there exists a random vector $U^{(n)}$ such that $(Y_i,X_i,U_i) \in \Gamma(\theta_1)$ and $U_i|X_i \sim Q_{U|X_i;\theta_2}$ almost surely for each i. Because $(Y_i,X_i) \in \Gamma_y(U_i,X_i;\theta_1)$, we know that the cost matrix $C(\theta)$ defined in (2.3), which enters the test statistic $T_n(\theta)$, belongs to the set $\mathcal{C}_\theta(U^{(n)})$ of cost matrices defined as in (2.5). Therefore,

$$T_n(\theta) = \mathcal{D}_n(C(\theta)) \leq \sup_{C \in \mathcal{C}_\theta(U^{(n)})} \mathcal{D}_n(C). \tag{A.1}$$

By Definition 3, $(X^{(n)}, U^{(n)})$ and $(X^{(n)}, \tilde{U}'^{(n)})$ are identically distributed. Hence, the $1-\alpha$ quantile $c_{n,1-\alpha}(\theta)$ of $\sup\{\mathcal{D}_n(C) : C \in \mathcal{C}_\theta(\tilde{U}'^{(n)})\}$ is also the $1-\alpha$ quantile of $\sup\{\mathcal{D}_n(C) : C \in \mathcal{C}_\theta(U^{(n)})\}$, so that (2.6) follows from (A.1).

Proof that (2.6) holds as an equality. Fix $\epsilon > 0$. We show below that for any $\beta \in (0,1)$, there exists some $P^{(n)} \in \mathcal{P}_\theta^{(n)}$ such that

$$P^{(n)}(T_n(\theta) \leq c_{n,1-\beta}(\theta) - \epsilon \mid \mathcal{V}^{(n)} = v^{(n)}) \leq 1 - \beta. \tag{A.2}$$

Suppose the cdf of $\tilde{T}_n(\theta)$ is continuous and increasing in a neighborhood of $c_{n,1-\alpha}(\theta)$. For any small enough $\eta > 0$, $c_{n,1-\alpha+\eta}(\theta) - c_{n,1-\alpha} > 0$. Let $\epsilon = c_{n,1-\alpha+\eta}(\theta) - c_{n,1-\alpha}(\theta)$. Then, (A.2) applied to $\beta = \alpha - \eta$ implies that there exists some $P^{(n)} \in \mathcal{P}_\theta^{(n)}$ such that

$$\begin{aligned} P^{(n)}(T_n(\theta) \leq c_{n,1-\alpha}(\theta) \mid \mathcal{V}^{(n)} = v^{(n)}) &= P^{(n)}(T_n(\theta) \leq c_{n,1-\alpha+\eta}(\theta) - \epsilon \mid \mathcal{V}^{(n)} = v^{(n)}) \\ &\leq 1 - \alpha + \eta. \end{aligned}$$

The above inequality holds for arbitrary small $\eta > 0$, and the result follows.

Proof of (A.2). By assumption, $\mathcal{P}_\theta^{(n)}$ is nonempty under the null hypothesis. Hence, there exists a marginal distribution $P_{X,n}$ such that $\Gamma_y(U, X; \theta_1)$ is almost surely non-empty if $X \sim P_{X,n}$ and $U|X \sim Q_{U|X;\theta_2}$. Let $(X^{(n)}, U^{(n)})$ be a vector of n i.i.d. draws from $P_{X,n} \times Q_{U|X;\theta_2}$. Write $X^{(n)} = (X_1, \ldots, X_n)$ and $U^{(n)} := (U_1, \ldots, U_n)$.

We will construct a map $\varphi : \mathcal{X}^n \times \mathcal{U}^n \to \mathcal{Y}^n$ such that the distribution $P^{(n)}$ of $(\varphi(X^{(n)}, U^{(n)}), X^{(n)}, U^{(n)})$ is in $\mathcal{P}_\theta^{(n)}$ and satisfies (A.2). Note that restriction (2) in the definition of $\mathcal{P}_\theta^{(n)}$ (Definition 1) is satisfied by the construction of $(X^{(n)}, U^{(n)})$.

In addition, the map φ we construct must satisfy the following.

(i) It must be measurable. To show this, we will rely on a classical theorem on the existence of measurable selections of correspondences, namely Theorem 17.40 page 184 of Bertsekas and Shreve [1996].

(ii) It must be a selection from the correspondence

$$\mathcal{Y}^{(n)}(X^{(n)}, U^{(n)}) := \{(y_1, \ldots, y_n) \in \mathcal{Y}^n : \forall j, (y_j, X_j, U_j) \in \Gamma(\theta_1)\},$$

so support restriction (1) in the definition of $\mathcal{P}_\theta^{(n)}$ (Definition 1) is satisfied. This will be imposed in the construction.

(iii) The distribution of $(\varphi(X^{(n)}, U^{(n)}), X^{(n)}, U^{(n)})$ must have identical marginals, for each observation unit $i = 1, \ldots, n$. We will construct φ in such a way that distribution of $(\varphi(X^{(n)}, U^{(n)}), X^{(n)}, U^{(n)})$ is exchangeable, hence has identical marginals.

(iv) The distribution $P^{(n)}$ of $(\varphi(X^{(n)},U^{(n)}),X^{(n)})$ must satisfy (A.2). By definition of $T_n(\theta)$ and $\tilde{T}_n(\theta)$, the latter is satisfied if $Y^{(n)} := \varphi(X^{(n)},U^{(n)})$ satisfies

$$\mathcal{D}_n(C(Y^{(n)},X^{(n)};\theta) \geq \sup_{C\in\mathcal{C}_\theta(U^{(n)})} \mathcal{D}_n(C) \; - \; \epsilon. \tag{A.3}$$

In the display above, $\mathcal{C}_\theta(U^{(n)})$ is defined as in (2.5), $X^{(n)}$ and $v^{(n)}$ determine the Monte Carlo latent sample $\tilde{U}^{(n)}$ according to Definition 3, and $C(Y^{(n)},X^{(n)};\theta)$ is the cost matrix with (i,j)th component $d((\tilde{U}_i,X_i),\Gamma_u(Y_j,X_j;\theta_1))$.

By the definition of $\mathcal{Y}^{(n)}(X^{(n)},U^{(n)})$, we have:

$$\sup_{C\in\mathcal{C}_\theta(U^{(n)})} \mathcal{D}_n(C) \;=\; \sup_{y^{(n)}\in\mathcal{Y}^{(n)}(X^{(n)},U^{(n)})} \mathcal{D}_n(C(y^{(n)},X^{(n)};\theta)) \;<\; \infty.$$

Thus (A.3) is equivalent to

$$\mathcal{D}_n(C(y^{(n)},X^{(n)};\theta) \geq \sup_{\tilde{y}^{(n)}\in\mathcal{Y}^{(n)}(X^{(n)},U^{(n)})} \mathcal{D}_n(C(\tilde{y}^{(n)},X^{(n)};\theta) \; - \; \epsilon. \tag{A.4}$$

Define the correspondence $\Phi : \mathcal{X}^n \times \mathcal{U}^n \rightrightarrows \mathcal{Y}^n$ by

$$\Phi\left(X^{(n)},U^{(n)}\right) \;:=\; \left\{y^{(n)} \in \mathcal{Y}^{(n)}(X^{(n)},U^{(n)}) : \text{ (A.4) holds}\right\}.$$

We fulfill requirements (i), (ii), (iii) and (iv) by showing that Φ admits a measurable selection φ such that $(\varphi(X^{(n)},U^{(n)}),X^{(n)},U^{(n)})$ is exchangeable. The rest of the proof is devoted to proving this fact.

Existence of φ. Consider any element $(x^{(n)},u^{(n)})$ of $\mathcal{X}^n\times\mathcal{U}^n$ such that $\Gamma_y(u_j,x_j;\theta_1)\neq\varnothing$ for each $j\leq n$, and such that $x_1 \preccurlyeq \ldots \preccurlyeq x_n$, where $\preccurlyeq$ is the order of Definition 3. For any permutation σ of $\{1,\ldots,n\}$, call $(\mathcal{X}^n\times\mathcal{U}^n)_\sigma$ the subset of $\mathcal{X}^n\times\mathcal{U}^n$ with elements $(x^{(n)},u^{(n)})$ such that $x_{\sigma(1)} \preccurlyeq \ldots \preccurlyeq x_{\sigma(n)}$. In particular, $(\mathcal{X}^n\times\mathcal{U}^n)_{id}$ is the subset of $\mathcal{X}^n\times\mathcal{U}^n$ with elements $(x^{(n)},u^{(n)})$ such that $x_1 \preccurlyeq \ldots \preccurlyeq x_n$. By Theorem 17.40 page 184 of Bertsekas and Shreve [1996], the correspondence Φ admits a universally measurable selection φ on $(\mathcal{X}^n\times\mathcal{U}^n)_{id}$.

Extend φ to the whole of $\mathcal{X}^n\times\mathcal{U}^n$ by setting $\varphi(x^{\sigma(n)},u^{\sigma(n)}) := y^{\sigma(n)}$, for all permutations σ of $\{1,\ldots,n\}$, where $\xi^{\sigma(n)} = (\xi_{\sigma(1)},\ldots,\xi_{\sigma(n)})$, for $\xi\in\{y,x,u\}$. We show that φ is universally measurable. Let B be a Borel subset of $\mathcal{Y}^{(n)}(x^{(n)},u^{(n)})$. The set $\varphi^{-1}(B)$ is the finite union of sets $\varphi^{-1}(B) \cap (\mathcal{X}^n \times \mathcal{U}^n)_\sigma$ over all permutations σ of $\{1, \ldots, n\}$. It suffices to show that $\varphi^{-1}(B) \cap (\mathcal{X}^n \times \mathcal{U}^n)_\sigma$ is universally measurable for any permutations σ of $\{1, \ldots, n\}$. Define the permutation operator ς associated with permutation σ by $\varsigma(X^{(n)}, U^{(n)}) := (X^{\sigma(n)}, U^{\sigma(n)})$. By construction, we have

$$\varphi^{-1}(B) \cap (\mathcal{X}^n \times \mathcal{U}^n)_\sigma = \varsigma\left(\varphi^{-1}(B) \cap (\mathcal{X}^n \times \mathcal{U}^n)_{id}\right). \tag{A.5}$$

Since the right-hand-side of (A.5) is a universally measurable set, we conclude that φ is universally measurable.

Next, we have to show that the extension thus defined still selects a $y^{(n)}$ that is almost extremal, in the sense of $\varphi(x^{\sigma(n)}, u^{\sigma(n)}) \in \Phi(x^{\sigma(n)}, u^{\sigma(n)}))$. This follows from the following two facts. First, $y^{\sigma(n)}$ is in $\mathcal{Y}^{(n)}(x^{\sigma(n)}, u^{\sigma(n)})$ if and only if $y^{(n)}$ is in $\mathcal{Y}^{(n)}(x^{(n)}, u^{(n)})$, which follows from the definition of $\mathcal{Y}^{(n)}$. Second,

$$\mathcal{D}_n(C(y^{\sigma(n)}, x^{\sigma(n)}; \theta)) = \mathcal{D}_n(C(y^{(n)}, x^{(n)}; \theta)), \tag{A.6}$$

which we now show. Given the construction of $\tilde{U}^{(n)}$ in Definition 3, a permutation of indices of $x^{(n)}$ induces an identical permutation of indices of $\tilde{U}^{(n)}$. This ensures that the cost matrix $C(y^{\sigma(n)}, x^{\sigma(n)}; \theta)$ has components $d((\tilde{U}_{\sigma(i)}, x_{\sigma(i)}), \Gamma_u(y_{\sigma(j)}, x_{\sigma(j)}; \theta_1))$. Since the optimal transport problem $\mathcal{D}_n$ is invariant to permutations of indices, (A.6) follows. Therefore, $\varphi(x^{\sigma(n)}, u^{\sigma(n)}) \in \Phi(x^{\sigma(n)}, u^{\sigma(n)}))$ as desired.

Finally, we define $Y^{(n)} := \varphi(X^{(n)}, U^{(n)})$. By the construction of φ, $Y^{\sigma(n)} := \varphi(X^{\sigma(n)}, U^{\sigma(n)})$. We now show that it implies exchangeability of $Y^{(n)}$. Indeed, setting $W_i := (X_i, U_i)$ for i–$1, \ldots, n$, we have

$$\begin{aligned}
\mathbb{P}((Y_1, \cdots, Y_n) \leq (y_1, \cdots, y_n)) &= \mathbb{P}((Y_{\sigma(1)}, \ldots, Y_{\sigma(n)}) \leq (y_{\sigma(1)}, \ldots, y_{\sigma(n)})) \\
&= \mathbb{P}(\varphi(W_{\sigma(1)}, \ldots, W_{\sigma(n)}) \leq (y_{\sigma(1)}, \ldots, y_{\sigma(n)})) \\
&= \mathbb{P}(\varphi(W_1, \ldots, W_n) \leq (y_{\sigma(1)}, \ldots, y_{\sigma(n)})) \\
&= \mathbb{P}((Y_1, \ldots, Y_n) \leq (y_{\sigma(1)}, \ldots, y_{\sigma(n)})),
\end{aligned}$$

where the third equality follows from the fact that the distribution of $\{W_1, \ldots, W_n\}$ is exchangeable.

We have therefore proved that the distribution $P^{(n)}$ of $(Y^{(n)}, X^{(n)})$ is in $\mathcal{P}_\theta^{(n)}$ and satisfies (A.2) as desired. $\square$

Proof of Theorem 3. Fix a sequence $(v_i)_{i\geq 1}$ of realizations of ν_i in Definition 3. Call $(\tilde{u}_i)_{i\geq 1}$ the realizations of the corresponding Monte Carlo latent sample. For each $n \in \mathbb{N}$, write $v^{(n)} := (v_i)_{i\leq n}$ and call $Q^*_{U,n} := \Sigma_{i\leq n}\delta_{\tilde{u}_i}/n$ the empirical distribution associated with the Monte Carlo latent sample. Choose the sequence $(v_i)_{i\geq 1}$ such that $Q^*_{U,n}$ converges in distribution to Q^*_U. Fix an arbitrary realizations of the triangular array $(Y_{i,n}, X_{i,n})_{i\leq n}$, $n \in \mathbb{N}$. Let $\{n_k, k \in \mathbb{N}\}$, be a subsequence such that $\liminf_{n\to\infty} T^*_n(\theta) = \lim_{k\to\infty} T^*_{n_k}(\theta)$. Since $\{P_{0n} : n \geq 1\}$ is tight, we can extract a further subsequence, still denoted n_k, such that P_{0n_k} converges to some distribution $P^* := P^*_Y \times P^*_X$ as $k \to \infty$. Then, $P_{X,0n_k} \times Q^*_{U,n_k}$ converges to $P^*_X \times Q^*_U$. Because $(y, x) \mapsto \Gamma_u(y, x; \theta)$ is continuous, $d((\tilde{u}, \tilde{x}), \Gamma_u(y, x; \theta))$ is continuous in $(\tilde{u}, \tilde{x}, y, x)$. Hence, by Theorem 5.20 in Villani [2009], we have

$$\begin{aligned}\mathcal{D}(Q^*_U \times P^*_X, P^*; \theta) &= \lim_{k\to\infty} \mathcal{D}(Q^*_{U,n_k} \times P^*_{X,0n_k}, P^*_{0n_k}; \theta)\\ &\geq \liminf_{n\to\infty} \mathcal{D}(Q^*_{U,n} \times P^*_{X,0n}, P^*_{0n}; \theta) \; > \; 0.\end{aligned}$$

On the other hand, Assumption 3 implies that $\hat{P}_n$ also converges in distribution to P^* with probability 1. Hence, by Theorem 5.20 in Villani [2009], we also have

$$\lim_{n\to\infty} T^*_n(\theta) \;=\; \mathcal{D}(Q^*_U \times P^*_X, P^*; \theta) \; > \; 0.$$

There remains to show that $\lim_{n\to\infty} \tilde{T}^0_n = 0$. Indeed, by Theorem 5.20 in Villani [2009],

$$\lim_{n\to\infty} \tilde{T}^0_n \;=\; \min_{\pi\in\mathcal{M}(Q^*_U\times P^*_X, Q^*_U\times P^*_X)} \mathbb{E}_\pi \, d((U, X), (U', X')) \;=\; 0.$$

$\square$

References

C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Springer, second edition, 1999.

T. Anderson and H. Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20: 46–63, 1949.

D. Andrews and X. Shi. Inference based on conditional moment inequalities. *Econometrica*, 81:609–666, 2013.

G. Barnard. Comment on 'the spectral analysis of point processes' by M.S. Bartlett. *Journal of the Royal Statistical Society, Series B*, 25:294, 1963.

L. Barseghyan, M. Coughlin, F. Molinari, and J. C. Teitelbaum. Heterogeneous choice sets and preferences. forthcoming in *Econometrica*, 2021.

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

D. Bertsekas. The auction algorithm: a distributed relaxation method for the assignment problem. *Annals of Operations Research*, 14:105–123, 1988.

D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case*. Athena Scientific: Belmont, Massachusetts, 1996.

T. Bresnahan and P. Reiss. Entry and competition in concentrated markets. *Journal of Political Economy*, 99:977–1009, 1991.

I. Canay and A. Shaikh. Practical and theoretical advances for inference in partially identified models. In B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, editors, *Advances in Economics and Econometrics*, volume 2 of *Econometric Society Monographs*, pages 271–306. Cambridge University Press, 2018.

V. Chernozhukov, S. Lee, and A. Rosen. Inference on intersection bounds. *Econometrica*, 81:667–737, 2013.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45:223–256, 2017.

V. Chernozhukov, D. Chetverikov, and K. Kato. Inference on causal and structural parameters using many moment inequalities. *Review of Economic Studies*, 86: 1867–1900, 2019.

A. Chesher and A. Rosen. Generalized instrumental variable models. *Econometrica*, 85:959–989, 2017.

A. Chesher and A. Rosen. Generalized instrumental variable models, methods and applications. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, editors, *Handbook of Econometrics*, volume 7A, pages 1–110. Elsevier, 2020.

A. Chesher, A. Rosen, and K. Smolinski. An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4:157–196, 2013.

F. Ciliberto, C. Murry, and E. Tamer. Market structure and competition in airline markets. *Journal of Political Economy*, 129:2995–3038, 2021.

A. de Paula, S. Richards-Shubik, and E. Tamer. Identifying preferences in networks with bounded degree. *Econometrica*, 86:263–288, 2018.

S. Dhouib, I. Redko, T. Kerdoncuff, R. Emonet, and M. Sebban. A swiss army knife for minimax optimal transport. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7613–7622, 2020.

J.-M. Dufour. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133:443–477, 2006.

J.-M. Dufour and L. Khalaf. Monte Carlo test methods in econometrics. In B. Baltagi, editor, *A Companion to Econometric Theory*, pages 494–519. Blackwell, 2001.

M. Dwass. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 28:181–187, 1957.

I. Ekeland, A. Galichon, and M. Henry. Optimal transportation and the falsifiability of incompletely specified economic models. *Economic Theory*, 42:355–374, 2010.

L. Epstein, H. Kaido, and K. Seo. Robust confidence regions for incomplete models. *Econometrica*, 84:1799–1838, 2016.

A. Galichon and M. Henry. Inference in incomplete models. unpublished manuscript, 2006.

A. Galichon and M. Henry. Set identification in models with multiple equilibria. *Review of Economic Studies*, 78:1264–1298, 2011.

R. Giacomini and T. Kitagawa. Robust Bayesian inference for set-identified models. *Econometrica*, 89:1519–1556, 2021.

R. Giacomini, T. Kitagawa, and M. Reid. Robust Bayesian inference in proxy SVARs. *Journal of Econometrics*, 2021.

P. Haile and E. Tamer. Inference in an incomplete model of English auctions. *Journal of Political Economy*, 111:1–51, 2003.

B. Jovanovic. Observable implications of models with multiple equilibria. *Econometrica*, 57:1431–1437, 1989.

D. Kédagni, L. Li, and I. Mourifié. Discordant relaxations of misspecified models. unpublished manuscript, 2020.

Y. Luo and H. Wang. Core determining class and inequality selection. *American Economic Review, Papers & Proceedings*, 107:274–277, 2017.

F. Molinari. Microeconometrics with partial identification. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, editors, *Handbook of Econometrics*, volume 7A, pages 355–486. Elsevier, 2020.

C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization, Algorithms and Complexity.* Dover, second edition edition, 1998.

G. Peyré and M. Cuturi. *Computational Optimal Transport with Applications to Data Science*. NOW: Boston, 2019.

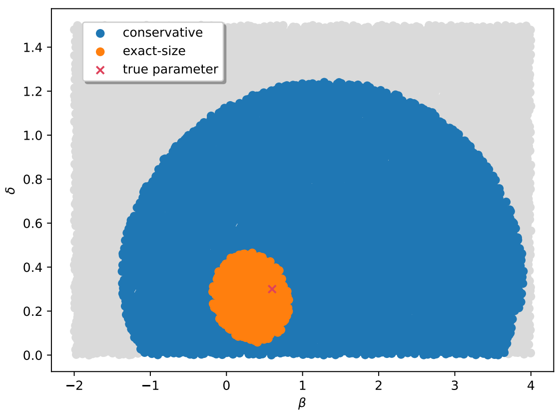
D. Schuhmacher, B. Bähre, C. Gottschlich, V. Hartmann, F. Heinemann, and B. Schmitzer. *Computation of Optimal Transport Plans and Wasserstein Distances*, R package version 0.12-2 edition, 2020.

E. Tamer. Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70:147–165, 2003.

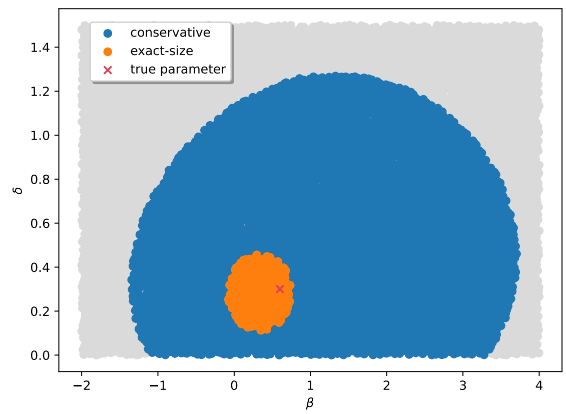
C. Villani. *Topics in Optimal Transportation*. Providence: American Mathematical Society, 2003.

C. Villani. *Optimal Transport: Old and New*. Springer, 2009.

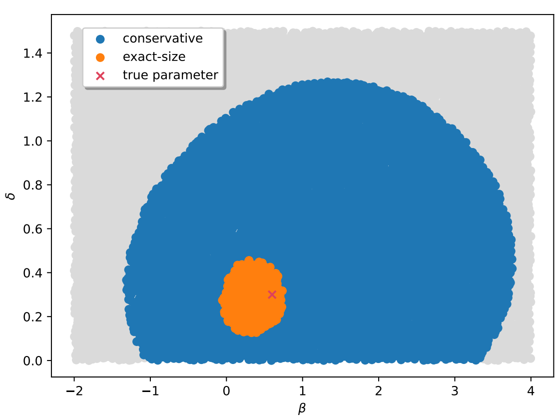
Figure 1. Confidence regions for the pair (β_1, δ) in the 6 player entry game, when the true value of $\beta_0 = 1$ is known a priori. The true parameter value is $(\beta_1, \delta) = (0.3, 0.6)$ and outcomes are selected uniformly within the predicted set of equilibrium. Confidence region CR_n is pictured in orange/light grey, while the conservative outer region CR_n^0 is pictured in blue/dark grey.



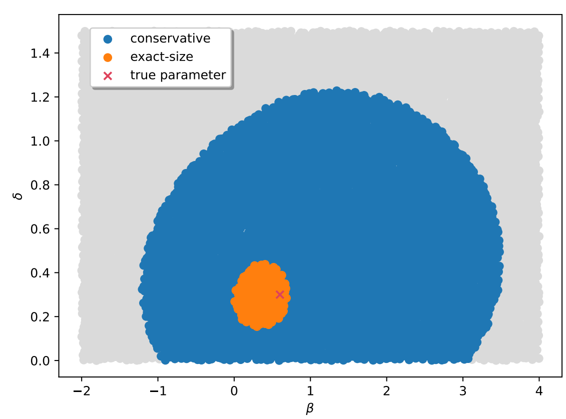
(a) $n = 100$



(b) $n = 500$



(c) $n = 1000$



(d) $n = 5000$

Johns Hopkins and Penn State