Dilation Bootstrap

A methodology for constructing confidence regions with partially identified models

Alfred Galichon and Marc Henry

Harvard University and Columbia University

First draft: May 27, 2006
This draft[1]: October 3, 2006

Abstract

We propose a methodology for constructing confidence regions with partially identified models of general form. The region is obtained by inverting a test of internal consistency of the econometric structure. We develop a dilation bootstrap methodology to deal with sampling uncertainty without reference to the hypothesized economic structure, and apply a duality principle to reduce the dimensionality of the remaining deterministic problem. As a result, the confidence region becomes easily computable, and the methodology can be applied to the estimation of models with sample selection, censored observations and to games with multiple equilibria.

JEL Classification: C10, C12, C13, C15, C34, C52, C61.

Keywords: Partial identification, incomplete specification test, duality, dilation bootstrap, nearest neighbours.

[1]We thank Gary Chamberlain, Victor Chernozhukov, Pierre-André Chiappori, Ivar Ekeland, Rustam Ibragimov, Guido Imbens, Shakeeb Khan, Geert Ridder, Joe Romano, Bernard Salanié and seminar participants at Harvard for helpful comments. Financial support from NSF Grant SBR 9729559 is gratefully acknowledged by both authors. The first author also gratefully acknowledges financial support from the Conseil Général des Mines. Correspondence addresses: Department of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138, USA. galichon@fas.harvard.edu and Department of Economics, Columbia University, 420 W 118th Street, New York, NY 10027, USA. mh530@columbia.edu.

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 (Shaikh and Vytlacil (2005), Magnac and Maurin (2005)) and measurement error (Chen, Hong, and Tamer (2005)), ad hoc imputation rules lead to fragile inference. In demand estimation based on revealed preference (Blundell, Browning, and Crawford (2003)) the data is generically insufficient for identification. In the analysis of social interactions (Brock and Durlauf (2002), 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, Berry, and Jia (2003), Pakes, Porter, Ho, and Ishii (2004)), 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 data 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. 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) will be compatible with it.

In such cases, many traditional estimation and testing techniques become inapplicable and a framework for inference in incomplete models is developing, with an initial focus on estimation. A question of particular relevance in applied work is how to construct valid confidence regions for identified sets of parameters. Formal methodological proposals include Chernozhukov, Hong, and Tamer (2002), Imbens and Manski (2004), Andrews, Berry, and Jia (2003), Pakes, Porter, Ho, and Ishii (2004), Shaikh (2005), Beresteanu and Molinari (2006), Galichon and Henry (2006). Some of these contributions are reviewed in section 1.4. The picture is far from complete and there is space for alternative proposals, possibly computationally more efficient, more general or less conservative.

In the present work, we propose a methodology that not only reduces the computational burden and increases the scope of applicability relative to the methods above, but also contributes to a better understanding of inference in incomplete models by clearly distinguishing how to deal with sampling uncertainty on the one hand, and model uncertainty on the other. The key to this separation is to deal with sampling variability without any reference to the hypothesized structure, using a methodology we call *dilation bootstrap*. This consists in dilating each point in the space of observable variables in such a way that the empirical probability (which is known) of dilated sets dominates the true probability (which is unknown) of the original sets. The unknown true probability (i.e. the true data generating mechanism) is then removed from the analysis, and we can proceed as if the problem were purely deterministic.

This deterministic problem is that of determining whether a parameter value is compatible with a known distribution, and the resulting confidence region is the set of such parameters. In great generality (i.e. an economic structure defined by a correspondence between observable and unobservable variables, and moment conditions on the latter), we show that this requirement of compatibility can be construed as a programming problem, the dual of which provides a function of the parameters that takes value zero for compatible parameters, and is non-negative everywhere. Hence, the confidence region we propose is the set of zeros of a function that is relatively easy to compute with standard optimization techniques.

Section 1 defines the economic structure, the identified set and the confidence region. Section 2 defines the test statistic. Section 2.3 discusses the power of the test and the corresponding tightness of the confidence regions. Section 2.4 explains the dilation bootstrap procedure. The last section concludes.

1 General setup

1.1 Econometric structure specification

We define the economic structure as in Jovanovic (1989), who pioneered the study of empirical implications of multiple equilibria. Variables under consideration are divided into two groups. Latent variables, $u \in \mathcal{U} \subseteq \mathbb{R}^{d_u}$, are typically not observed by the analyst, but some of their components may be observed by the economic actors. Observable variables, $y \in \mathcal{Y}$, with $\mathcal{Y}$ a finite set or $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$, are observed by the analyst and the economic actors. P is the true data generating process for the observable variables, and ν the hypothesized data generating processes for the latent variables. The econometric structure under consideration is given by a relation between observable and latent variables, i.e. a subset of $\mathcal{Y} \times \mathcal{U}$, which we shall write as a correspondence from $\mathcal{Y}$ to $\mathcal{U}$ denoted by $\Gamma_\theta : \mathcal{Y} \rightrightarrows \mathcal{U}$, where $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ is a vector of unknown parameters. The distribution ν of the unobservable variables U is assumed to satisfy a set of moment conditions, namely

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

and we denote by $\mathcal{V}_\theta$ the set of distributions that satisfy (1) and $\mathcal{M}(P, \mathcal{V}_\theta)$ the set of probabilities on $(\mathcal{Y} \times \mathcal{U})$ with marginal P on $\mathcal{Y}$ and marginal on $\mathcal{U}$ in $\mathcal{V}_\theta$. Finally, we call $\mathfrak{M}(\theta, \mathcal{V}_\theta)$ the structure defined by the correspondence Γ_θ and the moment conditions underlying the definition of $\mathcal{V}_\theta$. We make the additional technical assumptions that Γ_θ is non-empty and closed-valued and measurable, i.e. for each open set $\mathcal{O} \subseteq \mathcal{U}$, $\Gamma_\theta^{-1}(\mathcal{O}) = \{y \in \mathcal{Y} \mid \Gamma_\theta(y) \cap \mathcal{O} \neq \varnothing\} \in \mathcal{B}_\mathcal{Y}$, with $\mathcal{B}_\mathcal{Y}$ and $\mathcal{B}_\mathcal{U}$ denoting the Borel σ-algebras of $\mathcal{Y}$ and $\mathcal{U}$ respectively.

Example 1. *Games with multiple equilibria. A simple class of examples is that of models defined by a set of rationality constraints. Suppose 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 the opponents' strategies. X_j is a vector of observable characteristics of player j and U_j a vector of unobservable determinants of the payoff whose distributions are only known to satisfy a set of moment conditions as in (1). Finally θ is a vector of parameters. Pure strategy Nash equilibrium conditions $\Pi_j(S_j, S_{-j}, X_j, U_j; \theta) \geq \Pi_j(S, S_{-j}, X_j, U_j; \theta)$, for all S define a correspondence Γ_θ from unobservable variables U to observable variables $Y = (S', X')'$.*

Example 2. *Model defined by moment inequalities. A special case of the specification above is provided by models defined by moment inequalities.*

$$\mathbb{E}(\varphi_i(Y;\theta)) \leq 0, \quad \varphi_i : \mathcal{Y} \to \mathbb{R}, \;\; i = 1, \ldots, d_\varphi, \tag{2}$$

where Y denotes the whole vector of observable variables, including explanatory variables. This is a special case of our general structure, where

$$\Gamma_\theta(y) = \{u \in \mathcal{U} : \; u_i \geq \varphi_i(y;\theta), \;\; i = 1, \ldots, d_u\},$$

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

Example 3. *Model defined by conditional moment inequalities.*

$$\mathbb{E}(\varphi_i(Y;\theta)|X) \leq 0, \quad \varphi_i : \mathcal{Y} \to \mathbb{R}, \;\; i = 1, \ldots, d_\varphi, \tag{3}$$

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

$$\mathbb{E}(\varphi_i(Y;\theta)1\{t_1 \leq X \leq t_2\}) \leq 0, \quad \varphi_i : \mathcal{Y} \to \mathbb{R}, \;\; i = 1, \ldots, d_\varphi, \tag{4}$$

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 2.*

Example 4. *Unobserved random censoring (also known as accelerated failure time) model. A continuous variable* $Z = \mu(X, \theta) + \epsilon$, *where* μ *is known up to a vector of unknown parameters* θ, *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.*

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

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 3.*

Finally we turn to an example of binary response, and a simple unconditional inequalities example, which we shall use as a pilot examples to illustrate each step of our procedure.

Pilot Example 1. A Binary Response Model: *The observed variables* Y *and* X *are related by* $Z = 1\{X\theta + \varepsilon \leq 0\}$, *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* $\Gamma_\theta : \mathcal{Y} \rightrightarrows \mathcal{U}$ *characterized by* $\Gamma_\theta(1, x) = \{x\} \times (-2, -x\theta]$ *and* $\Gamma_\theta(0, x) = \{x\} \times (-x\theta, 2]$. *The two moment restrictions are* $m_\pm(x, \varepsilon) = (1\{\varepsilon \leq 0\} - \eta)(1 \pm x)$.

Pilot Example 2. A System of Moment Inequalities Model: *We consider the set of moment conditions:* $E\varphi(Y, \theta) \leq 0$ *with* $\theta' = (\theta_1, \theta_2)$, $\varphi' = (\varphi_1, \varphi_2, \varphi_3)$ *and* $\varphi_1(y, \theta) = \theta_1 + \theta_2 y - 1$, $\varphi_2(y, \theta) = -\theta_1 y$, $\varphi_3(y, \theta) = -\theta_2 y$. *So* $\Gamma_\theta(u) = \{u \in \mathbb{R}^3 : u_i \geq \varphi(y, \theta)\}$, *and* $m(u, \theta) = u$.

		$\Pr(Z=1\|X=-1)$		
		$<\eta$	$=\eta$	$>\eta$
$\Pr(Z=1\|X=1)$	$<\eta$	$\varnothing$	$\{\theta<0\}$	$\{\theta<0\}$
	$=\eta$	$\{\theta>0\}$	$\{\theta\in\mathbb{R}\}$	$\{\theta<0\}$
	$>\eta$	$\{\theta>0\}$	$\{\theta>0\}$	$\varnothing$

Table 1: Identified regions for the binary response pilot example.

1.2 Definition of the identified set

We argue that the structure is internally consistent[1] for some value of the parameter if and only if the correspondence Γ_θ it defines is almost surely respected, i.e. if $U \in \Gamma_\theta(Y)$ with probability one for some underlying probability structure. It captures the idea that the true data-generating process is compatible with the hypothesized structure. This statement is made precise in the following definition:

Definition 1. *For a given θ, model $\mathfrak{M}(\theta, \mathcal{V}_\theta)$ is internally consistent if and only if there exists a joint probability $\pi \in \mathcal{M}(P, \mathcal{V}_\theta)$ such that $\pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$.*

This prompts the natural definition of the identified set as the set of parameters such that the econometric structure is internally consistent.

Definition 2. *The identified set is $\Theta_I = \{\ \theta \in \Theta : \mathfrak{M}(\theta, \mathcal{V}_\theta)$ is internally consistent $\}$.*

Pilot example 1 continued One has $\Pr(Z=1|X) = \Pr(\varepsilon \leq -X\theta)$. Suppose $\theta > 0$. Then $\Pr(Z=1|X=1) = \Pr(\varepsilon \leq -\theta|X=1) \leq \Pr(\varepsilon \leq 0|X=1) = \eta$, and similarly, $\Pr(Z=1|X=-1) \geq \eta$. Symmetrical results hold for $\theta < 0$. The resulting identified regions for θ are summarized in table 1. This illustrates the fact that θ is *set-identified*: in this example, only a set of θ is identifiable, depending on features of distribution of (X,Z).

Example 5. *In the case of a model defined by moment inequalities as in example 2, the identified set is simply the set of θ such that the inequalities are satisfied, i.e. $\Theta_I = \{\theta \in \Theta :\ \mathbb{E}(\varphi_i(Y;\theta)) \leq 0,\ \varphi_i : \mathcal{Y} \to \mathbb{R}^{d_u},\ i = 1,\ldots,d_\varphi\}$*

So, in particular:

Pilot example 2 continued In this case, the identified region has the graphical representation given in figure 1.

[1]We owe the choice of terminology to Guido Imbens.

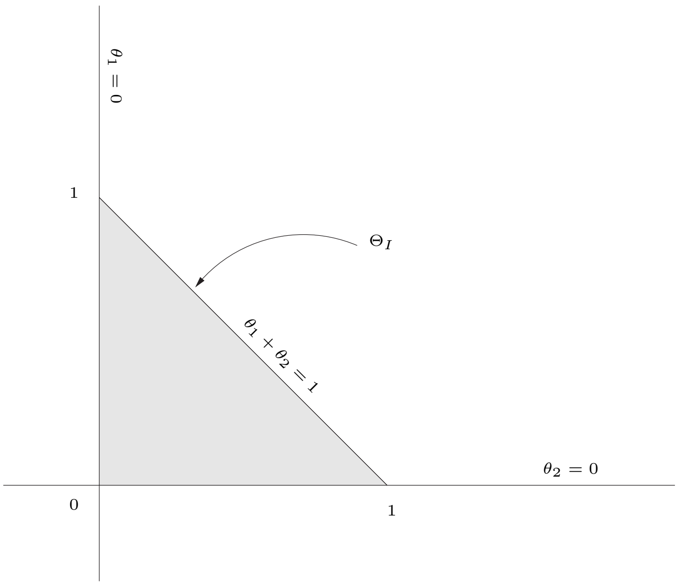

Figure 1: Identified region for the inequalities pilot example.

1.3 Confidence region for the identified set

Given a sample $(Y_1, \ldots, Y_n)$ of independently and identically distributed realizations of Y, our objective is to construct a sequence of random sets Θ_n^α such that for all $\theta \in \Theta_I$, $\lim_{n\to\infty} \Pr(\theta \in \Theta_n^\alpha) \geq 1-\alpha$. In other words, we are concerned with constructing a region Θ_n^α that covers each value of the identified set, as opposed to a region $\tilde{\Theta}$ that covers the identified set uniformly, i.e such that $\Pr(\Theta_I \subseteq \tilde{\Theta}) \geq 1-\alpha$. We do so (as in Chernozhukov, Hong, and Tamer (2002), Imbens and Manski (2004) and others) by including in Θ_n^α all the values of θ such that we fail to reject a test of internal consistency of $\mathfrak{M}(\theta, \mathcal{V}_\theta)$ with asymptotic level at least $1-\alpha$.

We shall demonstrate the construction of a test statistic $T_n^\alpha(\theta, \mathcal{V}_\theta)$ such that, conditionally on the structure $\mathfrak{M}(\theta, \mathcal{V}_\theta)$ being internally consistent, the probability that $T_n^\alpha(\theta, \mathcal{V}_\theta) = 0$ is at least $1-\alpha$ asymptotically, i.e.

$$\lim_{n\to\infty} \Pr\left(T_n^\alpha(\theta, \mathcal{V}_\theta) = 0 \mid \mathfrak{M}(\theta, \mathcal{V}_\theta) \text{ is internally consistent}\right) \geq 1-\alpha. \tag{5}$$

Hence we define our confidence region in the following way.

Definition 3. *The* $(1-\alpha)$ *confidence region for* Θ_I *is* $\Theta_n^\alpha = \{\theta \in \Theta : T_n^\alpha(\theta, \mathcal{V}_\theta) = 0\}$.

The full procedure is summarized in table 2. It is clear from equation 5 and the above definition that our confidence region covers each element of the identified set with proba-

Table 2: **SUMMARY OF THE PROCEDURE**

1. *Draw B bootstrap samples $(Y_1^b, \ldots, Y_n^b)$, and for each b and for each l, define W_l^b as the matrix with (i,j)-th entry equal to 1 if Y_j^b is among the l nearest neighbours of Y_i in Euclidian distance, and zero otherwise. Call l^b the smallest value of l such that $\min\{trace(W_l^b \Pi) : \Pi \text{ permutation}\} = 0$. Order the l^b and call l_n^α the $[B\alpha]$ largest, and take J_n^α to be the correspondence that to each sample point Y_j associates the l_n^α nearest neighbours of Y_j within the initial sample $(Y_1, \ldots, Y_n)$.*

2. *For each j, minimize $1\{u \notin \Gamma_\theta(J_n^\alpha(Y_j))\} - \lambda' m(u, \theta)$ over u, keeping λ fixed. Call the minimum $h_\lambda(Y_j)$. Maximize $\sup_\lambda \frac{1}{n} \sum_{j=1}^n h_\lambda(Y_j)$ over λ.*

3. *Define the confidence region Θ_n^α as the set of θ's for which the previous maximization returns zero.*

bility at least $1 - \alpha$ asymptotically. Hence, after a section devoted to discussing in detail our contribution within the literature on this question, the remainder of this paper will be concerned with the construction of the statistic $T_n^\alpha(\theta, \mathcal{V}_\theta)$ with the required property 5.

1.4 Review of the literature

The first general purpose method to construct confidence regions with partially identified models was proposed by Chernozhukov, Hong, and Tamer (2002). The idea is to extend M-estimation to cases where the criterion function $Q(\theta)$ is null on a set of parameter values (as opposed to a unique parameter value as in identified models) and non negative everywhere. The identified set Θ_I is simply defined as the set of parameter values for which $Q(\theta)$ is zero. The procedure consists in approximating the limiting distribution of the suitably normalized version of $\sup_{\theta \in \Theta_I} Q_n(\theta)$, where Q_n is the empirical criterion function, using an initial estimate of Θ_I and either a bootstrap, a sub-sampling or a simulation procedure (based on the asymptotic distribution), and constructing the $1 - \alpha$ confidence region for Θ_I using the α quantiles obtained with the approximation.

Although this approach is very elegant and very general in scope, it gives no theoretical guidance as to the choice of criterion function in a given situation, in contrast with the M-estimation literature which is motivated by a need for robust alternatives to maximum likelihood. Chernozhukov, Hong, and Tamer (2002) do propose a choice of Q in several leading examples, but it is difficult to asses the influence of this choice on the size of the confidence regions obtained within their framework.

One proposal to fill this gap appears in Galichon and Henry (2006). Within the framework presented in the present paper, except for the fact that the distribution of the unobservable variables is parametric, viz. ν_θ instead of $\nu \in \mathcal{V}_\theta$ here, they show that the null hypothesis of internal consistency of the econometric structure for any given parameter value θ is equivalent to $P(B) \leq \nu_\theta \Gamma_\theta(B)$ for all sets B in a class of sets $\mathcal{C}$ that they call core-determining (meaning that $P(B) \leq \nu_\theta \Gamma_\theta(B)$ for all $B \in \mathcal{C}$ implies $P(B) \leq \nu_\theta \Gamma_\theta(B)$ for all measurable sets B). Hence the $\inf_{B \in \mathcal{C}}[\nu_\theta(\Gamma_\theta(B)) - P(B)]$ is a natural choice of criterion function Q, with empirical version obtained by replacing P by P_n in the latter expression, thereby obtaining a generalized Kolmogorov-Smirnov statistic. Galichon and Henry (2006) show that the latter statistic can also be used directly to derive confidence regions that cover each element of Θ_I with a prescribed probability. Indeed they derive the asymptotic distribution of $\inf_{B \in \mathcal{C}}[\nu_\theta(\Gamma_\theta(B)) - P_n(B)]$ and propose to form the $1 - \alpha$ confidence region with all the values of the parameter such that the hypothesis of internal consistency of the structure is not rejected at the $1 - \alpha$ level of significance.

Andrews, Berry, and Jia (2003) also propose a heuristic for testing the inequality $P(B) \leq \nu_\theta \Gamma_\theta(B)$, and hence were the first to apply a Kolmogorov-Smirnov specification testing type of approach to problems of partial identification. However, they restrict attention to $\mathcal{Y}$ finite (as in the case of games with discrete strategies) and singleton B's. In other words, they test the sufficient conditions $P(\{y\}) \leq \nu_\theta \Gamma_\theta(\{y\})$ for each observable value y in the finite set $\mathcal{Y}$. Although it is natural to restrict attention to $\mathcal{Y}$ finite, since structures arising from games with discrete strategies are more likely to give rise to identification failures, restricting attention to the singletons leads to very conservative inference, since $\nu_\theta \Gamma_\theta(.)$ is an alternating capacity (i.e. a set function that satisfies $\nu_\theta(\Gamma_\theta(B \cup C)) \leq \nu_\theta(\Gamma_\theta(B)) + \nu_\theta(\Gamma_\theta(C))$ for all $B \cap C = \varnothing$).

The present contribution extends the results of Galichon and Henry (2006) to the case where no parametric assumption is made on the unobservable variables U, thereby including as a special case structures defined by moment inequalities. It also improves on the Galichon and Henry (2006) proposal very significantly. Indeed, the asymptotic distribution of the test statistic in the latter is not distribution free, which means that quantiles must be approximated via the simulation of a Brownian bridge or via a sub-sampling procedure, both of which are computationally costly and require some ad-hoc choices of deterministic sequences that are unpalatable to practitioners and always leave open the way for data-snooping. To avoid these drawbacks, we propose a (computationally economical) dilation bootstrap procedure which controls sample variability without any reference to the hypothesized econometric structure under scrutiny. Credit also goes to the pioneering insight of Andrews, Berry, and Jia (2003), who propose a procedure in a similar spirit. However, since they restrict attention to singletons, theirs is a simple bootstrap, whereas the dilation bootstrap presented here is a genuine statistical innovation.

After that dilation bootstrap is performed (once and for all, since it is independent of θ), one can proceed as if everything was known, so the remainder of the construction of the confidence region is a deterministic optimization. In addition to the obvious practical advantages, we believe this approach to be the most natural in cases where the hypothesized structure incompletely specifies the true data generating process. Note also that since the dilation bootstrap operates without reference to the parameterized structure, it is free of the drawbacks documented in Andrews (2000).

2 Internal consistency test statistic

This section is concerned with constructing the statistic $T_n^\alpha(\theta, \mathcal{V}_\theta)$ that satisfies requirement 5. In other words, we set out to construct a conservative test statistic for the hypothesis of internal consistency

$$\mathrm{H}_0(\theta): \qquad \exists \pi \in \mathcal{M}(P, \mathcal{V}_\theta): \ \pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0,$$

with rejection region $\mathbb{R}\backslash\{0\}$. the definition of alternatives will be discussed in section 2.3.

We define our test statistic and present our formal results before proceeding to a heuristic discussion of the principles involved. In all that follows, P_n refers to the empirical distribution of the sample $(Y_1, \ldots, Y_n)$, which gives mass $1/n$ to each of the sample points.

2.1 Internal consistency as an optimization problem

The null hypothesis of internal consistency $\mathrm{H}_0(\theta)$ is equivalent to the optimization problem

$$\min_{\pi \in \mathcal{M}(P, \mathcal{V}_\theta)} : \pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0.$$

This program is not computationally workable as such as it requires optimizing over an infinite-dimensional space. Fortunately, as shown in Step 2 of the proof of Theorem 1, this problem has a finite-dimensional dual formulation

$$\sup_{\lambda \in \mathbb{R}^{d_m}} \mathbb{E}_P \left[g_{\lambda,\theta}(Y) \right] = 0, \quad \text{where} \quad g_{\lambda,\theta}(y) = \inf_u \left(1\{u \notin \Gamma_\theta(y)\} - \lambda' m(u;\theta) \right),$$

and we use this dimensionality reduction principle to construct a feasible test statistic. This is illustrated with our pilot example:

Pilot example 1 continued Here, we have $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ and

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

Since we have seen that only the sign of θ can be identified, we assume that $\theta \in \Theta = \{-1,0,1\}$. We show at the end of the appendix that $\sup_{\lambda\in\mathbb{R}^2} \mathbb{E}_P(h_{\lambda,\theta=-1}(Z)) = 0$ if and only if $\eta \geq \Pr_{Z|X}(1|-1)$ and $\eta \leq \Pr_{Z|X}(1|1)$.

2.2 Definition of the test statistic

To define our test statistic, we need an auxiliary construction called a dilation. Call J_n^α a sequence of correspondences $J_n^\alpha : \mathcal{Y} \rightrightarrows \mathcal{Y}$ that satisfies

$$\Pr\left(\forall A \in \mathcal{B}_{\mathcal{Y}}:\ P(A) \leq P_n(J_n^\alpha(A))\right) \to 1-\alpha, \tag{6}$$

where $\mathcal{B}_{\mathcal{Y}}$ denotes the class of (Borel) measurable subsets of $\mathcal{Y}$.

We define the test statistic as follows:

$$T_n^\alpha(\theta,\mathcal{V}_\theta) = \sup_\lambda \left(\frac{1}{n}\sum_{j=1}^n h_\lambda(Y_j)\right) \text{ with } h_\lambda(y) = \inf_u \left(1\{u \notin \Gamma_\theta(J_n^\alpha(y))\} - \lambda' m(u,\theta)\right). \tag{7}$$

Pilot example 2 continued In that case, $J_n^\alpha(y) = [y-\delta_n^\alpha(y), y+\delta_n^\alpha(y)]$, where for each j, $\delta_n^\alpha(Y_j)$ is the distance between Y_j and its neighbour of rank l_n^α. The latter is determined with the algorithm of section 2.4.2 based on B bootstrap replications.

Defining

$$\begin{aligned} H(y) := \Gamma_\theta(J_n^\alpha(y)) = \{(u_1,u_2,u_3)' : \quad & u_1 \geq -\theta_1(y+\delta_n^\alpha(y)) \\ & u_2 \geq -\theta_2(y+\delta_n^\alpha(y)); \\ & u_3 \geq -1+\theta_1+\theta_2(y-\delta_n^\alpha(y))\}, \end{aligned}$$

we can write

$$h_\lambda(y) = \min\left(\inf_{u\in H(y)} -\lambda' u, \inf_{u\notin H(y)} (1-\lambda' u)\right),$$

so that the inner optimization, i.e. the computation of h_λ, is a simple linear programming problem, and the outer optimization, i.e. the maximization of $\frac{1}{n}\sum_{j=1}^n h_\lambda(Y_j)$, is a standard non-linear optimization problem.

We can now state the theorem that validates the construction of our confidence region.

Theorem 1. *Suppose the following assumptions hold: (A1) ν is absolutely continuous with respect to Lebesgue measure, and so is P when $\mathcal{Y} = \mathbb{R}^{d_y}$ and (A2) J_n^α satisfies requirement 6. Then* $\lim_{n\to\infty} Pr(T_n^\alpha(\theta, \mathcal{V}_\theta) = 0) \geq 1-\alpha$ *for all* $\theta \in \Theta_I$*. Hence,* $\lim_{n\to\infty} Pr(\theta \in \Theta_n^\alpha) \geq 1-\alpha$ *for all* $\theta \in \Theta_I$*.*

The formal proof is in the appendix. We give a detailed heuristic to explain the principle of the test. The hypothesis of internal consistency of the structure is that $U \in \Gamma_\theta(Y)$ almost surely, where U is the unobservable variable, and Y is the observable variable. A suitable choice of J_n^α ensures that, with probability larger than $(1-\alpha)$ asymptotically, we have $Y \in J_n^\alpha(Y^*)$, where Y^* is a bootstrap variable (i.e. a variable with distribution P_n), hence that $U \in \Gamma_\theta(J_n^\alpha(Y^*))$. Call the latter statement $(\mathcal{S}_n^\alpha)$. Since $(\mathcal{S}_n^\alpha)$ does not depend on the unknown data generating process P, there only remains to determine whether $(\mathcal{S}_n^\alpha)$ is true or not. Formally, $(\mathcal{S}_n^\alpha)$ states the existence of a joint law $\pi \in \mathcal{M}(P_n, \mathcal{V}_\theta)$ such that $\pi(\{(u,v) : u \notin \Gamma_\theta(J_n^\alpha(v))\}) = 0$. Hence $(\mathcal{S}_n^\alpha)$ can be written as the solution of a programming problem $\min_{\pi \in \mathcal{M}(P_n,\mathcal{V}_\theta)} \pi(\{(u,v) : u \notin \Gamma_\theta(J_n^\alpha(v))\}) = 0$. Since $T_n^\alpha(\theta, \mathcal{V}_\theta)$ is the dual formulation of this programming problem, the level of the test of the internal consistency hypothesis based on the rejection region $\{T_n^\alpha(\theta, \mathcal{V}_\theta) \neq 0\}$ is larger than $1-\alpha$ asymptotically.

2.3 Power of the test against fixed alternatives

Remember that for a fixed value of the parameter vector θ, our null hypothesis of internal consistency is the following:

$$\mathrm{H}_0(\theta): \qquad \exists \pi \in \mathcal{M}(P, \mathcal{V}_\theta): \ \pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$$

which is equivalent to

$$\min_{\pi \in \mathcal{M}(P,\mathcal{V}_\theta)} : \pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$$

We need to distinguish two kinds of fixed alternatives, the

- *Quasi-consistent* alternatives

$$\mathrm{H}_{\mathrm{QC}}(\theta): \qquad \inf_{\pi \in \mathcal{M}(P,\mathcal{V}_\theta)} : \pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$$

  but the infimum is not attained,

- *Inconsistent* alternatives

$$\mathrm{H}_{\mathrm{IC}}(\theta): \qquad \inf_{\pi\in\mathcal{M}(P,\mathcal{V}_\theta)} : \pi(\{(u,y)\in\mathbb{R}^{d_y}\times \mathrm{R}^{d_u} : u\notin\Gamma_\theta(y)\})>0.$$

2.3.1 *Quasi-consistent* alternatives

Since our procedure is insensitive to the fact that the infimum is attained or not, we need to make sure quasi-consistent alternatives are ruled out. The following example shows that the infimum is not always attained.

Example 6. *Let $P = N(0,1)$, $\mathcal{U} = \mathcal{Y} = \mathbb{R}$, $\mathcal{V}_\theta = \{\nu : \mathbb{E}_\nu(U) = 0\}$, and $\Gamma_\theta(y) = \{1\}$ for all $y\in\mathcal{Y}$, and consider the distribution $\pi_m = P\otimes\nu_m$ such that $\nu_m(\{1\}) = 1-1/m$, and $\nu_m(\{1-m\}) = 1/m$. The π_m probability of $U\notin\Gamma_\theta(y)$ is $1/m$ which indeed tends to zero as $m\to\infty$, but it is clear that there exists no distribution ν which puts all mass on $\{1\}$ and has expectation 0.*

It is clear from example 6 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 $m(u,\theta)$ need to be bounded. In all that follows, we assume without loss of generality that $\mathcal{U}\subset\mathbb{R}^{d_u}$ is actually the domain of U.

Assumption 1. *For all $\theta\in\Theta$, $\lim_{M\to\infty}\sup_{\nu\in\mathcal{V}_\theta}\nu\left[\| m(U,\theta)\| 1_{\{\|m(U,\theta)\|>M\}}\right]=0$.*

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

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

Example 7. *In example 2, by Theorem 1.6 page 9 of Rockafellar and Wets (1998), we know that assumption 3 is satisfied when the moment functions φ_j, $j=1,\dots,d_\varphi$ are lower semi-continuous.*

We can now state the result:

Theorem 2. *Under assumptions 1, 2 and 3, H_0 implies H_{IC}, ie. H_{QC} is ruled out.*

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

Pilot example 1 continued Here, U=[-2,2] is compact, and $m_{\pm}(\varepsilon, x) = [1_{\{\varepsilon \leq 0\}} - \eta](1 \pm x)$ are bounded. Thus the admissible set of measures $\mathcal{M}(P, \mathcal{V}_\theta)$ is uniformly tight, and we can replace the "min" by an "inf" in the expression of H_0.

2.3.2 *Inconsistent* alternatives

Now that we have ruled out quasi-consistent alternatives that our procedure is insensitive to, we give conditions for our procedure to have power against inconsistent alternatives, i.e. that our test rejects inconsistent alternatives with probability tending to one.

First, this requires that the dilation used for the confidence region shrinks to the identity mapping at a suitable rate to ensure that

$$\Pr\left(\min_{\pi \in \mathcal{M}(P_n, \mathcal{V}_\theta)} \pi\{(v,u) : u \notin \Gamma_\theta(J_n^\alpha(v))\} = 0 \mid \mathrm{H}_{\mathrm{IC}}\right) \to 0 \tag{8}$$

for almost all samples and for all $\theta \in \Theta \backslash \Theta_I$.

Second, to ensure that the dual formulation used in the test statistic is not smaller than the primal program, we need an additional assumption (generally called a *Slater condition* in the optimization literature):

Assumption 4. *For almost all samples* $(Y_1, \ldots, Y_n)$*, there exists a measurable function* g *and a vector* λ *such that for all* $(u,v) \in \mathcal{U} \times \{Y_1, \ldots, Y_n\}$*,* $g(y) + \lambda' m(u, \theta) < 1\{u \notin \Gamma_\theta(J_n^\alpha(y))\}$*.*

Remark 2. *This condition is an interior condition, i.e. it ensures there exists a feasible solution to the optimization problem in the interior of the constraints.*

Theorem 3. *Under 8 and assumptions 1, 2, 3 and 4,* $Pr(T_n^\alpha(\theta, \mathcal{V}_\theta) = 0) \to 0$ *for each* $\theta \in \Theta \backslash \Theta_I$*.*

Remark 3. *As described in the appendix, this 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.*

2.4 Dilation Bootstrap

2.4.1 Dilation

We now turn to the question of how to construct the dilation J_n^α that satisfies requirement 6. The idea is to dilate each point in space in such a way that we account for all sampling variability before introducing any reference to the hypothesized structure $\mathfrak{M}(\theta, \mathcal{V}_\theta)$. The suitable dilation is determined with an appeal to the empirical bootstrap principle. We use the discrepancy between the empirical distribution P_n and the bootstrap distribution P_n^* (the empirical distribution of a sample of n independent and identically distributed variables with distribution P_n) to approximate the discrepancy between the true distribution of observable variables P and the empirical distribution P_n. Hence we choose J_n^α such that:

$$\Pr\left(\forall A \in \mathcal{B}_{\mathcal{Y}} : P_n^*(J_n^\alpha(A)) \geq P_n(A)\right) \rightarrow 1 - \alpha \tag{9}$$

in probability conditionally on the original sample $(Y_1, \ldots, Y_n)$.

2.4.2 Dilation Bootstrap Procedure in the continuous case

Consider B iid samples $(Y_1^b, \ldots, Y_n^b)$, $b = 1, \ldots, B$ drawn from P_n and call P_n^b the empirical distribution of sample b.

1. For each bootstrap replication b and for each l, define:

$$W_{i,j}^b(l) = 1\{Y_j^b \text{ is not among the } l \text{ nearest neighbours of } Y_i \ (Y_i \text{ included})\}$$
$$c^b(l) = \min\{\text{trace}(W_l^b \Pi) : \Pi \text{ is a permutation matrix}\}$$
$$l^b = \min\{l : c^b(l) = 0\}$$

2. Let l_n^α be the $[B\alpha]$ largest among the l^b, $b = 1, \ldots, B$, and take J_n^α to be the correspondence that to each sample point Y_j associates the l_n^α nearest neighbours (in Euclidian distance) of Y_j (itself included) within the initial sample $(Y_1, \ldots, Y_n)$.

Discussion:

The matrix $W_l^b = (W_{i,j}^b(l))_{i,j=1}^n$ is a matrix of 0-1 weights which penalizes the pairs (Y_i, Y_j^b) where Y_i and Y_j^b are to far apart in nearest neighbour ranking. The quantity $\text{trace}(W_l^b \Pi)$ is the number of pairs that are more than l neighbours apart in the one-to-one matching of initial sample points $Y_1 \ldots, Y_n$ with bootstrap sample points $Y_1^b, \ldots, Y_n^b$ defined by the permutation matrix Π. The cost function $c^b(l)$ is the minimum matching cost among all possible matches (i.e. all possible permutation matrices).

A cost $c^b(l) = 0$ means that there is a matching where the pairs (Y_i, Y_j^b) are never more than l neighbours apart. Hence, if we define $J_{n,l}$ as the correspondence $y \rightrightarrows \{y$ and its $l-1$ nearest neighbours$\}$, we have for any $F \subseteq \mathcal{Y}$,

$$P_n^b(J_{n,l}(F)) \geq P_n(F)$$

since by construction $J_{n,l}(F)$ contains at least as many bootstrap draws as F contains original draws. As a result, by choosing J_n^α as in step 2 of the bootstrap procedure above, we ensure that for a proportion $1-\alpha$ of the bootstrap samples, we have $P_n^b(J_n^\alpha(F)) \geq P_n(F)$ for all $F \subseteq \mathcal{Y}$.

Remark 4. *Although the problem of finding the minimum cost matching (called the* assignment problem *or* marriage problem*) is very familiar to economists, as far as we know, its application within a bootstrap procedure is unprecedented. The best known algorithms to find the optimal assignment require $O(n^3)$ operations, which is computationally the most demanding element in our bootstrap procedure*[2] *(whose formal complexity is in $O(n^3 \ln n)$).*

2.4.3 Bootstrap Quantile

In the case where $\mathcal{Y}$ is the real line, the optimal assignment can be obtained in the following way (known as the *greedy algorithm*): for each order statistic $Y_{(i)}$ of the original sample, starting with the smallest $Y_{(1)}$, find the closest bootstrap draw that has not been matched yet, and match it with $Y_{(i)}$. Lemma 4 in the appendix justifies this construction and this optimal match has the simple graphical representation given in figure 2.

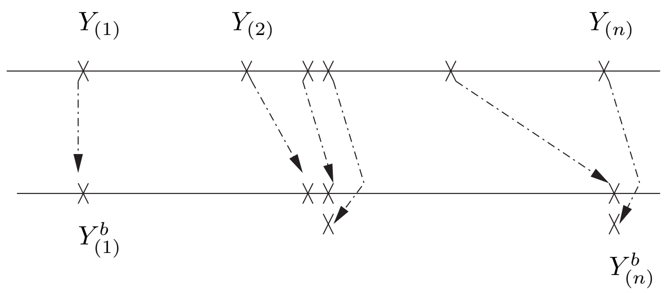

Figure 2: Optimal matching. The maximum number of neighbours is $l^b = 4$ since the longest arrow visible on the graph links a point with its third nearest neighbour in Euclidian distance.

[2]We propose a Matlab implementation of the whole bootstrap procedure, where the assignment procedure is borrowed from Markus Buehren's implementation of the Kuhn-Munkrés algorithm (also known as the Hungarian method, see Papadimitriou and Steiglitz (1982)). One bootstrap iteration takes about 15 minutes for $n = 1000$ on an AMD Opteron (tm) Processor 250 with 4G of RAM. Since the dilation bootstrap needs to be performed only once on a given data set, this seems very reasonable, especially if one uses parallel processing.

As shown in lemma 4 and seen in figure 2, where the $Y_{(j)}$'s $j = 1 \ldots, n$ are the order statistics of the original sample, and the $Y^b_{(j)}$'s $j = 1 \ldots, n$ are the order statistics of the b-th bootstrap sample, the matching that minimizes the maximum index distance between matched points is the bootstrap quantile function. Hence, by (5.3.1) page 163 of Csörgő and Révész (1981), the number of nearest neighbours $l_{n,i}$ is asymptotically equivalent to $(n \ln \ln n/2)^{1/2}$, which gives an upper bound on the rate at which the correspondence J_n^α shrinks to the identity.

2.4.4 Dilation Bootstrap procedure in the discrete case

The dilation bootstrap procedure outlined above relies on Euclidian geometry for the nearest neighbour matching. Hence it is not directly applicable to the case where the observable variable Y has a discrete distribution, or equivalently, when $\mathcal{Y}$ is a finite set, as in our pilot example 1. However, we can embed the discrete case within the continuous case using an extraneous randomization device, as we now explain.

Label the elements of $\mathcal{Y}$ so that $\mathcal{Y} \subseteq \mathbb{N}$. Let V be a uniformly distributed random variable on $[-\frac{1}{2}, \frac{1}{2}]$ independent of Y and U. Consider $\vartheta : \mathcal{Y} \times [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ defined by $\vartheta(y, v) = y + v$. Call $\tilde{Y} = \vartheta(Y, V)$ and $\tilde{U} = (U, V)$. Define $\Upsilon_\theta : \mathcal{U} \times [-\frac{1}{2}, \frac{1}{2}] \rightrightarrows \mathbb{R}$ by

$$\Upsilon_\theta(u, v) = \{\vartheta(y, v) : y \in \Gamma_\theta^{-1}(u)\}.$$

Finally, call $\tilde{\Gamma}_\theta = \Upsilon_\theta^{-1}$. Lemma 1 below, the proof of which is immediate, justifies using our procedure in table 2 with the Y replaced by $\tilde{Y}$, U by $\tilde{U}$, Γ_θ replaced by $\tilde{\Gamma}_\theta$ and m_θ by $\tilde{m}_\theta : (u, v) \mapsto (m(u), v)$.

Lemma 1. *For any $\theta \in \Theta$, $\tilde{U} \in \tilde{\Gamma}_\theta(\tilde{Y})$ is almost surely equivalent to $U \in \Gamma_\theta(Y)$.*

The idea is to embed the set $\mathcal{Y}$ into the richer set $\mathbb{R}$ in order to employ our dilation bootstrap procedure. As a result of lemma 1, we can treat the discrete case with the same dilation bootstrap procedure, as long as we suitably redefine the structure to take into account for the extraneous randomization device.

Conclusion

In this paper we contribute to the debate on estimation of partially identified models with a new proposal for constructing confidence regions for partially identified parameters. We argue that our method has the following advantages over existing approaches:

- It clearly distinguishes the treatment of sampling uncertainty and model uncertainty.
- It deals with the former with a (computationally reasonable) dilation bootstrap procedure that is performed once and for all on the data set, irrespective of the hypothesized economic structure to be evaluated.
- As for model uncertainty, it is treated with a duality principle, which allows to reduce the dimensionality of the optimization problem implicit in the definition of the identified region. As a result, the complexity of this part of the procedure is commensurate with the complexity of the hypothesized economic structure itself.
- The procedure involves no choice of deterministic sequence as do the sub-sampling and the adaptive approaches.

Since it is based on searching over the parameter space for values for which our statistic is equal to zero, it is still only applicable for low dimensionality parameter spaces. A method for dealing with nuisance parameters still has to be developed for the method to be applicable in a wider context.

Appendix

Proof of Theorem 1 :

Discussion and reader's guide to the proof

The key to the construction of the internal consistency test statistic is the observation that the hypothesis of internal consistency of definition 1 can be seen as an optimization problem: indeed, the following statements are trivially equivalent.

(i) There exists a joint probability $\pi \in \mathcal{M}(P, \mathcal{V}_\theta)$ such that $\pi(\{(u, y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$,

(ii) $\min_{\pi \in \mathcal{M}(P,\mathcal{V}_\theta)} : \pi(\{(u, y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$.

By duality, statement (ii) above can be shown to imply

(iii) $\sup_\lambda \int g_{\lambda,\theta}(y) dP(y) = 0$ where $g_{\lambda,\theta}(y) = \inf_u (1\{u \notin \Gamma_\theta(y)\} - \lambda' m(u))$.

One route (which we take in a companion note available on request) is to directly consider the empirical equivalent of formulation (iii), i.e. $\sup_\lambda \sum_{j=1}^n g_{\lambda,\theta}(Y_j)/n$ and show that with a root-n re-scaling, it converges to the supremum of a P-Brownian bridge over the sub-class comprising the functions $g_{\lambda,\theta}$ such that $\int g_{\lambda,\theta}(y)dP(y) = 0$. Then quantiles may be obtained for the test statistic with either a simulation or a sub-sampling procedure.

What we propose instead is to control sampling error prior to considering the optimization problem and taking its dual. Hence the idea to dilate each point in space using the correspondence $J_n^\alpha : \mathcal{Y} \rightrightarrows \mathcal{Y}$ in such a way that the probability of $Y \in J_n^\alpha(Y^*)$ when Y is distributed according to P and Y^* according to P_n is approximately $(1-\alpha)$ conditionally on the sample (these heuristic statements are made precise using coupling and decoupling arguments). Observing that if $Y \in J_n^\alpha(Y^*)$, then $\{U \notin \Gamma_\theta(J_n^\alpha(Y^*))\} \subseteq \{U \notin \Gamma_\theta(Y)\}$, we see that the event $U \notin \Gamma_\theta(J_n^\alpha(Y^*))$ has probability zero under the null that $U \in \Gamma_\theta(Y)$.

Formally, what we show is that under the hypothesis of internal consistency of $\mathfrak{M}(\theta, \mathcal{V}_\theta)$, there exists a probability $\pi \in \mathcal{M}(P_n, \mathcal{V}_\theta)$ such that $\pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(J_n^\alpha(y))\}) = 0$. We then apply the duality principle to show that the latter is equivalent to $T_n^\alpha(\theta, \mathcal{V}_\theta) = 0$, which completes the argument since J_n^α is constructed in such a way that this happens with probability at least $1-\alpha$ asymptotically.

Proof:

The assumptions of the theorem are the following:

1. The structure $\mathfrak{M}(\theta, \mathcal{V}_\theta)$ is internally consistent, hence there exists a joint probability $\pi \in \mathcal{M}(P, \mathcal{V}_\theta)$ such that $\pi(\{(u,y) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_u} : u \notin \Gamma_\theta(y)\}) = 0$,
2. The dilation J_n^α satisfies requirement 6, hence with probability tending to $1-\alpha$, there exists a joint probability $\pi^* \in \mathcal{M}(P, P_n)$ such that $\pi(\{(y,v) \in \mathbb{R}^{d_y} \times \mathrm{R}^{d_y} : y \notin J_n^\alpha(v)\}) = 0$.

For the remainder of the proof, all statements are with probability tending to $1-\alpha$. From the assumptions above, there is a coupling π of the pair (Y,U) such that $U \in \Gamma_\theta(Y)$ π-almost surely, and there is a coupling π^* of the pair (Y^*, Y) such that $Y \in J_n^\alpha(Y^*)$ π^*-almost surely.

First step:

We need to show that there exists a coupling $\tilde{\pi}$ of the pair (U, Y^*) such that $U \in \Gamma_\theta(J_n^\alpha(Y^*))$ $\tilde{\pi}$-almost surely. We know that the bootstrap variable Y^* can be written $f(Y, \eta)$, where

f is a measurable mapping and η is a random variable defined on the Polish space $\mathcal{E}$ and independent of Y (in other words, Y^* can be obtained from Y using an independent randomization device). Moreover, by theorem 13.1.1 page 487 of Dudley (2003), for each $y \in \mathcal{Y}$, there exists a bijective mapping $\phi_y : [0,1] \mapsto \mathcal{Y}$ such that ϕ_y and ϕ_y^{-1} are measurable, and $\mathfrak{m} \circ \phi_y = \nu_y$, where $\mathfrak{m}$ is Lebesgue measure on $[0,1]$ (i.e. the uniform distribution) and ν_y is the conditional distribution of U given $Y = y$. Hence, if we take ζ to be a uniformly distributed random variable on $[0,1]$ independent of Y, we have for all set B,

$$\Pr(\phi_Y(\zeta) \in B) = \mathbb{E}\left[\Pr(\phi_y(\zeta) \in B) \mid Y = y\right] = \mathbb{E}\left[\nu_y(B) \mid Y = y\right] = \nu(B),$$

where ν is the distribution of U. Call $\tilde{\pi}$ the distribution of $(U, Y^*) = (\phi_Y(\zeta), f(Y, \eta))$, where (Y, η, ζ) is defined on the product space $\mathcal{Y} \times \mathcal{E} \times [0,1]$ with Y, η and ζ mutually independent. This coupling is such that the marginal distributions are ν and P_n and $U \in \Gamma_\theta(J_n^\alpha)$ $\tilde{\pi}$-almost surely as required.

<u>Second Step:</u>

We now need to show that the existence of the coupling $\tilde{\pi}$ such that $U \in \Gamma_\theta(J_n^\alpha)$ $\tilde{\pi}$-almost surely, implies that $T_n^\alpha(\theta, \mathcal{V}_\theta) = 0$. The existence of $\tilde{\pi}$ is equivalent to

$$\min_{\pi \in \mathcal{M}(P_n, \mathcal{V}_\theta)} \pi\{(u,v) : u \notin \Gamma_\theta(J_n^\alpha(v))\} = 0.$$

The left hand side of the expression above is larger than

$$\sup_{\lambda, h} \sum_{j=1}^n h(Y_j) \quad \text{subject to} \quad \lambda' m(u) + h(v) \leq 1\{u \notin \Gamma_\theta(J_n^\alpha(v))\},$$

where $\lambda \in \mathbb{R}^{d_m}$, and h is a measurable function. Indeed, it suffices to apply $\tilde{\pi}$ to the constraint to obtain the claimed inequality: $\tilde{\pi}\left(\lambda' m(u) + h(v)\right) = \nu(\lambda' m(u)) + \frac{1}{n}\sum_{j=1}^n h(Y_j) \leq \int 1\{u \notin \Gamma_\theta(J_n^\alpha(v))\} d\tilde{\pi}(u,v) = \tilde{\pi}\{(u,v) : u \notin \Gamma_\theta(J_n^\alpha(v))\}$. Now, since the quantity to maximize does not depend on u, the program is equivalent to

$$\sup_\lambda \left(\frac{1}{n} \sum_{j=1}^n h_\lambda(Y_j) \right) \text{ with } h_\lambda(y) = \inf_u \left(1\{u \notin \Gamma_\theta(J_n^\alpha(y))\} - \lambda' m(u)\right).$$

Hence, the latter expression is smaller than zero. Since zero is achieved when $\lambda = 0$, the proof is complete. □

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

Proof of Lemma 2 For $M > 1$, by assumptions 1,

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

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

$$1 - \epsilon \leq \sup_{\nu \in \mathcal{V}_\theta} \nu \left(\{ \| m(U, \theta) \| \leq M \} \right)$$

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

Lemma 3. *If $\mathcal{V}_\theta$ is uniformly tight, then $\mathcal{M}(P, \mathcal{V}_\theta)$ is uniformly tight.*

Proof of Lemma 3 For $\epsilon > 0$, there exists a compact $K_Y \subset \mathcal{Y}$ such that $P(K_Y) \geq 1 - \epsilon/2$; by tightness of $\mathcal{V}_\theta$, there exists also a compact $K_U \subset \mathcal{U}$ such that $\nu(K_U) \geq 1 - \epsilon/2$ for all $\nu \in \mathcal{V}_\theta$. For every $\pi \in \mathcal{M}(P, \mathcal{V}_\theta)$, one has $\pi(K_Y \times K_U) \geq \max(P(K_Y) + \nu(K_U) - 1, 0)$ (Fréchet-Hoeffding lower bound), thus $\pi(K_Y \times K_U) \geq 1 - \epsilon$. □

Proof of Theorem 2 Suppose $\inf_{\pi \in \mathcal{M}(P,\mathcal{V}_\theta)} E_\pi \left[1 \{ U \notin \Gamma_\theta(Y) \} \right] = 0$, we shall show that the infimum is actually attained. Let $\pi_n \in \mathcal{M}(P, \mathcal{V}_\theta)$ a sequence of probability distributions of the joint couple (U, Y) such that

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

By Lemma 3, $\mathcal{M}(P, \mathcal{V}_\theta)$ is uniformly tight, hence by Prohorov's theorem it is relatively compact. Consequently there exists a subsequence $\pi_{\varphi(n)} \in M(P, \mathcal{V}_\theta)$ which is weakly convergent to π.

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

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

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

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

Proof of Theorem 3 We need to prove that there is no duality gap, i.e. that

$$\inf_{\pi \in \mathcal{M}(P_n, \mathcal{V}_\theta)} \pi\{(v,u) : u \notin \Gamma_\theta(J_n^\alpha(v))\} = \sup_{h,\lambda} \frac{1}{n} \sum_{j=1}^{n} h(Y_j)$$

where the sup is under the constraint $h(v) + \lambda' m(u) \leq 1\{u \notin \Gamma_\theta(J_n^\alpha(v))\}$ for all (u, v). This is a consequence of the first line of Proposition 3.4 page 150 of Shapiro (2001)). □

Pilot example 1 continued We have $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ and

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

Hence,

$$\lambda' m(\varepsilon,x) = [1_{\varepsilon \leq 0} - \eta](\lambda_1(1+x) + \lambda_2(1-x)),$$

thus

$$\begin{aligned} \lambda' m(\varepsilon,1) &= 2[1_{\varepsilon \leq 0} - \eta]\lambda_1, \\ \lambda' m(\varepsilon,-1) &= 2[1_{\varepsilon \leq 0} - \eta]\lambda_2. \end{aligned}$$

So, for $\theta = -1$,

$$\begin{aligned} g_{\lambda,\theta}(1,0) &= \min(\inf_{\varepsilon \geq 1}\{-2[1_{\varepsilon \leq 0} - \eta]\lambda_1\}; \inf_{\varepsilon \leq 1}\{1 - 2[1_{\varepsilon \leq 0} - \eta]\lambda_1\}) \\ &= 2\eta\lambda_1 + \min(0, 1 - 2\lambda_1), \end{aligned}$$

and similarly,

$$\begin{aligned} g_{\lambda,\theta}(-1,0) &= 2\eta\lambda_2 + \min(0, -2\lambda_2), \\ g_{\lambda,\theta}(1,1) &= 2\eta\lambda_1 + \min(0, -2\lambda_1), \\ g_{\lambda,\theta}(-1,1) &= 2\eta\lambda_2 + \min(1, -2\lambda_2). \end{aligned}$$

We have

$$\begin{aligned} \mathbb{E}_P[g_{\lambda,\theta}(Z)] &= P_X(-1)[P_{Z|X}(0|-1)g_{\lambda,\theta}(-1,0) + P_{Z|X}(1|-1)g_{\lambda,\theta}(-1,1)] \\ &+ P_X(1)[P_{Z|X}(0|1)g_{\lambda,\theta}(1,0) + P_{Z|X}(1|1)g_{\lambda,\theta}(1,1)]. \end{aligned}$$

Compute

$$\begin{aligned} &[P_{Z|X}(0|-1)g_{\lambda,\theta}(-1,0) + P_{Z|X}(1|-1)g_{\lambda,\theta}(-1,1)] \\ &\quad = 2\eta\lambda_2 + P_{Y|X}(1|-1) \text{ for } \lambda_2 < -1/2, \\ &\quad = 2\lambda_2(\eta - \Pr_{Z|X}(1|-1)) \text{ for } \lambda_2 \in [-1/2, 0], \\ &\quad = 2(\eta - 1)\lambda_2 \text{ for } \lambda_2 > 0. \end{aligned}$$

Similarly,

$$\begin{aligned} &[P_{Z|X}(0|1)g_{\lambda,\theta}(1,0) + P_{Z|X}(1|1)g_{\lambda,\theta}(1,1)] \\ &\quad = 2\eta\lambda_1 \text{ for } \lambda_1 < 0, \\ &\quad = 2\lambda_1(\eta - \Pr_{Z|X}(1|1)) \text{ for } \lambda_1 \in [0, 1/2], \\ &\quad = 2(\eta - 1)\lambda_1 + \Pr_{Z|X}(0|1) \text{ for } \lambda_1 > 1/2. \end{aligned}$$

The expression attains its maximum for $\lambda_2 \in [-1/2, 0]$, the second one for $\lambda_1 \in [0, 1/2]$. Therefore $\sup_{\lambda \in \mathbb{R}^2} \mathbb{E}_P(h_{\lambda,\theta=-1}(Z)) = 0$ if and only if $\eta \geq \Pr_{Z|X}(1|-1)$ and $\eta \leq \Pr_{Z|X}(1|1)$.

Lemma 4. *For all $x_1, x_2, x_1^*, x_2^* \in \mathbb{R}$, if the graph of a bijection between (x_1, x_2) and (x_1^*, x_2^*) crosses, then* $\max(|x_1 - x_1^*|, |x_2 - x_2^*|) \leq \max(|x_2 - x_1^*|, |x_1 - x_2^*|)$.

Proof of Lemma 4 In the trapezium (x_1, x_2, x_1^*, x_2^*), where the parallel sides are (x_1, x_2) and (x_1^*, x_2^*), it is easily seen that when the angle $(x_1^* x_1, x_1 x_2)$ is larger than 90^o, as on the left side of figure 3, then $|x_1^* - x_2| \geq |x_1^* - x_1|$, and when the angle $(x_1^* x_1, x_1 x_2)$ is smaller than 90^o, as on the right side of figure 3, then $|x_1^* - x_2| \geq |x_1 - x_2^*|$. Hence $|x_1 - x_1^*| \leq \max(|x_2 - x_1^*|, |x_1 - x_2^*|)$. Likewise for $|x_2 - x_2^*|$, which proves the result. □

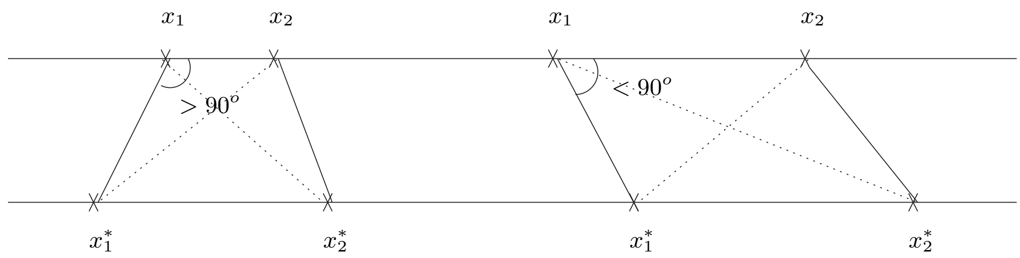

Figure 3: Trapezium configurations.

References

Andrews, D. (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 68, 399–405.

Andrews, D., S. Berry, and P. Jia (2003): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," unpublished manuscript.

Beresteanu, A., and F. Molinari (2006): "Asymptotic properties for a class of partially identified models," Cemmap Working Papers, CWP10/06.

Bierens, H. (1990): "A Consistent Conditional Moment test for Functional Form," *Econometrica*, 58, 1443–1458.

Blundell, R., M. Browning, and I. Crawford (2003): "Nonparametric engel curves and revealed preference," *Econometrica*, 71, 205–240.

Brock, B., and S. Durlauf (2002): "A Multinomial choice model of neighborhood effects," *American Economic Review*, 92, 298–303.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement error models with auxiliary data," *Review of Economic Studies*, 22, 343–366.

Chernozhukov, V., H. Hong, and E. Tamer (2002): "Inference on Parameter Sets in Econometric Models," unpublished manuscript.

Csörgő, M., and P. Révész (1981): *Strong approximations in probability and statistics.* Academic Press.

Dudley, R. (2003): *Real Analysis and Probability.* Cambridge University Press.

Galichon, A., and M. Henry (2006): "Inference in Incomplete Models," Columbia University Discussion Paper 0506-28.

Imbens, G., and C. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1859.

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

Khan, S., and E. Tamer (2006): "Inference on Randomly Censored Regression Models using Conditional Moment Inequalities," unpublished manuscript.

Magnac, T., and E. Maurin (2005): "Partial identification in monotone binary models: discrete regressors and interval data," unpublished manuscript.

Manski, C. (1990): "Nonparametric bounds on treatment effects," *American Economic Review*, 80, 319–323.

Manski, C. (2004): "Social learning from private experiences: the dynamics of the selection problem," *Review of Economic Studies*, 71, 443–458.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004): "Moment Inequalities and Their Application," unpublished manuscript.

Papadimitriou, C., and K. Steiglitz (1982): *Combinatorial Optimization: Algorithms and Complexity.* Prentice Hall.

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

Shaikh, A. (2005): "Inference for a Class of Partially Identified Econometric Models," unpublished manuscript.

Shaikh, A., and E. Vytlacil (2005): "Threshhold crossing models and bounds on treatment effects: a nonparametric analysis," NBER Technical Working Paper 0307.

Shapiro, A. (2001): "On Duality Theory of Conic Linear Problems," in *Semi-Infinite Programming, M. Goberna and M. López, ed.*, pp. 135–165. Kluwer: Dordrecht.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

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