Journal of Econometrics 152 (2009) 186–196

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A test of non-identifying restrictions and confidence regions for partially identified parameters[☆]

Alfred Galichon[a,*], Marc Henry[b]

[a] Département d'économie, École polytechnique, 91128 Palaiseau, France

[b] Département de sciences économiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada

ARTICLE INFO

Article history:
Available online 23 January 2009

JEL classification:
C10
C12
C13
C14
C52
C61

Keywords:
Partial identification
Mass transportation
Specification test

ABSTRACT

We propose an easily implementable test of the validity of a set of theoretical restrictions on the relationship between economic variables, which do not necessarily identify the data generating process. The restrictions can be derived from any model of interactions, allowing censoring and multiple equilibria. When the restrictions are parameterized, the test can be inverted to yield confidence regions for partially identified parameters, thereby complementing other proposals, primarily Chernozhukov et al. [Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243–1285].

0. Introduction

In several rapidly expanding areas of economic research, the identification problem is steadily becoming more acute. In policy and program evaluation (Manski, 1990) and more general contexts with censored or missing data (Molinari, 2003; Magnac and Maurin, 2008) and measurement error (Chen et al., 2005), ad hoc imputation rules lead to fragile inference. In demand estimation based on revealed preference (Blundell et al., 2008) the data is generically insufficient for identification. In the analysis of social interactions (Brock and Durlauf, 2007; Manski, 2004), complex strategies to reduce the large dimensionality of the correlation structure are needed. In the estimation of models with complex strategic interactions and multiple equilibria (Tamer, 2003; Andrews et al., unpublished manuscript; Pakes et al., unpublished manuscript), assumptions on equilibrium selection mechanisms may not be available or acceptable.

More generally, in all areas of investigation with structural data insufficiencies or incompletely specified economic mechanisms, the hypothesized structure fails to identify a unique possible generating mechanism for the data that is actually observed. Hence, when the structure depends on unknown parameters, and even if a unique value of the parameter can still be construed as the true value in some well defined way, it does not correspond in a one-to-one mapping with a probability measure for the observed variables. We then call the structural restrictions non-identifying. In other words, even if we abstract from sampling uncertainty and assume the distribution of the observable variables is perfectly known, no unique parameter but a whole set of parameter values (hereafter called identified set in the terminology of Manski (2005)) will be compatible with it.

Once a theoretical description of an economic system is given, a natural question to consider is whether the structure can be rejected on the basis of data on its observable components. Marschak and Andrews (1944) construct a collection of production functions that are compatible with structural restrictions and

[☆] This research was partly carried out while the first author was visiting the Bendheim Center for Finance, Princeton University and financial support from NSF grant SES 0350770 to Princeton University, from NSF grant SES 0532398, from the Program for Economic Research at Columbia University and from Chaire EDF-Calyon "Finance et Développement Durable" is gratefully acknowledged. We are grateful to Victor Chernozhukov, Pierre-André Chiappori, Guido Imbens and Bernard Salanié for encouragement, support and many helpful discussions. We also thank three anonymous referees, whose detailed and insightful comments helped significantly improve the paper, and we thank conference participants at Econometrics in Rio and seminar participants at Berkeley, Chicago, Columbia, École polytechnique, Harvard-MIT, MIT Sloane OR, Northwestern, NYU, Princeton, SAMSI, Stanford, the Weierstrass Institute and Yale for helpful comments (with the usual disclaimer).

* Corresponding author.
E-mail addresses: alfred.galichon@gmail.com, alfred.galichon@polytechnique.edu (A. Galichon), marc.henry@umontreal.ca (M. Henry).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.01.010

are not rejected by the data. We extend this approach within the general formulation of Koopmans and Reiersol (1950), who define a structure as the combination of a binary relation between observed socioeconomic variables (market entry, insurance coverage, winning bids in auctions, etc.) and unobserved ones (productivity shocks, risk level, or risk attitude, valuations or information depending on the auction paradigm, etc.) and a generating mechanism for the unobserved variables. This setup is employed by Roehrig (1988) and Matzkin (1994), who analyze conditions for nonparametric identification of structures where the endogenous observable variables are functions of unobservable variables and exogenous observable ones.

Here, following Jovanovic (1989), we allow the relation between observable and unobservable variables to be many-to-many, thereby including structures with multiple equilibria (when a value of the latent variables is associated with a set of values of the observable variables) and censored endogenous observable variables (where a value of the observable variable is associated with set of values of the latent variables). We do not strive for identification conditions, but rather for the ability to reject such structures that are incompatible with data, as in the original work of Marschak and Andrews (1944).

We show that such a goal can be attained in all generality (i.e. for any structure, involving discrete as well as continuous observable variables), through an appeal to the duality of mass transportation (see Villani (2003) for a comprehensive account of the theory). Given any set of (possibly non-identifying) restrictions on the relation between latent and observable variables, and given the distribution ν of latent variables, the structure thus defined is compatible with the true distribution P of the observable variables if and only if there exists a joint distribution with marginals P and ν and such that the restrictions are almost surely respected. Otherwise, the data could not have been generated in a such a way. We show that the latter condition can be formulated as a mass transportation problem (the problem of transporting a given distribution of mass from an initial location to a different distribution of mass in a final location while minimizing a certain cost of transportation, as originally formulated by Monge (1781)). We show that this optimization problem has a dual formulation, an empirical version of which is a generalized Kolmogorov–Smirnov test statistic. We base a test of the restrictions in the structure on this statistic, whose asymptotic distribution we derive, and approximate using the bootstrapped empirical process.

Once we have a test of the structure, we can form confidence regions for unknown parameters using the methodology of Anderson and Rubin (1949), which consists in collecting all parameter values for which the structure is not rejected by the test at the desired significance level. The construction of such confidence regions has been the focus of much research lately (see for instance the thorough literature review in Chernozhukov et al. (2007)). Unlike much of the econometric research on this issue, we do not restrict the analysis to models defined by moment inequalities. On the other hand, we consider structures in the sense of Koopmans and Reiersol (1950), and hence parametric distributions for the latent variables. This, however, is a common assumption in empirical work with game theoretic models, as exemplified by Andrews et al. (unpublished manuscript), Ciliberto and Tamer (unpublished manuscript), and more generally Ackerberg et al. (2007).

The paper is organized as follows: Section 1 is divided in four subsections. The first describes the setup; the second defines the hypothesis of compatibility of the structure with the data; the third explains how to construct a confidence region for the identified set, and the fourth reviews the related literature. Section 2 is divided in three subsections. The first subsection describes and justifies the generalized Kolmogorov–Smirnov test of compatibility of the structure with the data; the second shows consistency of the test, and the third investigates size properties of the test in a Monte Carlo experiment. Section 3 concludes.

1. Incomplete model specifications

1.1. Description of the framework

Consider the model of an economy which is composed of an observed variable Y and a latent, unobserved variable U. Formally, (Y, U) is a pair of random vectors defined on a common probability space. The pair (Y, U) has probability law π which is unknown. Y represents the variables that are observable, and U the variables that are unobservable. Y may have discrete and continuous components. Y may include variables of interest in their own right, and randomly censored or otherwise transformed versions of variables of interest. We call the law of the observable variables P. It is unknown, but the data available is a sample of independent and identically distributed vectors $(Y_1, \ldots, Y_n)$ with law P. U includes random shocks and other unobserved heterogeneity components. The law π of (Y, U) can be decomposed into the unconditional distribution P of Y and the conditional distribution of U given Y, namely $\pi_{U|Y}$. Throughout the paper it is supposed that $\pi_{U|Y}$ is unknown but fixed across observations.

The distribution of U is parameterized by a vector $\theta_1 \in \Theta_1$, where Θ_1 is an open subset of $\mathbb{R}^{d_1}$, and the law of U is denoted ν_{θ_1}. Finally, an economic model is given to us in the form of a set of restrictions on the vector (Y, U), which can be summarized without loss of generality by the relation $U \in \Gamma_{\theta_2}(Y)$ where Γ_{θ_2} is a many-to-many mapping, which is completely given except for the vector of structural parameters $\theta_2 \in \Theta_2$, where Θ_2 is an open subset of $\mathbb{R}^{d_2}$. θ_1 and θ_2 may contain common components. We call θ the combination of the two, so that $\theta \in \Theta$, with Θ an open subset of $\mathbb{R}^{d_\theta}$, and $d_\theta \leq d_1 + d_2$. From now on, we shall therefore denote the distribution of U by ν_θ and the many-to-many mapping by Γ_θ. In all that follows, we assume that Γ_θ is measurable (a very weak requirement which is defined in the Appendix), and has non-empty and closed values.

We are interested in testing the compatibility of the observed variables Y with the model described by (Γ, ν). A related question is set-inference in a parametric model $(\Gamma_\theta, \nu_\theta)$: a confidence region for θ can be obtained by inverting the specification test, namely retaining the values of θ which are not rejected. Note that if $\theta_2 = (\beta, \eta)$, where β are the parameters of interest and $\eta \in H$ are nuisance parameters, we can redefine the economic model restrictions as $U \in \Gamma_\beta(Y)$ where Γ_β is defined by $\Gamma_\beta(y) = \bigcup_{\eta \in H} \Gamma_{(\beta,\eta)}(y)$ for all $y \in \mathbb{R}^{d_y}$. Hence we can assume again without loss of generality that θ_2 is indeed the parameter of interest. As the main focus of the present paper is to derive a specification test, whenever there is no ambiguity we shall implicitly fix the parameter θ and drop it from our notations.

Example 1. A prominent example for this set-up is provided by the class of models defined by a static game of interaction. Consider a game where the payoff function for player j, $j = 1, \ldots, J$ is given by $\Pi_j(S_j, S_{-j}, X_j, U_j; \theta)$, where S_j is player j's strategy and S_{-j} is their opponents' strategies. X_j is a vector of observable characteristics of player j and U_j a vector of unobservable determinants of the payoff. Finally θ is a vector of parameters. Pure strategy equilibrium conditions define a many-to-many mapping Γ_θ from unobservable player characteristics U to observable variables $Y = (S, X)$. More precisely, $\Gamma_\theta(s, x) = \{u \in \mathbb{R}^J : \Pi_j(s_j, s_{-j}, x_j, u_j; \theta) \geq \Pi_j(s, s_{-j}, x_j, u_j; \theta),\ \text{for all } s \text{ and all } j\}$. When the strategies are discrete, this is the set-up considered by Andrews et al. (unpublished manuscript), Pakes et al. (unpublished manuscript), and Ciliberto and Tamer (unpublished manuscript).