SYMPOSIUM

Optimal transportation and the falsifiability of incompletely specified economic models

Ivar Ekeland · Alfred Galichon · Marc Henry

Abstract A general framework is given to analyze the falsifiability of economic models based on a sample of their observable components. It is shown that, when the restrictions implied by the economic theory are insufficient to identify the unknown quantities of the structure, the duality of optimal transportation with zero–one cost function delivers interpretable and operational formulations of the hypothesis of specification correctness from which tests can be constructed to falsify the model.

I. Ekeland
University of British Columbia and PIMS, Vancouver, Canada
e-mail: ekeland@math.ubc.ca

A. Galichon (✉)
École polytechnique, Paris, France
e-mail: alfred.galichon@polytechnique.edu

M. Henry
Université de Montréal, CIRANO and CIREQ, Montreal, Canada
e-mail: marc.henry@umontreal.ca

Introduction

In many contexts, the ability to identify econometric models often rests on strong prior assumptions that are difficult to substantiate and even to analyze within the economic decision problem. A recent approach has been to forego such prior assumptions, thus giving up the ability to identify a single value of the parameter governing the model, and allow instead for a set of parameter values compatible with the empirical setup. A variety of models have been analyzed in this way, whether partial identification stems from incompletely specified models (typically models with multiple equilibria) or from structural data insufficiencies (typically cases of data censoring). See Manski (2005) for a recent survey on the topic.

All these incompletely specified models share the basic fundamental structure that a set of unobserved economic variables and a set of observed ones are linked by restrictions that stem from the theoretical economic model. In this paper, we propose a general framework for conducting inference in such contexts. This approach is articulated around the formulation of a hypothesis of compatibility of the true distribution of observable variables with the restrictions implied by the model as an optimal transportation problem. Given a hypothesized distribution for latent variables, compatibility of the true distribution of observed variables with the model is shown to be equivalent to the existence of a zero cost transportation plan from the hypothesized distribution of latent variables to the true distribution of observable variables, where the zero–one cost function is equal to one in cases of violations of the restrictions embodied in the model.

Two distinct types of economic restrictions are considered here. On the one hand, the case where the distribution of unobserved variables is parameterized yields a traditional optimal transportation formulation. On the other hand, the case where the distribution of unobserved economic variables are only restricted by a finite set of moment equalities yields an optimization formulation which is not a classical optimal transportation problem, but shares similar variational properties. In both cases the inspection of the dual of the specification problem's optimization formulation has three major benefits.

First, the optimization formulation relates the problem of falsifying incompletely specified economic models to the growing literature on optimal transportation (Rachev and Rüschendorf 1998; Villani 2003), in particular with relation to the literature on probability metrics (Zolotarev 1997, Chap. 1). Second, the dual formulation of the optimization problem provides significant dimension reduction, thereby allowing the construction of computable test statistics for the hypothesis of compatibility of true observable data distribution with the economic model given. Thirdly, and perhaps most importantly, in the case of models with discrete outcomes, the optimal transportation formulation allows to tap into a very rich combinatorial optimization literature relative to the discrete transport problem (Papadimitriou and Steiglitz 1998) thereby allowing inference in realistic models of industrial organization and other areas of economics where sophisticated empirical research is being carried out.

The paper is organized as follows. The next section sets out the framework, notations and defines the problem considered. Section 1 considers the case of parametric restrictions on the distribution of unobserved variables, gives the optimal transportation
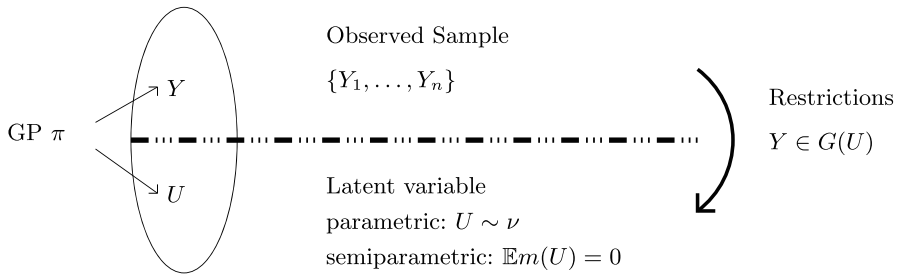
Fig. 1 Summary of the structure. GP stands for generating process, i.e. a joint distribution π generating the pairs (Y_i, U_i), $i = 1, \ldots, n$, the first component of which is observed

formulation of the compatibility of the distribution of observable variables with the economic model at hand, and discusses strategies to falsify the model based on a sample of realizations of the observable variables. Section 2 similarly considers the case of semiparametric restrictions on the distribution of unobservable variables and the last section concludes.

General framework and notations

We consider, as in Jovanovic (1989), an economic model that governs the behaviour of a collection of economic variables (Y, U), where Y is a random element taking values in the Polish space $\mathcal{Y}$ (endowed with its Borel σ-algebra $\mathcal{B}_\mathcal{Y}$) and U is a random element taking values in the Polish space $\mathcal{U}$ (endowed with its Borel σ-algebra $\mathcal{B}_\mathcal{U}$). Y represents the subcollection of observable economic variables generated by the unknown distribution P, and U represents the subcollection of unobservable economic variables generated by a distribution ν. The economic model provides a set of restrictions on the joint behaviour of observable and latent variables, i.e. a subset of $\mathcal{Y} \times \mathcal{U}$, which can be represented without loss of generality by a correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ (Fig. 1).

In all that follows, the correspondence will be assumed non-empty closed-valued and measurable, i.e. $G^{-1}(\mathcal{O}) := \{u \in \mathcal{U} : G(u) \cap \mathcal{O} \neq \varnothing\} \in \mathcal{B}_\mathcal{U}$ for all open subset $\mathcal{O}$ of $\mathcal{Y}$. A measurable selection of a measurable correspondence G is a measurable function g such that $g \in G$ almost surely, and Sel(G) denotes the collection of measurable selections of G (non-empty by the Kuratowski–Ryll–Nardzewski selection theorem). We shall denote by $c(y, u)$ a cost of transportation, i.e. a real valued function on $\mathcal{Y} \times \mathcal{U}$. For any set A, we denote by 1_A its indicator function, i.e. the function taking value 1 on A and 0 outside of A. $\mathcal{M}(\mathcal{Y})$ (resp. $\mathcal{M}(\mathcal{U})$) will denote the set of Borel probability measures on $\mathcal{Y}$ (resp. $\mathcal{U}$) and $\mathcal{M}(P, \nu)$ will denote the collection of Borel probability measures on $\mathcal{Y} \times \mathcal{U}$ with marginal distributions P and ν on $\mathcal{Y}$ and $\mathcal{U}$ respectively. We shall generally denote by π a typical element of $\mathcal{M}(P, \nu)$. For a Borel probability measure ν on $\mathcal{U}$ and a measurable correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$, we denote by νG^{-1} the set function that to a set A in $\mathcal{B}_\mathcal{Y}$ associates $\nu(G^{-1}(A)) = \nu(\{u \in \mathcal{U} : G(u) \cap A \neq \varnothing\})$. Note that the set function νG^{-1} is a *Choquet capacity functional* (Choquet 1954). The *Core* of a Choquet capacity functional νG^{-1}, denoted Core(νG^{-1}) is defined

as the collection of Borel probability measures set-wise dominated by νG^{-1}, i.e. $\text{Core}(\nu G^{-1}) = \{Q \in \mathcal{M}(\mathcal{Y}) : \forall A \in \mathcal{B}_{\mathcal{Y}}, Q(A) \leq \nu G^{-1}(A)\}$. In the terminology of cooperative games, if νG^{-1} defines a transferable utility game, $\nu G^{-1}(A)$ is the utility value or worth of coalition A and the Core of the game νG^{-1} is the collection of undominated allocations (Moulin 1995).

Example 1 (Social interactions with small groups) To illustrate this framework, consider the following model of discrete choice with social interactions, in the spirit of Manski (1993) and Brock and Durlauf (2001), but with the special feature that the interaction networks are small, so that multiple equilibria are more pervasive. Consider the variable $Y_i = 1$ if individual i smokes and 0 otherwise. Suppose the utility of individual i from smoking is given by $Y_i[-U_i + \sum_{j \in \mathcal{F}(i)} Y_j]$, where $\mathcal{F}(i)$ is the set of individuals that are directly connected to i in the social network (a graph which is given as a primitive of the problem), U_i is individual i's idiosyncratic disutility of smoking. Consider for instance the following simple configuration for the social network. There are three individuals A, B and C, connected in a line network A–B–C, so that $\mathcal{F}(A) = \mathcal{F}(C) = \{B\}$ and $\mathcal{F}(B) = \{A, C\}$. The following are the pure strategy Nash equilibria of the game, which define the equilibrium correspondence, hence the correspondence G.

- If $U_B > 2$ or ($U_B > 1$ and ($U_A > 1$ or $U_C > 1$)) or ($U_A > 1$ and $U_C > 1$), then "nobody smokes" is the unique equilibrium. So $G((U_A, U_B, U_C)) = \{(0, 0, 0)\}$.
- If ($U_A < 1$ and $U_B < 1$ and $U_C > 1$) then there are two equilibria, either "nobody smokes" or "A and B smoke" (and symmetrically if the roles of A and C are reversed). So $G((U_A, U_B, U_C)) = \{(0, 0, 0), (1, 1, 0)\}$.
- If ($U_A < 1$ and $U_B < 2$ and $U_C < 1$) then "everybody smokes" and "nobody smokes" are both equilibria. So $G((U_A, U_B, U_C)) = \{(0, 0, 0), (1, 1, 1)\}$.

Hence, the set of observable outcomes is $\mathcal{Y} = \{(0, 0, 0), (0, 1, 1), (1, 1, 0), (1, 1, 1)\}$. P is the true distribution of equilibria in a population of identical networks (true frequencies of elements in $\mathcal{Y}$), and ν is the distribution of idiosyncratic disutilities of smoking.

Example 2 (Diamond-type search model) Suppose there are N players searching for trading parters. Player i exerts effort $Y_i \in [0, 1]$ with cost $C(Y_i)$ to find a partner. A trader's probability of finding a partner is proportional to their own effort and the sum of other traders' efforts. Hence, the payoff function is

$$\pi_i(Y) = \epsilon Y_i \sum_{j \neq i} Y_j - C(Y_i),$$

where ϵ is the gains of trade observed by the players before making their effort choice, but not by the econometrician, who only knows it is distributed according to distribution ν, which is absolutely continuous with respect to Lebesgue measure. Assuming the cost function is increasing in effort, $x = 0$ is an equilibrium, and so is $Y = (\alpha(\epsilon), \ldots, \alpha(\epsilon))$, where $\alpha(\epsilon)$ satisfies $C'(\alpha) = \alpha(N-1)\epsilon$. In this case, $\mathcal{Y} = [0, 1]^N$ and the equilibrium correspondence is $G(\epsilon) = \{(0, \ldots, 0)^t, (\alpha(\epsilon), \ldots, \alpha(\epsilon))^t\}$. Note

that since both equilibrium are perfectly correlated, this is equivalent to the simplified formulation where $\mathcal{Y} = [0, 1]$ and $G(\epsilon) = \{0, \alpha(\epsilon)\}$.

Example 3 (Oligopoly entry models) A leading example of the framework above is that of empirical models of oligopoly entry, proposed in Bresnahan and Reiss (1990) and Berry (1992), and considered in the framework of partial identification by Tamer (2003), Andrews et al. (2003), Berry and Tamer (2006), Ciliberto and Tamer (2006), and Pakes (2004) among others. For illustration purposes, we describe the special case of this framework extensively studied in Tamer (2003), Berry and Tamer (2006) and Ciliberto and Tamer (2006). Two firms are present in an industry, and a firm decides to enter the market if it makes a non negative profit in a pure strategy Nash equilibrium. Y_i is firm i's strategy, and it is equal to 1 if firm i enters the market, and zero otherwise. Y denotes the vector (Y_1, Y_2) of strategies of both firms. In standard notation, Y_{-i} denotes the vector of strategies of firm $j = 3 - i$. In models of oligopoly entry, the profit π_i of firm i is allowed to depend on strategies Y_{-i} of the other firm, as well as on a profit shifter ϵ_i that is observed by both firms but not by the econometrician, and a vector of unknown structural parameters θ. Profit functions are supposed to have the following linear form $\pi_{im} = \delta_{-i} Y_{-i} + \epsilon_i$, where the unobserved profit shifters are distributed according to a known distribution and where parameters $\delta_{1,2}$ are given. Hence, $Y_i = 1$ if $\delta_{-i} Y_{-i} + \epsilon_i \geq 0$ and zero otherwise. As noted in Tamer (2003), if monopoly profits are larger than duopoly profits, i.e. $\delta_i < 0$, for $i = 1, 2$, and if $0 \leq \epsilon_i \leq -\delta_{-i}$, $i = 1, 2$, then there are multiple equilibria, since the model predicts either $Y_1 = 1$ and $Y_2 = 0$ or $Y_1 = 0$ and $Y_2 = 1$. The set of possible outcomes is $\mathcal{Y} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, and the correspondence G is given in Fig. 2.

We are interested in characterizing restrictions on the distribution of observables induced by the model, in order to devise methods to falsify the model based on a sample of repeated observations of Y. We shall successively consider two leading cases of this framework. First the case where the distribution ν of unobservable variables is given by the economic model, and second, the case where a finite collection of moments of the distribution ν of unobservable variables are given by the economic model.

The general principle we shall develop here in both parts is therefore the following. We want to test the compatibility of a *reduced-form model*, summarized by the distribution P of an observed variable Y, with a *structural model*, summarized by a set $\mathcal{V}$ of distributions ν for the latent variable U. Two leading cases will be considered for the set $\mathcal{V}$: the parametric case, where $\mathcal{V}$ contains one element $\mathcal{V} = \{\nu\}$, and the semiparametric case, where the distributions ν in $\mathcal{V}$ are specified by a finite number of moment restrictions $\mathbb{E}_\nu [m_i (U)] = 0$.

The restriction of the model defines compatibility between outcomes of the reduced-form and the structural models: such outcomes u and y are compatible if and only if the binary relation $y \in G (u)$ holds (this relation defines G).

Now we turn to the compatibility of the probabilistic models, namely of the specification of distributions for U and Y. The models $Y \sim P$ and $U \sim \nu \in \mathcal{V}$ are compatible if there is a joint distribution π for the pair (Y, U) with respective marginals P and some $\nu \in \mathcal{V}$ such that $Y \in G (U)$ holds π almost surely. In other words, P and $\mathcal{V}$ are
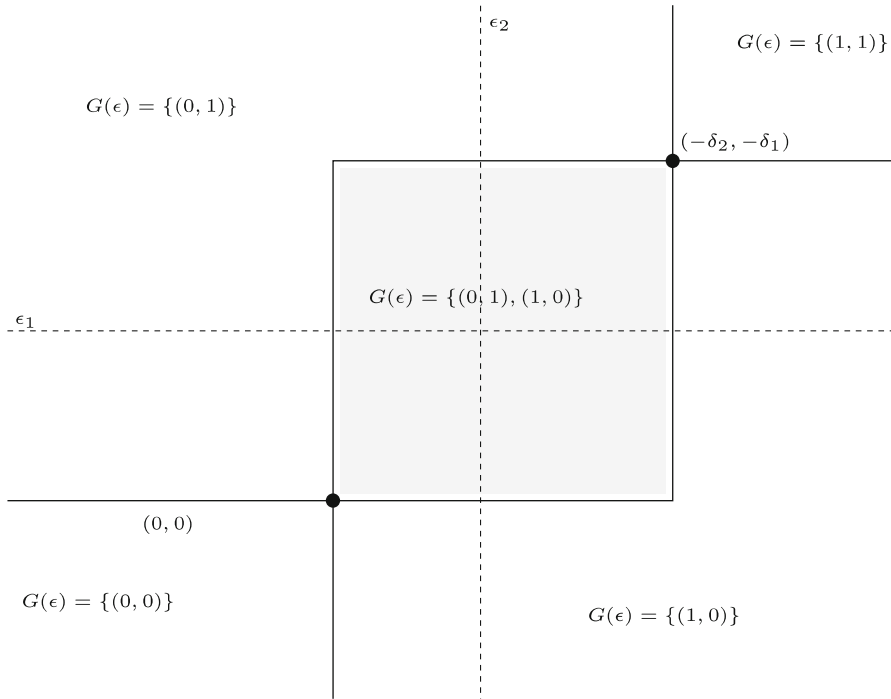
Fig. 2 Equilibrium correspondence in Example 3

compatible if and only if

$$\exists v \in \mathcal{V}, \quad \exists \pi \in \mathcal{M}(P, v) : \Pr_\pi \{Y \notin G(U)\} = 0.$$

In the sequel we shall examine equivalent formulations of this compatibility principle, first in the parametric case and then in the semiparametric case.

1 Parametric restrictions on unobservables

Consider first the case where the economic model consists in the correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ and the distribution v of unobservables. The observables are fully characterized by their distribution P, which is unknown, but can be estimated from data. The question of compatibility of the model with the data can be formalized as follows: Consider the restrictions imposed by the model on the joint distribution π of the pair (Y, U):

– its marginal with respect to Y is P,
– its marginal with respect to U is v,
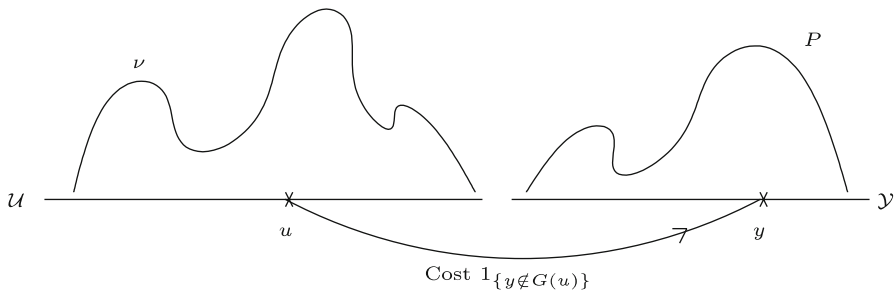– the economic restrictions $Y \in G(U)$ hold π-almost surely.

Fig. 3 Transportation plan from mass distribution ν on $\mathcal{U}$ to mass distribution P on $\mathcal{Y}$ with cost of transportation equal to 1 if the restrictions are violated, and 0 otherwise

A probability distribution π that satisfies the restrictions above may or may not exist. If and only if it does, we say that the distribution P of observable variables is compatible with the economic model (G, ν).

Definition 1 A distribution P is compatible with the model (G, ν) for (Y, U) if there exists a probability distribution π for the vector (Y, U) with marginals P and ν such that $\pi(\{Y \in G(U)\}) = 1$.

1.1 Optimal transportation formulation

This hypothesis of compatibility has the following optimization interpretation. The distribution P is compatible with the model (G, ν) if and only if

$$\exists \pi \in \mathcal{M}(P, \nu) : \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} \, d\pi(y, u) = 0,$$

and thus we see that it is equivalent to the existence of a zero cost transportation plan for the problem of transporting mass ν into mass P with zero–one cost function $c(y, u) = 1_{\{y \notin G(u)\}}$ associated with violations of the restrictions implied by the model (Fig. 3).

The two dual formulations of this optimal transportation problem are the following:

$$\text{(P)} \quad \inf_{\pi \in \mathcal{M}(P, \nu)} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} \, d\pi(y, u)$$

$$\text{(D)} \quad \sup_{f(y) + h(u) \leq 1_{\{y \notin G(u)\}}} \int_{\mathcal{Y}} f \, dP + \int_{\mathcal{U}} h \, d\nu$$

Through applications of optimal transportation duality theory, it can be shown that the two programs are equal and that the infimum in (P) is attained, so that the compatibility hypothesis of Definition 1 is equivalent to (D) $= 0$, which in turn can be shown to be equivalent to

$$\sup_{A \in \mathcal{B}_{\mathcal{Y}}} \left[P(A) - \nu \left(G^{-1}(A) \right) \right] = 0, \tag{1}$$

using the zero–one nature of the cost function to specialize the test functions f and h to indicator functions of Borel sets. Note that it is relatively easy to show necessity, since the definition of compatibility implies that $Y \in A \Rightarrow U \in G^{-1}(A)$, so that $1_{\{Y \in A\}} \leq 1_{\{U \in G^{-1}(A)\}}$, π-almost surely. Taking expectation, we have $\mathbb{E}_{\pi}(1_{\{Y \in A\}}) \leq \mathbb{E}_{\pi}(1_{\{U \in G^{-1}(A)\}})$, which yields $P(A) \leq \nu(G^{-1}(A))$. The converse relies on the duality of optimal transportation (Villani 2003, Theorem 1.27, p. 44 and Galichon and Henry 2008 for details). Note also that in the particular case where the spaces of the observed and latent variables are the same $\mathcal{Y} = \mathcal{U}$ and G is the identity function $G(u) = \{u\}$, then (1) defines the Total Variation metric between P and ν. When $\mathcal{Y} = \mathcal{U}$ and $G(u) = \{y \in \mathcal{Y} : d(y, u) \leq \varepsilon\}$, the above duality boils down to a celebrated theorem due to Strassen (Dudley 2002, section 11.6). A closely related result was proven by Artstein in Artstein (1983, Theorem 3.1), using an extension of the marriage lemma.

Example 1 continued: In the social network example, the set of equilibria was $\mathcal{Y} = \{(0, 0, 0), (0, 1, 1), (1, 1, 0), (1, 1, 1)\}$. Let p_{000} denote the true frequency of equilibrium $(0, 0, 0)$, p_{011} that of $(0, 1, 1)$, p_{110} that of $(1, 1, 0)$ and p_{111} that of $(1, 1, 1)$. The set of predicted combinations of equilibria is $\{\{(0, 0, 0)\}, \{(0, 0, 0), (1, 1, 0)\}, \{(0, 0, 0), (1, 1, 0)\}, \{(0, 0, 0), (1, 1, 1)\}\}$. Call q_{ω} the likelihood of equilibrium combination ω predicted by the model, so that for instance $q_{000,110} = \nu\{u \in \mathcal{U} : G(u) = \{(0, 0, 0), (1, 1, 0)\}\}$. The previous discussions shows that $(p_{000}, p_{011}, p_{110}, p_{111})$ is compatible with the model (G, ν) if and only if for all subset A of $\mathcal{Y}$, $P(A) \leq \nu G^{-1}(A)$. Take the subset $\{(0, 1, 1), (1, 1, 0)\}$ for instance, the constraint to check is

$$p_{011} + p_{110} \leq \nu G^{-1}(\{(0, 1, 1), (1, 1, 0)\}) = q_{000,011} + q_{000,110},$$

the ν probability of the region of $\mathcal{U}$ with incidence of the equilibrium $(0, 1, 1)$ or $(1, 1, 0)$.

Example 2 continued: In the case of the search model, it can be shown further that compatibility of the model with the distribution of observed equilibria is equivalent to $P([0, y] \leq \nu([0, \alpha^{-1}(y)])$ and $P([y, 1] \leq \nu([\alpha^{-1}(y), \alpha^{-1}(1)])$ for all $y \in [0, 1]$.

Example 3 continued: In the case of the duopoly entry game, the model is compatible with distribution P of observable variables (indicator of presence of firms in the market) if and only if the 16 inequalities $P(A) \leq \nu(G^{-1}(A))$, or in a different notation $\mathbb{P}(Y \in A) \leq \mathbb{P}(G(\epsilon) \cap A \neq \varnothing)$ for all sets A in $2^{\mathcal{Y}}$.

1.2 Game theoretic interpretation of the specification problem

The optimal transportation of the specification problem at hand leads to an interpretation of the latter as a game between the Analyst and a malevolent Nature. This

highlights connections between partial identification and robust decision-making (Hansen and Sargent 2001) and ambiguity (Maccheroni et al. 2006). As above, P and ν are given. In the special case where we want to test whether the true functional relation between observable and unobservable variables is γ_0 (i.e. the complete specification problem), and where P and ν are absolutely continuous with respect to Lebesgue measure, the optimal transportation formulation of the specification problem involves the minimization over the set of joint probability measures with marginals P and ν of the integral $\int 1_{\{y \neq \gamma_0(u)\}} \, d\pi(y, u)$. The latter can be written as the minimax problem

$$\min_{\tau} \max_{V} \int \left[1_{\{\tau(u) \neq \gamma_0(u)\}} - V(\tau(u)) \right] d\nu(u) + \int V(y) dP(y).$$

This yields the interpretation as a zero-sum game between the Analyst and Nature, where the Analyst pays Nature the amount

$$\int \left[1_{\{\tau(u) \neq \gamma_0(u)\}} - V(\tau(u)) \right] d\nu(u) + \int V(y) dP(y). \tag{2}$$

P and ν are fixed. The Analyst is asked to propose a plausible functional relation $y = \tau(u)$ between observed and latent variables, and Nature chooses V in order to maximize transfer (2) from the Analyst. This transfer can be decomposed into two terms. The first term $\int V(y) dP(y) - \int V(\tau(u)) d\nu(u)$ is a punishment for guessing the wrong distribution: this term can be arbitrarily large unless $P = \nu\tau^{-1}$. The second term, $\int 1_{\{\tau(u) \neq \gamma_0(u)\}} d\nu(u)$ is an incentive to guess τ close to the true functional relation γ_0 between u and y.

The value of this game for Nature is equal to $T(P) = \inf\{\mathbb{P}(\tau(U) \neq \gamma_0(U)) : U \sim \nu, \ \tau(U) \sim P\}$ and is independent of who moves first. This follows from the Monge–Kantorovitch duality. Indeed, if Nature moves first and plays V, the Analyst will choose τ to minimize $\int (1_{\{\tau(u) \neq \gamma_0(u)\}} - V(\tau(u))) d\nu(u)$. Denoting $V^*(u) = \inf_y \{1_{\{y \neq \gamma_0(u)\}} - V(y)\}$, the value of this game for Nature is

$$\sup_{V^*(u) + V(y) \leq 1_{\{y \neq \gamma_0(u)\}}} \int V^*(u) d\nu(u) + \int V(y) dP(y).$$

If, on the other hand, the Analyst moves first and plays τ, then Nature will receive an arbitrarily large transfer if $P \neq \nu\tau^{-1}$, and a transfer of $\int 1_{\{\tau(u) \neq \gamma_0(u)\}} d\nu(u)$ independent of V otherwise. The value of the game for Nature is therefore $\inf\{\mathbb{P}(\tau(U) \neq \gamma_0(U)) : U \sim \nu, \ \tau(U) \sim P\}$. The Monge–Kantorovitch duality states precisely that the value when Nature plays first is equal to the value when Analyst plays first.

Finally, we have an interpretation of the set of observable distributions P that are compatible with the model (G, ν) as the set of distributions P such that the Analyst is willing to play the game, i.e. such that the value of the game is zero for some functional relationship γ_0 among the selections of G.

1.3 Test of compatibility

We now consider falsifiability of the incompletely specified model through a test of the null hypothesis that P is compatible with (G, ν). Falsifying the model in this framework corresponds to the finding that a sample $(Y_1, \ldots, Y_n)$ of n copies of Y distributed according to the unknown true distribution P was not generated as part of a sample $((Y_1, U_1), \ldots, (Y_n, U_n))$ distributed according to a fixed π with marginal ν on $\mathcal{U}$ and satisfying the restrictions $Y \in G(U)$ almost surely. Using the results of the previous section, this can be expressed in the following equivalent ways.

Proposition 1 *The following statements are equivalent*:

   (i) *The observable distribution P is compatible with the model (G, ν)*,
  (ii) $\inf_{\pi \in \mathcal{M}(P, \nu)} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} d\pi(y, u) = 0$,
 (iii) $\sup_{A \in \mathcal{B}_{\mathcal{Y}}} [P(A) - \nu(G^{-1}(A))] = 0$.

Call P_n the empirical distribution, defined by $P_n(A) = \sum_{i=1}^{n} 1_{\{Y_i \in A\}}/n$ for all A measurable, and form the empirical analogues of the conditions above as

$$(\text{EP}) \quad \inf_{\pi \in \mathcal{M}(P_n, \nu)} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} d\pi(y, u)$$

$$(\text{ED}) \quad \sup_{A \in \mathcal{B}_{\mathcal{Y}}} \left[P_n(A) - \nu(G^{-1}(A)) \right].$$

Note first that by the duality of optimal transportation, the empirical primal (EP) and the empirical dual (ED) are equal. In the case $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$, Galichon and Henry (2008) propose a testing procedure based on the asymptotic treatment of the feasible statistic

$$T_n = \sqrt{n} \sup_{A \in \mathcal{C}_n} \left[P_n(A) - \nu G^{-1}(A) \right], \quad \text{with } \mathcal{C}_n = \{(-\infty, Y_i], (Y_i, \infty) : i = 1, \ldots, n\}.$$

More general families of test statistic for this problem can be derived from the following observation: consider the total variation metric defined by

$$d_{\text{TV}}(\mu_1, \mu_2) = \sup_{A \in \mathcal{B}_{\mathcal{Y}}} (\mu_1(A) - \mu_2(A))$$

for any two probability measures μ_1 and μ_2 on $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$, and

$$d_{TV}(P, \mathcal{Q}) = \inf_{Q \in \mathcal{Q}} d_{TV}(P, Q)$$

for a probability measure P and a set of probability measures $\mathcal{Q}$. Galichon and Henry (2008) derive conditions under which the equalities

$$d_{\mathrm{TV}}\left(P_n, \mathrm{Core}(\nu G^{-1})\right) = \inf_{g \in \mathrm{Sel}(G)} \sup_{A \in \mathcal{B}_{\mathcal{Y}}} \left(P_n(A) - \nu g^{-1}(A)\right)$$

$$= \sup_{A \in \mathcal{B}_{\mathcal{Y}}} \inf_{g \in \mathrm{Sel}(G)} \left(P_n(A) - \nu g^{-1}(A)\right)$$

$$= \sup_{A \in \mathcal{B}_{\mathcal{Y}}} \left(P_n(A) - \nu(G^{-1}(A))\right)$$

hold, so that the empirical dual is equal to the total variation distance between the empirical distribution P_n and $\mathrm{Core}(\nu G^{-1})$. Hence, (ED) yields a family of test statistics $d(P_n, \mathrm{Core}(\nu G^{-1}))$, for the falsification of the model (G, ν), where d satisfies $d(x, A) = 0$ if $x \in A$ and 1 otherwise.

Alternatively, a family of statistics can be derived from the empirical primal (EP) if the 0–1 cost is replaced by d as above, yielding the statistics

$$\inf_{\pi \in \mathcal{M}(P_n, \nu)} \int\int_{\mathcal{Y} \times \mathcal{U}} \mathrm{d}(y, G(u)) \mathrm{d}\pi(y, u)$$

generalizing goodness-of-fit statistics based on the Wasserstein distance (del Barrio et al. 1999).

1.4 Computational aspects of the transportation formulation

In addition to producing families of test statistics, hence inference strategies, for partially identified structures, the optimal transportation formulation has clear computational advantages. First of all, efficient algorithms for the computation of the optimal transport map rely on both primal and dual formulations of the optimization problem. More specifically, in cases with discrete observable outcomes, the Monge–Kantorovitch optimal transportation problem reduces to its discrete counterpart, sometimes called the Hitchcock problem [Hitchcock (1941); Kantorovich (1942); Koopmans (1949)]. This problem has a long history of applications in a vast array of fields, and hence spurred the development of many families of algorithms and implementations since Ford and Fulkerson (1957). The optimal transportation formulation therefore allows the development of procedures for testing incomplete structures and estimating partially identified parameters that are vastly more efficient than existing ones [see for instance (2008) for the efficient computation of the the identified set in discrete games].

2 Semiparametric restrictions on unobservables

As before, we consider an economic model that governs the behaviour of a collection of economic variables (Y, U). Here, Y is a random element taking values in the

Polish space $\mathcal{Y}$ (endowed with its Borel σ-algebra $\mathcal{B}_{\mathcal{Y}}$) and U is a random vector taking values in $\mathcal{U} \subseteq \mathbb{R}^{d_u}$. Y represents the subcollection of observable economic variables generated by the unknown distribution P, and U represents the subcollection of unobservable economic variables generated by a distribution ν. As before, the economic model provides a set of restrictions on the joint behaviour of observable and latent variables, i.e. a subset of $\mathcal{Y} \times \mathcal{U}$ represented by the measurable correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$. The distribution ν of the unobservable variables U is now assumed to satisfy a set of moment conditions, namely

$$\mathbb{E}_\nu(m_i(U)) = 0, \quad m_i : \mathcal{U} \to \mathbb{R}, \quad i = 1, \ldots, d_m \tag{3}$$

and we denote by $\mathcal{V}$ the set of distributions that satisfy (3), and by $\mathcal{M}(P, \mathcal{V})$ the collection of Borel probability measures with one marginal fixed equal to P and the other marginal belonging to the set $\mathcal{V}$. Note that a limit case of this framework, where an infinite collection of moment conditions uniquely determines the distribution of unobservable variables, i.e. when $\mathcal{V}$ is a singleton, we recover the parametric setup, with a classical optimal transportation formulation as in Sect. 1.

Example 4 (Model defined by moment inequalities.) A special case of the specification above is provided by models defined by moment inequalities.

$$\mathbb{E}(\varphi_i(Y)) \le 0, \quad \varphi_i : \mathcal{Y} \to \mathbb{R}, \quad i = 1, \ldots, d_\varphi. \tag{4}$$

This is a special case of our general structure, where $\mathcal{U} \subseteq \mathbb{R}^{d_u}$ and

$$G(u) = \{y \in \mathcal{Y} : \ u_i \ge \varphi_i(y), \ i = 1, \ldots, d_u\},$$

and $m_i(u) = u, i = 1, \ldots, d_\varphi$, with $d_u = d_\varphi$.

Example 5 Model defined by conditional moment inequalities.

$$\mathbb{E}(\varphi_i(Y)|X) \le 0, \quad \varphi_i : \mathcal{Y} \to \mathbb{R}, \quad i = 1, \ldots, d_\varphi, \tag{5}$$

where X is a sub-vector of Y. Bierens (1990) shows that this model can be equivalently rephrased as

$$\mathbb{E}(\varphi_i(Y)1\{t_1 \le X \le t_2\}) \le 0, \qquad \varphi_i : \mathcal{Y} \to \mathbb{R}, \quad i = 1, \ldots, d_\varphi, \tag{6}$$

for all pairs $(t_1, t_2) \in \mathbb{R}^{2d_x}$ (the inequality is understood element by element). Conditionally on the observed sample, this can be reduced to a finite set of moment inequalities by limiting the class of pairs (t_1, t_2) to observed pairs (X_i, X_j), $X_i < X_j$. Hence this fits into the framework of Example 4.

Example 6 Unobserved random censoring (also known as accelerated failure time) model. A continuous variable $Z = \mu(X) + \epsilon$, where μ is known, is censored by a random variable C. The only observable variables are X, $V = \min(Z, C)$ and $D = 1\{Z < C\}$. The error term ϵ is supposed to have zero conditional median

$P(\epsilon < 0|X) = 0$. Khan and Tamer (2006) show that this model can be equivalently rephrased in terms of unconditional moment inequalities.

$$\mathbb{E}\left[\left(1\{V \geq \mu(X)\} - \frac{1}{2}\right) 1\{t_1 \leq X \leq t_2\}\right] \leq 0$$

$$\mathbb{E}\left[\left(\frac{1}{2} - D \times 1\{V \leq \mu(X)\}\right) 1\{t_1 \leq X \leq t_2\}\right] \geq 0$$

for all pairs $(t_1, t_2) \in \mathbb{R}^{2d_x}$ (the inequality is understood element by element). Hence this fits into the framework of Example 5.

Example 1 continued In case of models with multiple equilibria such as Example 1, where the idiosyncratic disutility of smoking is only restricted by a finite collection of moment equalities, for instance $\mathbb{E}U = 0$, the model cannot be written in the familiar moment inequality formulation of Example 4. Hence the need for a new approach to the problem of falsifying the model.

Finally we turn to an example of binary response, which we shall use as pilot examples for illustrative purposes.

Pilot Example 1: a binary response model. The observed variables Y and X are related by $Z = 1\{X + \varepsilon \leq 0\}$, under the conditional median restriction $\Pr(\varepsilon \leq 0|X) = \eta$ for a known η. In our framework the vector of observed variables is $Y = (Z, X)'$, and to deal with the conditioning, we take the vector U to also include X, i.e. $U = (X, \varepsilon)'$. To simplify exposition, suppose X only takes values in $\{-1, 1\}$, so that $\mathcal{Y} = \{0, 1\} \times \{-1, 1\}$ and $\mathcal{U} = \{-1, 1\} \times [-2, 2]$, where the restriction on the domain of ε is to ensure compactness only. The multi-valued correspondence defining the model is $\Gamma : \mathcal{U} \rightrightarrows \mathcal{Y}$ characterized by $\Gamma(1, x) = \{x\} \times (-2, -x]$ and $\Gamma(0, x) = \{x\} \times (-x, 2]$. The two moment restrictions are $m_{\pm}(x, \varepsilon) = (1\{\varepsilon \leq 0\} - \eta)(1 \pm x)$.

We are now in the case where the economic model consists in the correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ and a finite set of moment restrictions on the distribution ν of unobservables. Denote the model $(G, \mathcal{V})$. Again, the observables are fully characterized by their distribution P, which is unknown, but can be estimated from data. Consider now the restrictions imposed by the model on the joint distribution π of the pair (Y, U):

– its marginal with respect to Y is P,
– its marginal with respect to U belongs to $\mathcal{V}$,
– the economic restrictions $Y \in G(U)$ hold π almost surely.

Again, a probability distribution π that satisfies the restrictions above may or may not exist. If and only if it does, we say that the distribution P of observable variables is compatible with the economic model $(G, \mathcal{V})$.

Definition 2 A distribution P is compatible with the model $(G, \mathcal{V})$ for (Y, U) if there exists a law π for the vector (Y, U) with marginals P with respect to Y and marginal $\nu \in \mathcal{V}$ with respect to U such that $\pi(\{Y \in G(U)\}) = 1$.

2.1 Optimization formulation

This hypothesis of compatibility has a similar optimization interpretation as in the case of parametric restrictions on unobservables. The distribution P is compatible with the model $(G, \mathcal{V})$ if and only if

$$\exists \pi \in \mathcal{M}(P, \mathcal{V}) : \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} d\pi(y, u) = 0,$$

or equivalently

$$\min_{\pi \in \mathcal{M}(P, \mathcal{V})} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} d\pi(y, u) = 0. \tag{7}$$

Although this optimization problem differs from the optimal transportation problem considered above, we shall see that inspection of the dual nevertheless provides a dimension reduction which will allow to devise strategies to falsify the model based on a sample of realizations of Y. However, before inspecting the dual, we need to show that the minimum in (7) is actually attained, so that compatibility of observable distribution P with the model $(G, \mathcal{V})$ is equivalent to

$$\inf_{\pi \in \mathcal{M}(P, \mathcal{V})} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} d\pi(y, u) = 0. \tag{8}$$

The following example shows that the infimum is not always attained.

Example 7 Let $P = N(0, 1)$, $\mathcal{U} = \mathcal{Y} = \mathbb{R}$, $\mathcal{V} = \{v : \mathbb{E}_v(U) = 0\}$, and $\Gamma(y) = \{1\}$ for all $y \in \mathcal{Y}$, and consider the distribution $\pi_m = P \otimes v_m$ such that $v_m(\{1\}) = 1 - 1/m$, and $v_m(\{1 - m\}) = 1/m$. The π_m probability of $Y \notin \Gamma(U)$ is $1/m$ which indeed tends to zero as $m \to \infty$, but it is clear that there exists no distribution v which puts all mass on $\{1\}$ and has expectation 0.

It is clear from Example 7 that we need to make some form of assumption to avoid letting masses drift off to infinity. The theorem below gives formal conditions under which quasi-consistent alternatives are ruled out. It says essentially that the moment functions $m(u)$ need to be bounded.

Assumption 1 (Uniform integrability)

$$\lim_{M \to \infty} \sup_{v \in \mathcal{V}} v \left[\| m(U) \| \, 1_{\{\| m(U) \| > M\}} \right] = 0,$$

where $\| m(U) \|$ denotes the norm of the vector with components $m_i(U)$, for $1 \leq i \leq d_m$.

Assumption 2 (Tightness) For every $K \geq 0$, the set $\{u : \| m(u) \| \leq K\}$ is included in a compact set.

Assumption 1 is an assumption of uniform integrability. It is immediate to note that Assumptions 1 and 2 are satisfied when the moment functions $m(u)$ are bounded and $\mathcal{U}$ is compact.

Assumption 3 (Closed graph) The graph of G, i.e. $\{(y, u) \in \mathcal{Y} \times \mathcal{U} : y \in G(u)\}$ is closed.

In Example 4, by (Rockafellar et al., 1998, Theorem 1.6, p. 9), we know that Assumption 3 is satisfied when the moment functions φ_j, $j = 1, \ldots, d_\varphi$ are lower semi-continuous.

We can now state the result:

Theorem 1 *Under Assumptions* 1, 2 *and* 3, (8) *is equivalent to the compatibility of observable distribution P with model* $(G, \mathcal{V})$.

The two dual formulations of this optimization problem are the following:

$$(\text{P}) \quad \inf_{\pi \in \mathcal{M}(P, \mathcal{V})} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} \mathrm{d}\pi(y, u)$$

$$(\text{D}) \quad \sup_{f(y) + \lambda' m(u) \leq 1_{\{y \notin G(u)\}}} \int_{\mathcal{Y}} f \, \mathrm{d}P.$$

Since u does not enter in the dual functional, the dual constraint can be rewritten as $f(y) = \inf_u \{1_{\{y \notin G(u)\}} - \lambda' m(u)\}$, so that the dual program can be rewritten

$$T(P, \mathcal{V}) := \sup_{\lambda \in \mathbb{R}^{d_m}} \int_{\mathcal{Y}} \left(\inf_{u \in \mathcal{U}} [1_{\{y \notin G(u)\}} - \lambda' m(u)] \right) \mathrm{d}P(y),$$

which does not involve optimizing over an infinite dimensional space as the primal program did.

Pilot Example 1 continued Here, we have $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ and the integrand in the integral defining T above becomes

$$g_\lambda(x, 0) = \min \left(\inf_{\varepsilon \geq -x} \{-\lambda' m(\varepsilon, x)\}; \inf_{\varepsilon \leq -x} \{1 - \lambda' m(\varepsilon, x)\} \right),$$

$$g_\lambda(x, 1) = \min \left(\inf_{\varepsilon \leq -x} \{-\lambda' m(\varepsilon, x)\}; \inf_{\varepsilon \geq -x} \{1 - \lambda' m(\varepsilon, x)\} \right).$$

However, the dual formulation is useless if primal and dual are not equal. Note first that taking expectation in the dual constraint immediately yields (D) $\leq$ (P), which is the weak duality inequality. The converse inequality is shown below.

Assumption 4 (Slater condition) There exists a P-integrable function f and a vector λ and $\epsilon > 0$ such that for all $(y, u) \in \mathcal{Y} \times \mathcal{U}$, $f(y) + \lambda' m(u) < 1\{y \notin G(u)\} - \epsilon$.

The Slater condition is an interior condition, i.e. it ensures there exists a feasible solution to the optimization problem in the interior of the constraints. Notice that when the m_i are bounded, the Slater condition is always satisfied.

Theorem 2 (No duality gap) *Under assumptions* 1, 2, 3 *and* 4, *the observable distribution is compatible with model* $(G, \mathcal{V})$ *if and only if* $T(P, \mathcal{V}) = 0$.

As described in the Appendix, this result is ensured by the fact that there is no duality gap, i.e. that the statistic obtained by duality is indeed positive when the primal is.

2.2 Test of compatibility

We now consider falsifiability of the model with semiparametric constraints on unobservables through a test of the null hypothesis that P is compatible with $(G, \mathcal{V})$. Falsifying the model in this framework corresponds to the finding that a sample $(Y_1, \ldots, Y_n)$ of n copies of Y distributed according to the unknown true distribution P was not been generated as part of a sample $((Y_1, U_1), \ldots, (Y_n, U_n))$ distributed according to a fixed π with U-marginal ν in $\mathcal{V}$ and satisfying the restrictions $Y \in G(U)$ almost surely. Using the results of the previous section, this can be expressed in the following equivalent ways.

Proposition 2 *The following statements are equivalent:*

(i)   *The observable distribution P is compatible with the model $(G, \mathcal{V})$,*
(ii) $\inf_{\pi \in \mathcal{M}(P, \mathcal{V})} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} \mathrm{d}\pi(y, u) = 0$,
(iii) $\sup_{\lambda \in \mathbb{R}^{dm}} \int_{\mathcal{Y}} \left(\inf_{u \in \mathcal{U}} [1_{\{y \notin G(u)\}} - \lambda' m(u)] \right) \mathrm{d}P(y)$.

Call P_n the empirical distribution, defined by $P_n(A) = \sum_{i=1}^{n} 1_{Y_i \in A} / n$ for all A measurable, and form the empirical analogues of the conditions above as

$$(\text{EP}) \quad \inf_{\pi \in \mathcal{M}(P_n, \mathcal{V})} \int_{\mathcal{Y} \times \mathcal{U}} 1_{\{y \notin G(u)\}} \mathrm{d}\pi(y, u)$$

$$(\text{ED}) \quad \sup_{\lambda \in \mathbb{R}^{dm}} \frac{1}{n} \sum_{i=1}^{n} \left(\inf_{u \in \mathcal{U}} [1_{\{Y_i \notin G(u)\}} - \lambda' m(u)] \right).$$

Note first that by the duality result of theorem 2, the empirical primal (EP) and the empirical dual (ED) are equal. As in the parametric case, the cost function $c(y, u) = 1_{\{y \notin G(u)\}}$ can be replaced by $c(y, u) = d(y, G(u)) > 0$ if $y \notin G(u)$ and equal to 0 if $y \in G(u)$, to yield a family of numerically equivalent test statistics. Quantiles of their limiting distribution, or obtained from a bootstrap procedure can be used to form a test of compatibility, however, since (ED) involves two consecutive optimizations, a computationally more appealing procedure called dilation is proposed in Galichon and Henry (2006). The idea is to control the size of the test nonparametrically so as to compute (ED) only once. For a test with level $1 - \alpha$, compute a correspondence

$J_n : \mathcal{Y} \rightrightarrows \mathcal{Y}$ such that there exist a pair of random vectors Y and Y^* with marginal distributions P and P_n, respectively, and satisfying $Y^* \in J_n(Y)$ with probability $1 - \alpha$. The test then consists in rejecting compatibility of the unknown distribution P of the observables with the model $(G, \mathcal{V})$ if and only if the known empirical distribution P_n is not compatible with the model $(J_n \circ G, \mathcal{V})$, i.e. if

$$\sup_{\lambda \in \mathbb{R}^{d_m}} \frac{1}{n} \sum_{i=1}^n \left(\inf_{u \in \mathcal{U}} [1_{\{Y_i \notin J_n \circ G(u)\}} - \lambda' m(u)] \right) \neq 0.$$

3 Conclusion

We have proposed an optimal transportation formulation of the problem of testing compatibility of an incompletely specified economic model with the distribution of its observable components. In addition to relating this problem to a rich optimization literature, it allows the construction of computable test statistics and the application of efficient combinatorial optimization algorithms to the problem of inference in discrete games with multiple equilibria. A major application of tests of incomplete specifications is the construction of confidence regions for partially identified parameters. In this respect, the optimal transportation formulation proposed here allows the direct application of the methodology proposed in the seminal paper of Chernozhukov et al. (2007) to general models with multiple equilibria.

Appendix

Proof of results in the main text

Lemma 1 *Under assumptions* 1 *and* 2, $\mathcal{V}$ *is uniformly tight.*

Proof of Lemma 1 For $M > 1$, by Assumptions 1,

$$\sup_{v \in \mathcal{V}} v(\{\|m(U)\| > M\}) \leq \sup_{v \in \mathcal{V}} v\left[\|m(U)\| 1_{\{\|m(U)\| > M\}}\right] \to 0 \text{ as } M \to \infty,$$

hence for $\epsilon > 0$, there exists $M > 0$ such that

$$1 - \epsilon \leq \sup_{v \in \mathcal{V}} v(\{\|m(U)\| \leq M\})$$

but by Assumption 2, there exists a compact set K such that $\{\|m(U)\| \leq M\} \subset K$.

Lemma 2 *If* $\mathcal{V}$ *is uniformly tight, then* $\mathcal{M}(P, \mathcal{V})$ *is uniformly tight.*

Proof of Lemma 2 For $\epsilon > 0$, there exists a compact $K_Y \subset \mathcal{Y}$ such that $P(K_Y) \geq 1 - \epsilon/2$; by tightness of $\mathcal{V}$, there exists also a compact $K_U \subset \mathcal{U}$ such that $v(K_U) \geq 1 - \epsilon/2$ for all $v \in \mathcal{V}$. For every $\pi \in \mathcal{M}(P, \mathcal{V})$, one has $\pi(K_Y \times K_U) \geq \max(P(K_Y) + v(K_U) - 1, 0)$ (Fréchet-Hoeffding lower bound), thus $\pi(K_Y \times K_U) \geq 1 - \epsilon$.

Proof of Theorem 1 Suppose $\inf_{\pi \in \mathcal{M}(P,\mathcal{V})} E_\pi \left[1 \{Y \notin G(U)\} \right] = 0$, we shall show that the infimum is actually attained. Let $\pi_n \in \mathcal{M}(P,\mathcal{V})$ a sequence of probability distributions of the joint couple (U, Y) such that

$$E_{\pi_n} \left[1 \{Y \notin G(U)\} \right] \to 0.$$

By Lemma 2, $\mathcal{M}(P,\mathcal{V})$ is uniformly tight, hence by Prohorov's theorem it is relatively compact. Consequently there exists a subsequence $\pi_{\varphi(n)} \in M(P,\mathcal{V})$ which is weakly convergent to π.

One has $\pi \in \mathcal{M}(P,\mathcal{V})$. Indeed, clearly $\pi_Y = P$, and by Assumption 2 the sequences of random variables $m(U_{\varphi(n)})$ are uniformly integrable, therefore by van der Vaart (1998, Theorem 2.20), one has $\pi_{\varphi(n)}[m(U_{\varphi(n)})] \to \pi[m(U)]$, thus $\pi[m(U)] = 0$. Therefore, $\pi \in \mathcal{M}(P,\mathcal{V})$.

By Assumption 3, the set $\{Y \notin G(U)\}$ is open, hence by the Portmanteau lemma (van der Vaart 1998, Lemma 2.2, formulation (v)),

$$\liminf \pi_{\varphi(n)} \left[\{Y \notin G(U)\} \right] \geq \pi \left[\{Y \notin G(U)\} \right]$$

thus $\pi \left[\{Y \notin G(U)\} \right] = 0$.

Proof of Theorem 2 We need to show that the following two optimization problems $(\mathcal{P})$ and $(\mathcal{P}^*)$ have finite solutions, and that they are equal.

$$(\mathcal{P}) : \sup_{(f,\lambda) \in \mathcal{C}^0 \times \mathbb{R}^{dm}} \langle P, f \rangle \quad \text{subject to } Lf \leq \delta - \lambda' m$$

and

$$(\mathcal{P}^*) : \sup_{(\pi,\gamma) \in \mathcal{M} \times \mathbb{R}^{dm}} \langle \pi, \delta \rangle \quad \text{subject to } L^*\pi = P, \ \pi \geq 0, \ \langle \pi, m \rangle = 0.$$

where $\mathcal{C}^0$ is the space of continuous functions of y and u, equipped with the uniform topology, its dual with respect to the scalar product $\langle Q, f \rangle = \int f \, dQ$ is the space $\mathcal{M}$ of signed (Radon) measures on $\mathcal{Y} \times \mathcal{U}$ equipped with the vague topology (the weak topology with respect to this dual pair), L is the operator defined by $L(f)(y, u) = f(y)$ for all u, and its dual L^* is the projection of a measure π on $\mathcal{Y}$, and the function δ is defined by $\delta(y, u) = 1\{y \notin G(u)\}$. Note that $\delta(y, u)$ is not continuous, and hence is not included in the dual of $\mathcal{M}$. However, since G has a closed graph, δ is lower semi-continuous, hence, so is the restriction of the function

$$\langle \delta, \pi \rangle := \int f \, d\pi$$

to non-negative measures, and the set of continuous functions such that $Lf \leq \delta$ is closed.

We now see that $(\mathcal{P}^*)$ is the dual program of $(\mathcal{P})$: indeed, we have

$$\sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \langle P, f \rangle \quad \text{subject to } Lf \leq \delta - \lambda'm$$

$$= \sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \inf_{\pi\geq 0,\ \pi\in\mathcal{M}} \langle P, f \rangle + \langle \pi, \delta - \lambda'm - Lf \rangle$$

$$= \sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \inf_{\pi\geq 0,\ \pi\in\mathcal{M}} \langle P, f \rangle + \langle \pi, \delta \rangle - \lambda'\langle \pi, m \rangle - \langle \pi, Lf \rangle$$

$$= \sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \inf_{\pi\geq 0,\ \pi\in\mathcal{M}} \langle P, f \langle +\langle \pi, \delta \rangle - \lambda'\langle \pi, m \rangle - \langle L^*\pi, f \rangle$$

$$= \sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \inf_{\pi\geq 0,\ \pi\in\mathcal{M}} \langle \pi, \delta \rangle - \lambda'\langle \pi, m \rangle + \langle P - L^*\pi, f \rangle,$$

and

$$\inf_{\pi\geq 0,\ \pi\in\mathcal{M}} \sup_{(f,\lambda)\in\mathcal{C}^0\times\mathbb{R}^{dm}} \langle \pi, \delta \rangle - \lambda'\langle \pi, m \rangle + \langle P - L^*\pi, f \rangle$$

$$= \inf_{(\pi,\gamma)\in\mathcal{M}\times\mathbb{R}^{dm}} \langle \pi, \delta \rangle \quad \text{subject to } \langle \pi, m \rangle = 0, \quad L^*\pi = P, \quad \pi \geq 0.$$

We now proceed to prove that the strong duality holds, i.e. that the infimum and supremum can be switched. Under condition (4), by Ekeland and Temam (1976, Proposition (2.3), p. 52), $(\mathcal{P})$ is stable. Hence, by Ekeland and Temam (1976, Proposition (2.2), p. 51), $(\mathcal{P})$ is normal and $(\mathcal{P}^*)$ has at least one solution. Finally, since $f \mapsto <P, f>$ is linear, hence convex and lower semi-continuous, by Ekeland and Temam (1976, Proposition (2.1), p. 51), the two programs are equal and have a finite solution.

References

Andrews, D., Berry, S., Jia, P.: Placing bounds on parameters of entry games in the presence of multiple equilibria (2003, unpublished manuscript)

Artstein, Z.: Distributions of random sets and random selections. Israel J Math **46**, 313 (1983)

del Barrio, E., Cuestas-Albertos, J., Matrán, C., Rodríguez-Rodríguez, J.: Tests of goodness-of-fit based on the l2 wasserstein distance. Ann Stat **27**, 1230–1239 (1999)

Berry, S.: Estimation of a model of entry in the airline industry. Econometrica **60**, 889–917 (1992)

Berry, S., Tamer, E.: Identification in models of oligopoly entry. In: Advances in Economics and Econometrics, pp. 46–85. London: Cambridge University Press (2006)

Bierens, H.: A consistent conditional moment test for functional form. Econometrica **58**, 1443–1458 (1990)

Bresnahan, T., Reiss, P.: Entry in monopoly markets. Rev Econ Stud **57**, 531–553 (1990)

Brock, B., Durlauf, S.: Discrete choice with social interactions. Rev Econ Stud **68**, 235–265 (2001)

Chernozhukov, V., Hong, H., Tamer, E.: Estimation and confidence regions for parameter sets in econometric models. Econometrica **75**, 1243–1285 (2007)

Choquet, G.: Theory of capacities. Annales de l'Institut Fourier **5**, 131–295 (1954)

Ciliberto, F., Tamer, E.: Market structure and multiple equilibria in airline markets (2006, unpublished manuscript)

Dudley, R.: Real Analysis and Probability. London: Cambridge University Press (2002)

Ekeland, I., Temam, R.: Convex Anaysis and Variational Problems. NY, USA: North Holland/Elsevier (1976)

Ford, L., Fulkerson, D.: A simple algorithm for finding maximal network flows and an application to the hitchcock problem. Can J Math **9**, 210–218 (1957)

Galichon, A., Henry, M.: Dilation bootstrap: a methodology for constructing confidence regions with partially identified models (2006, unpublished manuscript). Available from ssrn.com, id = 934442

Galichon, A., Henry, M.: Inference in models with multiple equilibria (2008, unpublished manuscript). Available from ssrn.com, id = 1134762

Galichon, A., Henry, M.: A test of non-identifying restrictions and confidence regions for partially identified parameters. J Econom (2008, forthcoming)

Galichon, A., Henry, M.: Universal power of Kolmogorov–Smirnov tests of under-identifying restrictions (2008, unpublished manuscript). Available from ssrn.com, id = 1123825

Hansen, L., Sargent, T.: Robust control and model uncertainty. Am Econ Rev **91**, 60–66 (2001)

Hitchcock, F.: The distribution of a product from several sources to numerous localities. J Math Phys **20**, 224–230 (1941)

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

Kantorovich, L.: On the translocation of mass. Doklady Academii Nauk SSSR **37**, 199–201 (1942)

Khan, S., Tamer, E.: Inference on randomly censored regression models using conditional moment inequalities (2006, unpublished manuscript)

Koopmans, T.: Optimum utilization of the transportation system. Econometrica **17**, 136–146 (1949)

Maccheroni, F., Marinacci, M., Rustichini, A.: Ambiguity aversion, robustness and the variational representation of preferences. Econometrica **74**, 1447–1498 (2006)

Manski, C.: Identification of endogenous social effects: the reflection problem. Rev Econ Stud **60**, 531–542 (1993)

Manski, C.: Partial identification in econometrics. New Palgrave Dictionary of Economics, 2nd edn. New York: Macmillan (2005)

Moulin, H.: Cooperative Microeconomics. Princeton: Princeton University Press (1995)

Pakes, A., Porter, J., Ho, K., Ishii, J.: Moment inequalities and their application (2004, unpublished manuscript)

Papadimitriou, C., Steiglitz, K.: Combinatorial Optimization: Algorithms and Complexity. NY, USA: Dover (1998)

Rachev, S., Rüschendorf, L.: Mass Transportation Problems. Theory, vol. I. Berlin: Springer (1998)

Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Berlin: Springer (1998)

Tamer, E.: Incomplete simultaneous discrete response model with multiple equilibria. Rev Econ Stud **70**, 147–165 (2003)

van der Vaart, A.: Asymptotic Statistics. London: Cambridge University Press (1998)

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

Zolotarev, V.: Modern Theory of Summation of Random Variables. Utrecht: VSP (1997)