INSTITUTT FOR SAMFUNNSØKONOMI

DEPARTMENT OF ECONOMICS

SAM 1 2013
ISSN: 0804-6824
January 2013

Discussion paper

Identification in Models with Discrete Variables

BY
Lukáš Lafférs

This series consists of papers with limited circulation, intended to stimulate discussion.

Norges Handelshøyskole
NORWEGIAN SCHOOL OF ECONOMICS

Identification in Models with Discrete Variables*

Lukáš Lafférs†

January 8, 2013

Abstract

This paper provides a new simple and computationally tractable method for determining an identified set that can account for a broad set of economic models when economic variables are discrete. Using this method it is shown on a simple example how can imperfect instruments affect the size of the identified set when strict exogeneity is relaxed. It could be of great interest to know to what extent are the results driven by the exogeneity assumption which is often a subject of controversy. Moreover, flexibility gained from the new proposed method suggests that the determination of the identified set need not be application-specific anymore. This paper presents a unifying framework that approaches identification in an algorithmic way.

JEL: C10, C21, C26, C61.

Introduction and Motivation

Identification plays a central role in economic research. In most economic models latent variables such as unobserved heterogeneity, ability or preference shocks are introduced in order to explain relations of interest such that the model mimics reality. Given data that reveals the distribution of observable variables we would like to learn as much as possible about the relations or features of the economic model. This information is often embedded in an unknown parameter. Since latent variables are not directly observable, certain assumptions have to be made about them in order to use data to say something about an unknown parameter or about a feature of interest. Depending on the strength of these assumptions knowledge of the true data generating process of observed variables can have either (1) no identifying power, (2) may shrink the set of potential parameter candidates, in which case the model is said to be partially identified, (3) these assumptions are sufficient to pin down one potentially true parameter which is the point identified case or (4) the assumptions are too strict and the model can be refuted.

*I would like to thank Magne Asphjell, Andrew Chesher, Gernot Doppelhofer, Marc Henry, Toru Kitagawa, Kai Liu, Peter Molnár, Alexei Onatski, Adam Rosen, Erik Sørensen, Konrad Smolinski, Ivan Sutoris and participants at UCL seminar, Cambridge Econometrics Workshop and NHH lunch seminar for helpful comments. All errors are my own.

†Department of Economics, NHH - Norwegian School of Economics, E-mail: `lukas.laffers@nhh.no`, Web: `http://sites.google.com/site/lukaslaffers`

In practice, often strong assumptions that guarantee point identification are made. However such assumptions might include knowledge of the family of probability distributions of unobserved variables which rarely can be justified on economic grounds. The only reason is to make inference tractable. It is interesting to ask what would happen had these restrictions not been imposed and try to develop an inferential procedure that is robust to assumptions that are controversial or made for technical convenience. The first necessary step is to know what is the set of models (or parameters) that are compatible with the set of assumptions made and compatible with the data in case we have perfect information on the probability distribution of observable variables in other words if we have a data sample of infinite length. This is the question of identification. Once this is resolved one can proceed to inference and find out how to use an imperfect data to build confidence regions or hypothesis tests.

The contribution of this paper is threefold. First, a new simple identification method is presented. Second, it is shown how can the method nest several existing results from the literature. Third, we show how can this method approach the identification in cases when strict exogeneity of instruments is relaxed. Advantages compared to the previous literature are that an economic model is not restricted to be linear and at the same time the degree of violation of the exogeneity assumption can be controlled for.

This paper presents a new method that is an extension of an existing framework by Galichon and Henry (2011, 2009) and Ekeland, Galichon, and Henry (2010) (henceforth GH framework) that traces identified set in a richer set of economic problems when observed variables are discrete. As a motivating example we study the impact of violation of the strict exogeneity assumption in single equation endogenous binary response model. Complementing existing results on imperfect instruments (Nevo and Rosen, 2012), (Conley et al., 2012) this method can control the departure from the strict exogeneity of the instrument and allows us to study non-linear models.

The proposed method can reproduce some other results in the partial identification literature that were obtained by different approaches. These include the single equation endogenous binary response model of Chesher (2009) and Chesher (2010), triangular system of equations with binary dependent variables of Shaikh and Vytlacil (2011), treatment effects in studies with imperfect compliance of Balke and Pearl (1997), and binary choice models with zero median restriction of Komarova (2009). In the first and the fourth example the original GH framework[1] applies but the extension help us to formulate the problem such that it is possible to relax the strict exogeneity of instruments in a simple way as is done in section 3. In the other examples, the extension is essential as some of the assumptions that are made can not be formulated within the original GH framework. The extension therefore enriches the set of problems that can be addressed.

The major advantage of the new method is its algorithmic structure: identifying restrictions enter the setup in a straightforward manner and effective algorithms to determine the identified set are employed. Instead of using distinct strategies for different applications, this method provides a unifying framework which is conceptually

[1] After a mild modification.

simple. The presented framework is not application specific it applies to a wide range of problems with discrete variables when identification is only partial.

There are several limitations of the method that is presented in this paper. The method describes how identified set can be found given perfect information on the data generating process of observed variables, yet inference is not studied here. Observable variables in the model are restricted to be discrete. Models with continuous observable variables can be discretized, but this discretization will always bring some degree of arbitrariness to the problem and the impact of this is not studied here. Unobservable variables are *not* restricted to be discrete, a continuous unobservable variable can be transformed into a discrete one and it is shown that this will not affect the identified set.

The study of partial identification was initiated by Manski (1990), however these ideas were not fully appreciated at the beginning. Monographs include Manski (1995) and Manski (2003) and recent comprehensive survey papers are Manski (2008) and Tamer (2010). Among many interesting applications prominent ones are e.g. Returns to schooling (Manski and Pepper, 2000), Demand for fish (Chernozhukov et al., 2009) or Discrete choice with social interactions (Brock and Durlauf, 2001). Determination of identified set is studied in Galichon and Henry (2011, 2009) by means of optimal transportation formulation, in Beresteanu and Molinari (2008), Beresteanu et al. (2012, 2011) and Chesher, Rosen, and Smolinski (2011) using random set theory, and in Chesher (2010) structural quantile functions are used. Reader interested in inference in the partially identified setting might refer to Galichon and Henry (2011, 2009), Chernozhukov, Hong, and Tamer (2007), Imbens and Manski (2004), Beresteanu and Molinari (2008), Beresteanu et al. (2012, 2011), Chernozhukov, Lee, and Rosen (2012), Andrews and Shi (2012), Romano and Shaikh (2010),Bugni (2010) and Rosen (2008).

Section 1 describes the identification strategy of Galichon and Henry with the proposed extension. In section 2 examples are given on how the extended framework can nest different identification approaches. Section 3 explains how one of the examples can be modified in order to study the impact of imperfect instruments. Section 4 concludes and an Appendix consists of proofs (Appendix A), technical details on presented examples (Appendix B), and implementation issues (Appendix C).

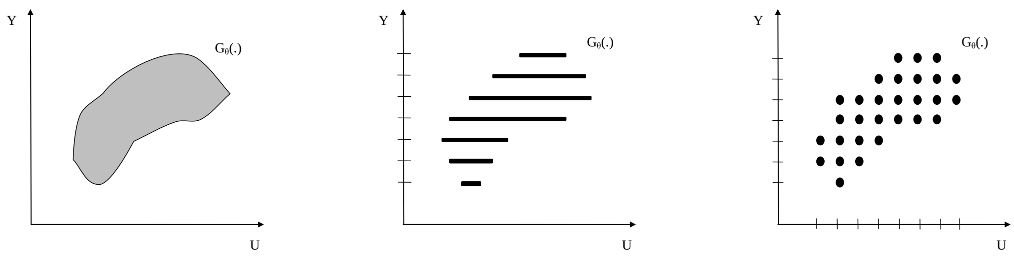
1 Methods

This section first explains the basic elements of the partial identification framework of Galichon and Henry, later on my extension is motivated and presented.

1.1 Galichon and Henry's framework

Here basic ingredients of GH identification setup are presented. Let

- $Y \in \mathcal{Y}$ be a random vector of observable variables with probability density function or probability mass function (pdf or pmf) p,
- $U \in \mathcal{U}$ be a random vector of unobservable variables with pdf or pmf ν,



(a) Y and U continuous (b) Y discrete and U continuous (c) Y and U discrete

Figure 1: Illustration of the correspondence G that carries the information about the economic model. The joint distribution of (Y, U) is restricted to have support in grey area with probability one.

- $G : \mathcal{U} \mapsto \mathcal{Y}$ be a measurable correspondence[2] that restricts the co-occurrence of pairs (Y,U) to those that are compatible with an economic model at hand, formally $Y \in G(U)$. This is how economic restrictions are modeled within the GH setup.

The fact that G is a many-to-many correspondence enables us to work with censored data (for a given Y we contemplate different values of U) or multiple equilibria (for a given U we consider different values of Y). Figure 1 illustrates many-to-many mapping G. Note that point-identification is typically achieved if both Y and U are continuous and an inverse of the many-to-many mapping G^{-1} is a function. In this case a knowledge of probability behavior of observed variables tells us *exactly* the probability of the unobserved component.

First a concept of a **Structure** which groups all available restrictions is defined.

Definition 1. *A Structure S is defined as a triplet $S = (G, \nu, p)$.*

Another important notion to be defined is an **internal consistency** of a structure. The structure is internally consistent if there exists a joint distribution which potentially could have generated the probability of observed variables p and latent variables ν and satisfies economic restriction defined by G almost surely. If there is no such joint distribution the structure can clearly be refuted.

Definition 2. *Structure S is said to be internally consistent if and only if there exists a joint probability distribution π of (Y, U) on $\mathcal{Y} \times \mathcal{U}$ with marginal distributions p and ν respectively such that $Pr_\pi(\{Y \in G(U)\}) = 1$.*[3]

In practice, most models are parametrized so let us now consider the situation when $\nu = \nu_\theta$ and $G = G_\theta$ are parametrized with a vector of parameters $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^d$.[4] Finally, our object of interest, an **Identified set**, is defined. It is a collection of all parameters θ that guarantee internal consistency of the structure.

[2] Therefore for all open subsets A of $\mathcal{Y}$, $G^{-1}(A) := \{U \in \mathcal{U} : G(U) \cap A \neq \varnothing\}$ is well defined.
[3] Definition 1 in Galichon and Henry (2009)
[4] The parameter θ may consist of two parts, $\theta = [\theta_1, \theta_2]$, so we can have G_{θ_1} and ν_{θ_2}.

Definition 3. *An identified set for θ, $\Theta_I(p)$, is defined as*
$\Theta_I(p) := \{\theta \in \Theta : (G_\theta, \nu_\theta, p)$ is internally consistent$\}$.[5]

Note that all members of the identified set correspond to structures that could have generated the probability of observed variables p, so in this sense they are *observationally equivalent*, and no amount of data would ever help us to distinguish between them. The identified set

- could be empty: $\Theta_I(p) = \{\varnothing\}$, hence the structure $(G_\theta, \nu_\theta, p)$ is *refuted* for all $\theta \in \Theta$,
- may consist of a single point: $\Theta_I(p) = \{\theta\}$, in this case θ is *point-identified*,
- can be a subset of Θ: $\Theta_I(p) = \{I \subset \Theta\}$ and θ is *partially identified*,
- may not shrink Θ at all: $\Theta_I(p) = \Theta$, so the structure $(G_\theta, \nu_\theta, p)$ places *no identifying* restrictions on θ.

For a fixed parameter θ, if all variables in the model are discrete the problem of finding a joint distribution of (Y, U) compatible with the economic model described by G_θ with appropriate marginals can be formulated as a linear program, and it will be shown how. Note that in most economic applications the latent component U is continuous. If the observed variables are discrete it is however possible to discretize U in a way that leaves the identified set unchanged as was proved in Galichon and Henry (2011). Suppose $\mathcal{Y} = \{y_1, ..., y_i, ..., y_n\}$ with corresponding probabilities p_i, $\mathcal{U} = \{u_1, ..., u_j, ..., u_m\}$ with probabilities ν_j. Economic model enters the problem as a set of restrictions on the support of (Y, U). Let us define a zero-one penalty on the support of all joint probabilities on $\mathcal{Y} \times \mathcal{U}$,

$$c_{ij} = 1(y_i \notin G_\theta(u_j)) = \begin{cases} 0, & \text{if } y_i \in G_\theta(u_j), \\ 1, & \text{otherwise,} \end{cases}$$

so penalty is put on those pairs (Y, U) that are incompatible with the economic model. The $n \times m$ matrix of the zero-one penalties $\{c_{ij}\}$ carries the same information as the mapping $G_\theta(.)$ and we denote the nm vector of this stacked matrix as $\boldsymbol{c}$.

Now a question of an existence of a joint probability distribution which assures internal consistency can be answered by means of the following linear program:[6]

$$\begin{aligned} &\min_{(\pi)} \textstyle\sum_{i,j} \pi_{ij} c_{ij} \\ &\text{s.t.} \\ &\textstyle\sum_j \pi_{ij} = p_i, \quad \forall i \qquad (1) \\ &\textstyle\sum_i \pi_{ij} = \nu_j, \quad \forall j \qquad (2) \\ &\pi_{ij} \geq 0, \quad \forall i, j. \qquad (3) \end{aligned}$$

where the minimum is taken across the all joint probability distributions π (nm vector of the stacked $n \times m$ matrix with elements $\{\pi_{ij}\}$). A structure is internally consistent if and only if the optimized value of the objective function is equal to 0. If this is

[5] Definition 2 in Galichon and Henry (2009), the dependence of the identified set $\Theta_I(p)$ on the distribution of observable variables p is made explicit.

[6] The dependence of c_{ij} and ν_j on parameter θ is omitted for the sake of brevity.

the case, it means that we have found a proper joint distribution π that is compatible with with the data (1) and the assumptions made on latent variables (2), and that the probability of an event not compatible with the economic model is zero.

The necessary and sufficient condition for the inclusion of the parameter θ in the identified set is

$$0 = \max_{A \subset \mathcal{Y}} (Pr(A) - \nu_\theta(G_\theta^{-1}(A))), \tag{4}$$

where the maximum is taken across all possible subsets of $\mathcal{Y}$. A similar result was first proven by Artstein (1983) and is based on an extension of the marriage lemma. Alternative proofs of (4) were given in Galichon and Henry (2009) which relied on optimal transportation theory, and in Henry et al. (2011) based on combinatorial optimization methods. Equation (4) can then also be used for hypothesis testing or building confidence regions for θ as proposed in Galichon and Henry (2009) and Henry et al. (2011). The latter allows for efficient confidence regions construction using a combinatorial bootstrap.

The properties of the approach:

- Flexible way how to access many problems when partial identification occurs.
- For discrete cases linear program nature makes is computationally convenient.
- If only U is continuous, problem can be transformed into the discrete one.
- Economic model is described by restriction on the support of observables and unobservables.

1.2 The extension of the Galichon and Henry framework

I aim to extend the GH method to entertain additional distribution restrictions. Even though the GH setup can address many problems, certain type of problems cannot be formulated within the GH framework. There are two ways how can our prior information enter the structure: the marginal distribution of unobservables ν and the support of (Y, U) via the correspondence G (or equivalently $\boldsymbol{c}$). Not all distributional assumptions we might believe can enter the structure. In many economic models some notion of independency is assumed.[7]

Because the problem is accessed at the lowest level, by constructing a joint distribution compatible with all the information researcher may have, it is possible to restrict this joint distribution to satisfy *any* type of distributional assumptions one may wish to make. If the distributional assumption can be written as a linear function of the joint probability π, the problem remains computationally attractive. Modeling the joint distribution gives us full control on utilizing the information at hand. This flexibility delivers a solution to cases where the GH setup is too restrictive, and this is the main contribution of this paper.

For illustrative purposes: suppose that in addition to information about G, it is known that $E(\phi_\theta(Y, U)) = 0$ and $|cov(Y, U)| \leq 0.1$. Such assumptions simply cannot

[7]One may also be willing to make some assumption about a distribution of variable in form of moment equality or moment inequality. It is important to note here that GH setup can handle moment inequalities $E(\phi(Y)) \leq 0$ if $E(m(U)) = 0$ is assumed (Ekeland et al. (2010) and Mourifié and Henry (2012)). In this case correspondence G is restricted to take specific form. However within the GH framework it is not possible to consider moment inequality *and* further information given by G.

be formulated as the restriction on the support of (Y, U), so there is no way how these assumptions can be embedded into the framework via G or ν. In this sense the original GH framework is too restrictive. The way to incorporate these assumptions is simply to restrict the set of joint distributions (all π-s) to only those that are compatible with this piece of information.

A question whether the extended set of restrictions is compatible with the observed data reduces to checking whether the optimized value is equal to zero in the following linear program:

$$
\begin{aligned}
&\min_{(\pi)} \textstyle\sum_{i,j} \pi_{ij} c_{ij} \\
&\text{s.t.} \\
&\textstyle\sum_j \pi_{ij} = p_i, \qquad \forall i \\
&\textstyle\sum_i \pi_{ij} = \nu_j, \qquad \forall j \\
&\textstyle\sum_{i,j} \pi_{ij} \phi_\theta(y_i, u_j) = 0, \qquad (5) \\
&\textstyle\sum_{i,j} \pi_{ij} y_i u_j - \sum_i p_i y_i \sum_j \nu_j u_j \leq 0.1, \qquad (6) \\
&-\textstyle\sum_{i,j} \pi_{ij} y_i u_j + \sum_i p_i y_i \sum_j \nu_j u_j \leq 0.1, \qquad (7) \\
&\pi_{ij} \geq 0, \qquad \forall i, j.
\end{aligned}
$$

Equation (5) restricts the joint distribution π to satisfy $E(\phi_\theta(Y, U)) = 0$, whereas inequalities (6) and (7) ensure that $|cov(Y, U)| \leq 0.1$ is satisfied.

As another example suppose that we have two observed variables $Y = (X, Z)$ with probabilities p_{ij} and unobserved variable U and instead of assuming the full knowledge of its distribution, we assume that it has zero mean, its 75% quantile is 0.8 and it is independent of Z. Now the problem would be formulated as follows

$$
\begin{aligned}
&\min_{(\pi)} \textstyle\sum_{i,j,k} \pi_{ijk} c_{ijk} \\
&\text{s.t.} \\
&\textstyle\sum_k \pi_{ijk} = p_{ij}, \qquad \forall i, j \\
&\textstyle\sum_{i,j,k} \pi_{ijk} u_k = 0, \\
&\textstyle\sum_{i,j,k} \pi_{ijk} 1(u_k \leq 0.8) = 0.75, \\
&\textstyle\sum_i \pi_{ijk} - \sum_i p_{ij} \sum_{i,j} \pi_{ijk} = 0, \qquad \forall j, k \\
&\pi_{ijk} \geq 0, \qquad \forall i, j, k.
\end{aligned}
$$

These examples are somewhat artificial but explain the main point. Economically interesting examples follow in section 2. It is important to note that if the additional constraints are such that the problem is within the linear programming framework, it remains computationally feasible.

The crucial step is to prove that discretization of unobserved variables is possible *even* when additional distributional restrictions are entertained. This is done for a certain class of distributional restrictions, and is discussed in detail in subsection 1.3 with a proof given in Appendix A.

We now state the proposed extension formally. We recall that $\mathcal{Y}$ and $\mathcal{U}$ are supports of discrete observable variable and continuous or discrete unobservable variables respectively. The set of all probabability distributions on $\mathcal{Y} \times \mathcal{U}$ is denoted by $\mathbf{\Pi}(\mathcal{Y}, \mathcal{U})$

and $\psi_\theta(\mathcal{Y},\mathcal{U},p,\nu)$ is the set of all $\pi \in \mathbf{\Pi}(\mathcal{Y},\mathcal{U})$ satisfying additional restrictions imposed. If information about the probability distribution ν of unobserved variables is not available, we have $\psi_\theta(\mathcal{Y},\mathcal{U},p)$, A set of all restrictions imposed is compatible with the data if and only if the optimal solution of the following optimization procedure is zero:

$$\begin{aligned} &\min_{(\pi)} \pi\{1(Y \notin G_\theta(U))\} \\ &\text{s.t.} \\ &\pi\{1(Y = y_i)\} = p_i, \qquad \forall i \\ &\pi \in \psi_\theta(\mathcal{Y},\mathcal{U},p,\nu). \end{aligned}$$

Note that if U is discrete and the set ψ_θ consists of restrictions that are linear in π, linear programming routines may be used.

The additional restrictions for the two examples given above are:

$$\psi_\theta(\mathcal{Y},\mathcal{U},p,\nu) = \left\{ \pi \in \mathbf{\Pi}(\mathcal{Y},\mathcal{U}) : \begin{array}{l} \forall u \in \mathcal{U} : \pi\{1(U=u)\} = \nu(u), \\ E_\pi \phi_\theta(Y,U) = 0, \\ \left| E_\pi YU - \sum_i p_i y_i \sum_j \nu_j u_j \right| \leq 0.1. \end{array} \right\} \tag{8}$$

and

$$\psi(\mathcal{X}\times\mathcal{Z},\mathcal{U},p) = \left\{ \pi \in \mathbf{\Pi}(\mathcal{X}\times\mathcal{Z},\mathcal{U}) : \begin{array}{l} E_\pi U = 0, \\ E_\pi 1(U \leq 0.8) = 0.75, \\ \forall z \in \mathcal{Z}, u \in U : \pi\{1(Z=z, U=u)\} = \\ \qquad = \sum_i p_{ij} \pi\{1(U=u)\ \}. \end{array} \right\} \tag{9}$$

where in the second example $\mathcal{Y} = \mathcal{X}\times\mathcal{Z}$ and ψ does not depend on θ.

The notion of structure and identified set now have to be redefined. To enrich the concept of the original structure we denote a triplet (G,ψ,p) as a **Generalized Structure** which groups all the restrictions placed on π.

Definition 4. *A Generalized Structure S is defined as a triplet $S = (G,\psi,p)$.*

Internal consistency and **identified set** are then defined similarly as in definitions 2 and 3.

Definition 5. *Generalized Structure S is said to be internally consistent if and only if there exists a joint probability distribution π of (Y,U) on $\mathcal{Y}\times\mathcal{U}$ in $\psi(\mathcal{Y},\mathcal{U},p)$ with Y-marginal distributions p such that $\pi(\{Y \in G(U)\}) = 1$.*

Definition 6. *An identified set for θ, $\Theta_I(p)$, is defined as*
$\Theta_I(p) := \{\theta \in \Theta : (G_\theta,\psi_\theta,p) \text{ is internally consistent}\}$.

We will refer to this formulation as the *extended Galichon and Henry framework*. If the latent variable U is discrete and the set ψ can be written as linear restrictions in π, effective algorithms can be employed to solve this linear program.

1.3 Discretization of Unobserved Variables

In most economic problems the unobserved component is continuous, hence in order to make the search in the space of joint probability functions tractable it may be convenient to discretize the unobserved component and then show that this discretization leaves the identified set unaffected. This is not true in general. We will show that if distributional restrictions ψ take specific forms, that nest all examples presented in this paper, discretization of the unobserved variable is possible and harmless. These sets of restrictions for the problems with continuous unobservables are

$$\psi_1(\mathcal{Y},\mathcal{U},p,\nu)=\left\{\pi\in\mathbf{\Pi}(\mathcal{Y},\mathcal{U}):\begin{array}{c}\forall u\in\mathcal{U}:\pi\{1(U=u)\}=\nu(u),\\ \forall I\in\mathbb{I};\forall u\in\mathcal{U}:\\ |\sum_{i\in I}\pi(y_i,u)-\sum_{i\in I}p_i\nu(u)|\leq\alpha\sum_{i\in I}p_i\nu(u),\end{array}\right\} \tag{R1}$$

and

$$\psi_2(\mathcal{Y},\mathcal{U},p,\cdot)=\left\{\pi\in\mathbf{\Pi}(\mathcal{Y},\mathcal{U}):\begin{array}{c}E_\pi\phi(U)=0,\\ \forall I\in\mathbb{I};\forall u\in\mathcal{U}:|\sum_{i\in I}\pi(y_i,u)-\\ -\sum_{i\in I}p_i\pi\{1(U=u)\}|\leq\alpha\sum_{i\in I}p_i\pi\{1(U=u)\},\end{array}\right\} \tag{R2}$$

where $\phi:\mathcal{U}\mapsto\mathcal{M}$ has a finite range $\mathcal{M}$ and $\mathbb{I}$ is a fixed set of indices.[8]

The the first restriction (R1) requires π to be compatible with assumed distribution of unobserved variables and hence nests original GH framework and the second will help us to restrict part of the observed component to be independent or "close to being independent"[9] of the unobserved component and the first line in (R2) will allow us to work with quantiles of U.

Let us denote a question of internal consistency of a generalized structure (G,ψ,p) with continuous unobserved variable as $\mathcal{P}_1$:

Y discrete with support $\mathcal{Y}=\{y_1,...,y_n\}$ and with probability $p=\{p_1,...,p_n\}$,
U continuous with support $\mathcal{U}$ (and with positive probability density ν for (R1)),
$G:\mathcal{U}\mapsto\mathcal{Y}$.

The aim is to find a function $\pi_1:\mathcal{Y}\times\mathcal{U}\mapsto[0,1]$ that satisfies

$$\sum_{i=1}^{n}\int_{u\in\mathcal{U}}\pi_1(y_i,u)1(y_i\in G(u))du=1, \tag{1.1}$$

$$\forall i=1,...,n:\quad \int_{u\in\mathcal{U}}\pi_1(y_i,u)du=p_i, \tag{1.2}$$

$$\pi_1\in\psi(\mathcal{Y},\mathcal{U},p,\cdot). \tag{1.3}$$

[8]If observed variable is multidimensional we can stack it into one vector. Summing across some sets of indices allows us formulate a restriction for one dimension only. As an example suppose that observed variables are (Y,X,Z), then we can place a restriction on X only, so that X is independent of U.

[9]The way how the independency restriction is relaxed will be discussed in section 3.

Problem $\mathcal{P}_1$ is computationally unfeasible because of its continuous component U. We can however transform the problem $\mathcal{P}_1$ with continuous U to the problem $\mathcal{P}_2$ with discrete U, such that it will not affect the identified set.

We will partition $\mathcal{U}$ into subsets that deliver the same $G(U)$ for the set of restrictions (R1) and into those that deliver the same $G(U)$ and $\phi(U)$ for (R2). It is then easy to show that if we group all Us in these subsets into atoms and proceed as if U were discrete, the identified set stays unchanged.

Formally, the partitioning of the $\mathcal{U}$ space is the following

$$\mathcal{G} \equiv \{\Delta^* \subset \mathcal{U} : \forall g_I \in \Delta^*, \forall g_{NI} \in \Delta^{*C} : G(g_I) \neq G(g_{NI})\} \tag{PartU1}$$

for (R1) and

$$\mathcal{S} \equiv \{\Delta^* \subset \mathcal{U} : \forall s_I \in \Delta^*, \forall s_{NI} \in \Delta^{*C} : G(s_I) \neq G(s_{NI}),\ \phi(s_I) \neq \phi(s_{NI})\}. \tag{PartU2}$$

for (R2).

The assumption of a finite range of ϕ is crucial as it implies a finite $\mathcal{S}$. Let m denotes the cardinality of either $\mathcal{G}$ or $\mathcal{S}$ depending on which one is in use. Then a new random variable U^* is defined. For every $j \in \{1, ..., m\}$, we choose a point of support u_j^* to be any $u \in \Delta_j^*$, a representative of the set Δ_j^*

$$\mathcal{U}^* \in \Delta_1^* \times \cdots \times \Delta_m^*. \tag{U}$$

To obtain a probability distribution ν^* of U^*, needed for restrictions (R1), we integrate $\nu(u)$ across the corresponding regions Δ_j^* of $\mathcal{U}$:

$$\forall j = 1, ..., m : \qquad \nu_j^* \equiv \int_{\Delta_j^*} \nu(u) du. \tag{P}$$

The discretized problem $\mathcal{P}_2$ is the following:

Y with support $\mathcal{Y} = \{y_1, ..., y_n\}$ with probability $p = \{p_1, ..., p_n\}$

U^* with support $\mathcal{U}^* = \{u_1^*, ..., u_m^*\}$ (with probability $\nu^* = \{\nu_1^*, ...\nu_m^*\}$ for (R1))

$G : \mathcal{U}^* \mapsto \mathcal{Y}$

The question is whether there exists a function $\pi_2 : \mathcal{Y} \times \mathcal{U}^* \mapsto [0, 1]$ such that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_2(y_i, u_j^*) 1(y_i \in G(u_j^*)) du = 1, \tag{2.1}$$

$$\forall i = 1, ..., n : \qquad \sum_{j=1}^{m} \pi_2(y_i, u_j^*) = p_i, \tag{2.2}$$

$$\pi_2 \in \psi(\mathcal{Y}, \mathcal{U}^*, p, \cdot), \tag{2.3}$$

Lemma 1. *If (R1),(P) and (PartU1) hold then a generalized structure $(G, \psi(\mathcal{Y}, \mathcal{U}, p, \nu), p)$ is internally consistent if and only if a generalized structure $(G, \psi(\mathcal{Y}, \mathcal{U}^*, p, \nu^*), p)$ is internally consistent.*

Lemma 2. *If* (R2) *and* (PartU2) *hold then a generalized structure* $(G, \psi(\mathcal{Y}, \mathcal{U}, p, \cdot), p)$ *is internally consistent if and only if a generalized structure* $(G, \psi(\mathcal{Y}, \mathcal{U}^*, p, \cdot), p)$ *is internally consistent.*

Proofs are given in Appendix A. Lemmata 1 and 2 say that for the internal consistency of a generalized structure, proposed discretization is innocuous.

It immediately follows that if G and ψ were parametrized by some $\theta \in \Theta$, problem $\mathcal{P}_1$ and problem $\mathcal{P}_2$ would lead to the *same* identified set for θ for both (R1) and (R2).

2 Motivating Examples

The present section introduces some examples for how the extended GH framework applies to problems in partial identification literature. The replication of existing results illustrates that the proposed extension indeed works. There is no computational gain from employing the presented method compared to other frameworks that derive analytical solutions. The greatest advantage of this method is its *generality*. Instead of deriving the identified set and proving that it is sharp from case to case, we propose one unifying framework that will trace the identified set regardless of the application. It is sufficient to formulate the economic model with restrictions in the extended GH setup and let the computer do the work. Also, if extra information becomes available, it is straightforward to incorporate it into the setup. Unlike the existing application-specific approaches, where incorporating further restrictions or changing the existing restriction, may cause significant difficulties for tracing the identified set, adding extra assumptions or changing the existing ones in the extended GH framework is trivial. Moreover, if the distributional restrictions are linear in the joint probability π, linear programming routines can be used. This is particularly interesting as linear programming is well understood, and ready-to-use computer codes are widely available.

The four examples presented in this section not only demonstrate that the method nests a few existing identification strategies and can replicate their results but they also illustrate how to formulate the economic problem at hand into the extended GH framework.

The four considered examples include single equation endogenous binary response model of Chesher (2009, 2010), bounds on treatment effects in triangular models with binary dependent variables (Shaikh and Vytlacil, 2011) and in studies with imperfect compliance of Balke and Pearl (1997) and binary choice models with zero median restriction of Komarova (2009).

For each example, first the problem and the notation is introduced, second discretization of unobserved variables is presented, third the problem is formulated within the extended GH framework, and fourth the results are compared. The original identification strategy is briefly outlined in Appendix B together with technical details on the examples.

Example 1: Single Equation Endogenous Binary Response Model

Illustrative example of a single equation endogenous binary response model is taken from Chesher (2010). Consider a probit model where discrete explanatory variable X

is possibly correlated with an unobserved U and an instrument Z which is independent of U is available.[10] Such model is in general not point-identified.

Suppose that the set of assumptions that define our model is the following:

- $Y = 1(U > t(X))$ (10)
- $U \perp\!\!\!\perp Z$ - the unobserved U is independent of the instrument Z
- $U \sim Unif(0,1)$ - U is uniformly distributed on $[0,1]$ interval
- $t(X) = \Phi(-\theta_0 - \theta_1 X)$ - the threshold-crossing function is assumed to take a particular form, where $\Phi(.)$ is a cumulative distribution function of the standard normal distribution.[11]

An interesting question one may want to ask is: Given that we have perfect information on the distribution of observables, the question is what can we say about the function $t(X)$ or equivalently about the coefficient $\theta = (\theta_0, \theta_1)$ from our economic model and from our assumptions.

Discretization of Unobservables

The discretization as explained in section 1.3 in this case boils down to the discretization employed by Galichon and Henry (2011) in the original GH setup. This is because the additional assumption $E(\phi(U)) = 0$ is not present. It is demonstrated for illustrative purposes.

Suppose that $\theta_1 > 0$ then the only subsets of (Y, X) that are compatible with (10) are $\{(0,0),(0,1)\}$ for $U \leq t(1)$, $\{(0,0),(1,1)\}$ for $t(1) < U \leq t(0)$ and $\{(1,0),(1,1)\}$ for $U < t(0)$. We assign to these three sets of Us three points (u_1^*, u_2^*, u_3^*) with probabilities $(t(1), t(0) - t(1), 1 - t(0))$. A similar procedure applies for $\theta_1 < 0$. On figure 2 we can see the case for $\theta_1 > 0$ on the left and for $\theta_1 < 0$ on the right side. Upper panes show the original support restriction G_θ and lower panes their discrete counterparts.

Formulation in the Extended GH framework

The distribution of observables (Y, X, Z) is assumed to be known and is denoted as p_{ijk} and U is assumed to be uniformly distributed on $[0,1]$.[12] For a given (θ_0, θ_1), the aim is to find joint probability π_{ijkl} of (Y, X, Z, U) that is compatible with the support restrictions and the distributional restrictions - marginals of π_{ijkl} are p_{ijk} and ν_l respectively and Z and U are independent.

The support restrictions are defined as follows

$$c_{ijkl} = 1(y_i \neq 1(u_l > t(x_j))) = \begin{cases} 0, & y_i = 1(u_l > t(x_j)), \\ 1, & \text{otherwise.} \end{cases} \tag{11}$$

[10] In case when X is continuous, the parameter is point identified and could be obtained by e.g. STATA's `ivprobit`.

[11] It is possible to determine lower and upper bound of the threshold-crossing function $t(X)$ without making this parametric assumption as it was done in Chesher (2009) and assume monotonicity of $t(X)$ instead. For the sake of simplicity the parametric example is presented.

[12] We could also assume that we observe the probability of Y, X given Z, for the sake of exposition probability of (Y, X, Z) is known.

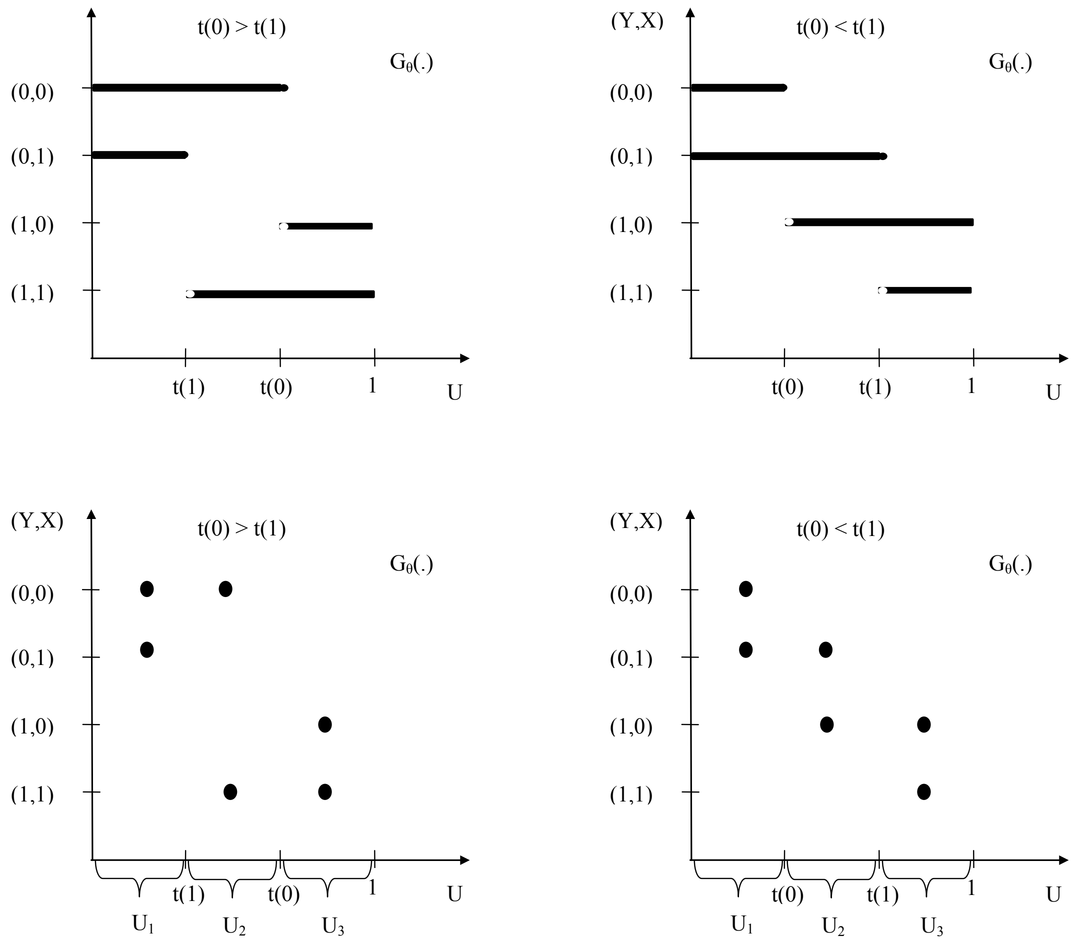


Figure 2: Discretization of unobservables in example 2. Left pane is for $\theta_1 > 0$, right one for $\theta_1 < 0$. Under the original continuous formulation of G_θ is its discretized counterpart.

so basically (Y, X, Z, U)s are restricted to those that satisfy (10).

We now turn into the formulation of the problem in the extended GH framework:

$$\begin{array}{rl}
\min_{(\pi)} \sum_{i,j,k,l} \pi_{ijkl} c_{ijkl} & \\
\text{s.t.} & \\
\sum_l \pi_{ijkl} = p_{ijk}, & \forall i,j,k \\
\sum_{i,j,k} \pi_{ijkl} = \nu_l, & \forall l \\
\sum_{i,j} \pi_{ijkl} = \sum_{i,j} p_{ijk}\nu_l, & \forall k,l \\
\pi_{ijkl} \geq 0, & \forall i,j,k,l.
\end{array} \tag{12}$$

If for a given (θ_0, θ_1) the optimum is achieved at 0, this (θ_0, θ_1) is added into the identified set.[13] [14]

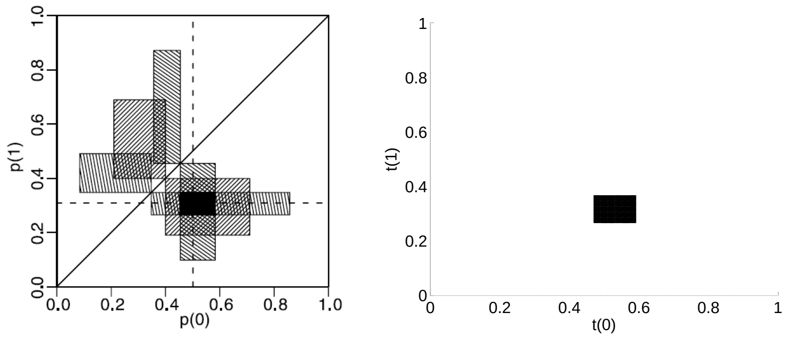


Figure 3: Identified set obtained by Chesher's approach (Chesher, 2010) is compared with our solution.

Results (binary X)

Identified set is expressed in terms of threshold-crossing function at 0 and 1, $t(0)$ and $t(1)$, rather than in the parameter space.[15] Figure 3 documents that extended GH setup does work for instruments in the case with binary endogenous variable.

Results (continuous X discretized)

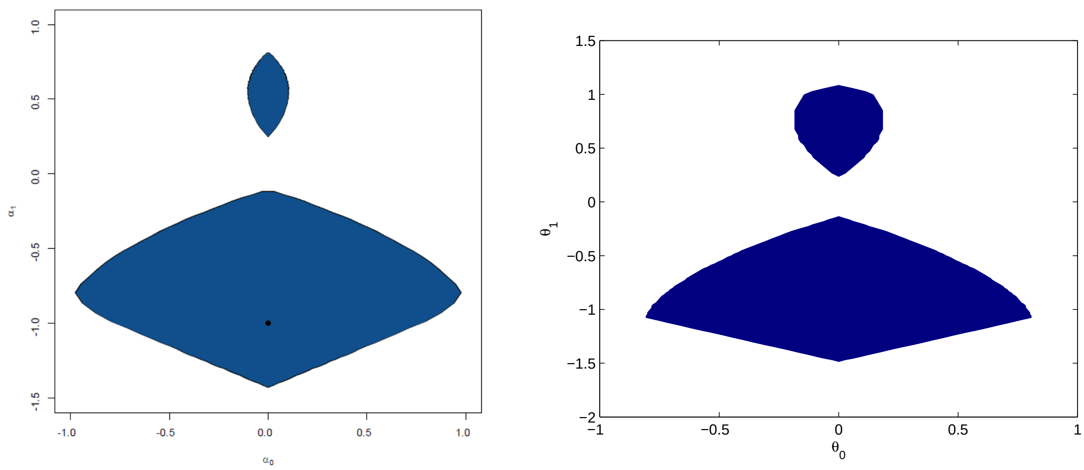


Figure 4: Chesher's result Chesher (2009) (Figure 8, p.37) for problem (10) with parameters given by (23) compared with the result obtained by extended GH approach.

Figure 4 compares the results obtained by Chesher (2009) and extended GH framework. Note that even though the shapes of the identified sets are similar, they are different. Methods of discrete approximations of continuous observed variables have to be developed in order to get reliable results.

---

[13]In this case parameter θ affects the support restrictions (10) only.

[14]Note that even though π is four dimensional the problem still lies within the linear programming framework since elements of π can be stacked to make a vector of size $n_Y \cdot n_X \cdot n_Z \cdot n_U$.

[15]In order to avoid confusion with probabilities p_{ijk} of observed variables, the threshold-crossing function is denoted as $t(.)$ unlike in Chesher (2009) who set it as $p(.)$.

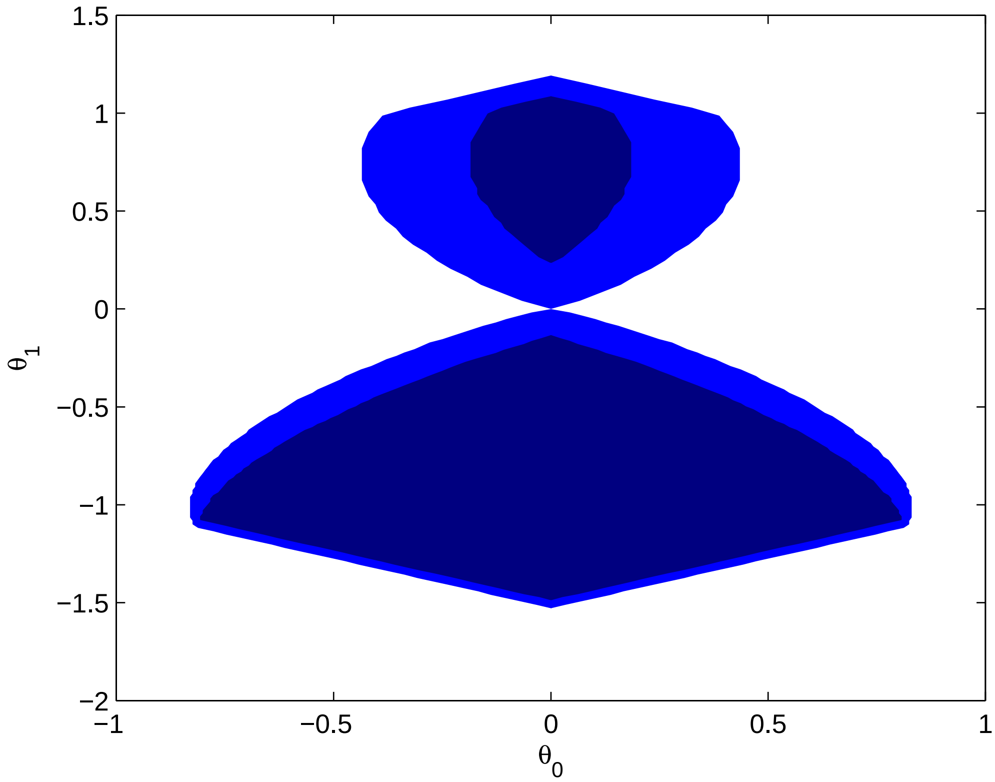


Figure 5: Dark blue - with independency restriction, light blue - without assuming independency.

Identifying power of the independency restriction

The identifying strength of the independency condition itself can now be studied.

Figure 5 shows the strength of the independency restriction.[16] It is clear that this extra information shrinks the identified region. It is also worth to notice that even if the instruments are entirely endogenous, some parameter values are excluded from the identified set. For these, no joint probability π_{ijkl} of observables and unobservables that is compatible with the data generating process p_{ijk} and with ν_l exists.

Objective function

On figures 6 and 7 the minimized objective function and its contours are shown.

Zeros of this function correspond to the identified set, however the values outside the identified set have interesting interpretation too": they stand for the minimal probability of a event incompatible with the economic model. If for instance, for a certain parameter value the minimized value of the objective function is 0.2 then it means that for any data-generating process at least 20% of the pairs of observed

[16]That means with the second last restriction omitted: $\sum_{i,j} \pi_{ijkl} = \sum_{i,j} p_{ijk}\nu_l \;\; \forall k, l$.

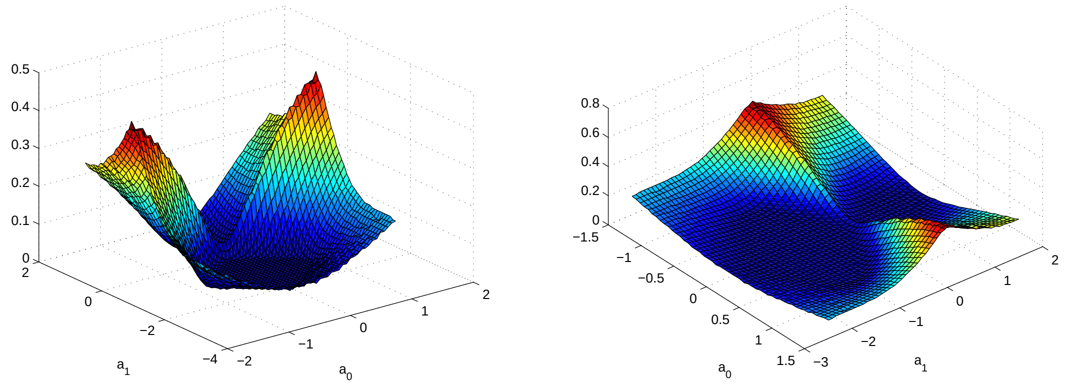


Figure 6: Minimized objective function.

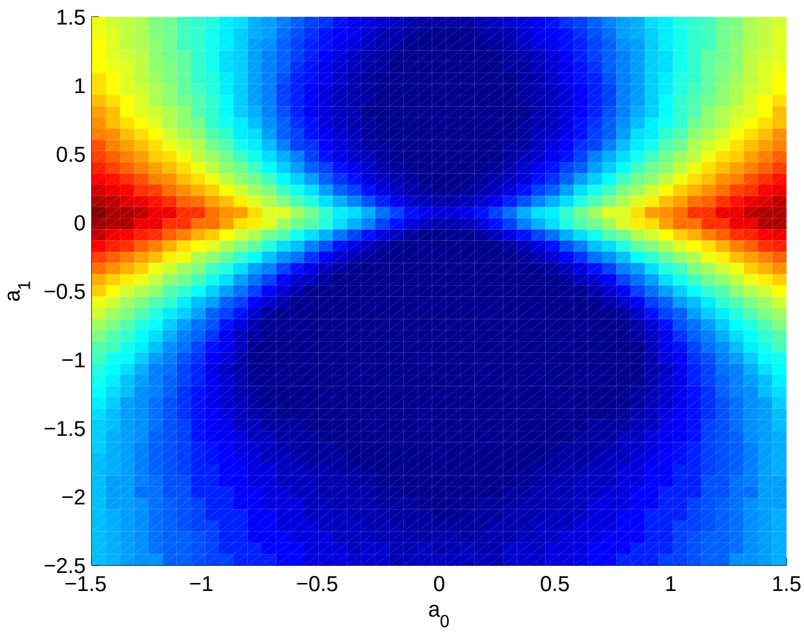


Figure 7: Contours of the minimized objective function.

and unobserved variables violate the support restrictions.[17] This may serve as an appealing measure of misspecification with respect to the support restrictions.

Example 2: Triangular System of Equations with Binary Dependent Variables

Following Shaikh and Vytlacil (2011) the object of interest is an Average Treatment Effect (ATE) in triangular system of equations.

The collection of assumptions is as follows:

- $Y = 1(\alpha D + \beta X - \epsilon_1 \geq 0),$ (13)
- $D = 1(\delta Z - \epsilon_2 \geq 0),$ (14)
- $(X, Z) \perp\!\!\!\perp (\epsilon_1, \epsilon_2),$

where Y is a binary outcome variable, D is a treatment identificator, X is an exogenous covariate and Z is an instrument. Note that no parametric distributional assumptions on (ϵ_1, ϵ_2) are made.

Formulation in the Extended GH framework

We have four observed variables (Y, X, D, Z) with probabilities p_{ijkl} and two unobserved variables (ϵ_1, ϵ_2). Let us denote $\pi_{ijklmn} = Pr(Y = y_i, X = x_j, D = d_k, Z = z_l, \epsilon_1 = u^1_m, \epsilon_2 = u^2_m)$. The penalty on the points of support not compatible with the economic restrictions G is given by

$$c_{ijklmn} = \begin{cases} 0, & (y_i, x_j, d_k, z_l, u^1_m, u^2_n) : y_i = 1(\alpha d_k - u^1_m \geq 0) \text{ and } d_k = 1(\delta z_l - u^2_n \geq 0), \\ 1, & \text{otherwise.} \end{cases}$$

A particular value of $ATE = \theta$ is compatible with the list of assumptions and with data (p_{ijkl}) if and only if zero is the optimal solution of the following optimization problem:

$$\begin{array}{cl} \min_{(\pi)} \sum_{i,j,k,l,m,n} \pi_{ijklmn} c_{ijklmn} & \\ \text{s.t.} & \\ \sum_{m,n} \pi_{ijklmn} = p_{ijkl}, & \forall i,j,k,l \\ \sum_{i,k} \pi_{ijklmn} = \sum_{i,k} p_{ijkl} \sum_{i,j,k,l} \pi_{ijklmn}, & \forall k,l,m,n \\ \sum_m \left[1(\alpha \geq u^1_m) - 1(0 \geq u^1_m)\right] \sum_{i,j,k,l,n} \pi_{ijklmn} = \theta, & \\ \pi_{ijklmn} \geq 0, & \forall i,j,k,l,m,n. \end{array}$$

Results

Figures 8 and 9 compare the results of Shaikh and Vytlacil (2011) with the extended GH framework.

[17] From Lemma 2 we can observe that this interpretation is not affected by the discretization of the unobserved variables.

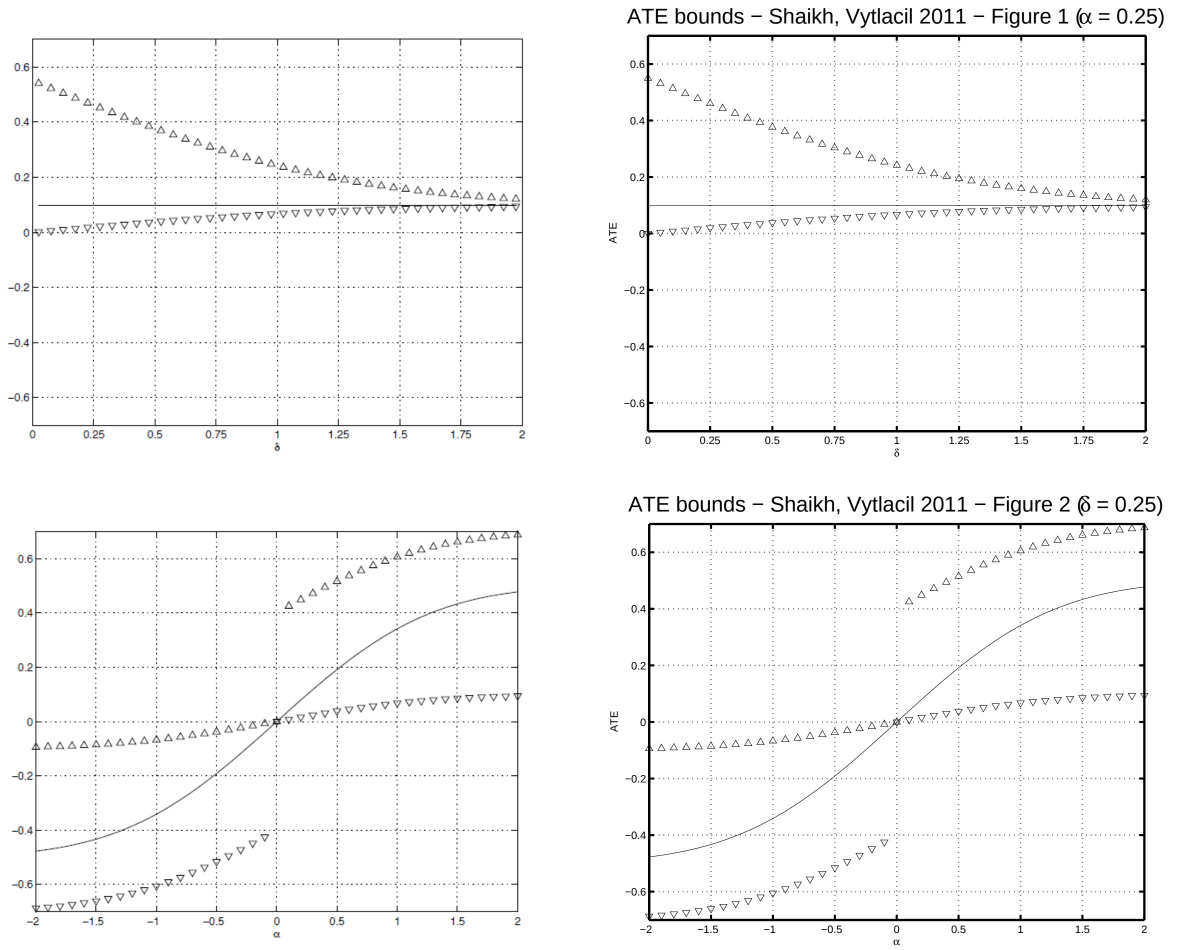


Figure 8: Bounds on ATE are compared using Shaikh and Vytlacil (2011) approach (left) and Extended GH framework (right), with X fixed ($X = 0$) and α fixed ($\alpha = 0.25$, upper pane) or δ fixed ($\delta = 0.25$, lower pane).

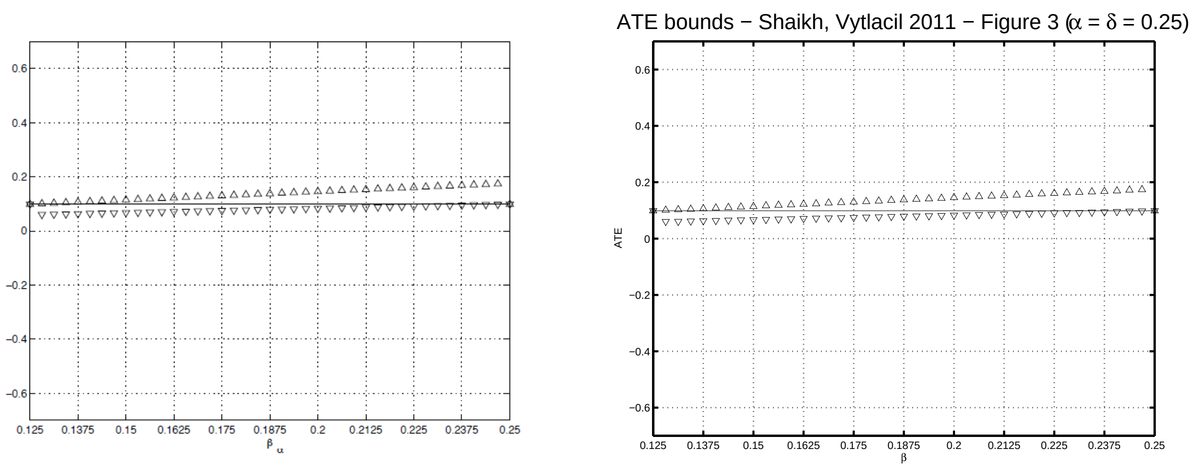


Figure 9: Bounds on ATE are compared using Shaikh and Vytlacil (2011) approach (left) and Extended GH framework (right), with variation in X ($supp(X) = \{-2, -1, 0, 1, 2\}$) and $\alpha = \beta = 0.25$ fixed.

Example 3: Bounds on Treatment Effects with Imperfect Compliance

The following subsection presents how the extended GH framework can determine sharp bounds on average causal effect when imperfect compliance is present. This was done in celebrated works of Balke and Pearl (1997, 1994) and this section replicates their results.

Consider three type of observed variables: $Y \in \{y_0, y_1\}$ is an outcome variable where y_0 stands for positive observed response, $D \in \{d_0, d_1\}$ indicates whether treatment was received (d_1) or not (d_0), and $Z \in \{z_0, z_1\}$ indicates whether a treatment was offered (z_1) or was not (z_0). An existence of unobserved U that captures individual characteristics that affects receiving of the treatment and outcome variable is also assumed. The quantity of interest is average causal effect of D on Y denoted as

$$ACE(D \to Y) = Pr(Y = y_1 | D = d_1) - Pr(Y = y_1 | D = d_0). \tag{15}$$

Restrictions that are imposed

- $Z \perp\!\!\!\perp Y | \{D, U\}$, Treatment assignment only affects the outcome variable through actual treatment D.
- $Z \perp\!\!\!\perp U$, Z and U are independent, randomization of the treatment assignments Z may deliver this property.
- no interactions between individuals or Stable Unit Treatment Value Assumption (known as SUTVA Assumption (Rubin, 1974)).

Formulation in the Extended GH framework

Following the notation of Balke and Pearl (1994), decompose the unobserved type U of an individual into two response function variables $R_D \in \{0, 1, 2, 3\}$ and $R_Y \in \{0, 1, 2, 3\}$. Pair (R_Y, R_D) is now the unobserved type (U) of the individual. Treatment D is a deterministic function of Z and R_D,

$$D = f_D(Z, R_D)$$

, where

$$\begin{aligned} f_D(z,0) = d_0 \quad , \quad f_D(z,1) &= \begin{cases} d_0, & \text{if } z = z_0, \\ d_1, & \text{if } z = z_1, \end{cases} \\ f_D(z,2) = d_1 \quad , \quad f_D(z,3) &= \begin{cases} d_1, & \text{if } z = z_0, \\ d_0, & \text{if } z = z_1. \end{cases} \end{aligned}$$

Similarly, the outcome Y is a deterministic function of D and R_Y:

$$D = f_Y(D, r_Y)$$

, where

$$\begin{aligned} f_Y(d,0) = y_0 \quad , \quad f_Y(d,1) &= \begin{cases} y_0, & \text{if } d = d_0, \\ y_1, & \text{if } d = d_1, \end{cases} \\ f_Y(d,2) = y_1 \quad , \quad f_Y(d,3) &= \begin{cases} y_1, & \text{if } d = d_0, \\ y_0, & \text{if } d = d_1. \end{cases} \end{aligned}$$

This is basically a discretization of the unobserved component U into the discrete (R_Y, R_D).

The quantity of interest is the Average Causal Effect $\theta = ACE(D \to Y) = Pr(R_Y = 1) - Pr(R_Y = 3)$, we would like to find sharp bounds on θ given $Pr(Y, D, Z)$, let also denote the probability of observed variables $p_{ijk} = Pr(Y = y_i, D = d_j, Z = z_j)$.[18] There are 5 variables in the model: observed Y, D, Z and unobserved R_Y, R_D. The mapping G between unobserved variables and observed variables is defined as

$$G(R_Y, R_D) = \{(Y, D, Z) : f_D(Z, R_D) = D, f_Y(D, R_Y) = Y\}.$$

Now we denote the joint probability distribution of observed and unobserved variables as $\pi_{ijklm} = Pr(Y = y_i, D = d_j, Z = z_k, R_Y = l, R_D = m)$.

Penalty on the points of support not compatible with G is given by

$$c_{ijklm} = \begin{cases} 0, & (y_i, d_j, z_k) \in G(l, m), \\ 1, & \text{otherwise.} \end{cases}$$

Finally, parameter θ is included in the identified set if and only if the optimized value of the following problem is equal to zero:

$$\begin{array}{rl}
\min_{(\pi)} \sum_{ijklm} \pi_{ijklm} c_{ijklm} & \\
\text{s.t.} & \\
\sum_{lm} \pi_{ijklm} = p_{ijk}, & \forall i, j, k \\
\pi_{ijklm} \sum_{ik} \pi_{ijklm} = \sum_i \pi_{ijklm} \sum_k \pi_{ijklm}, & \forall i, j, k, l, m \\
\sum_{ij} \pi_{ijklm} = \sum_{ij} p_{ijk} \sum_{ijk} \pi_{ijklm}, & \forall i, j, k, \\
\sum_{ijkm} \pi_{ijk1m} - \sum_{ijkm} \pi_{ijk3m} = \theta, & \\
\pi_{ijklm} \geq 0, & \forall i, j.
\end{array}$$

The first restriction says that the π_{ijklm} has to be compatible with p_{ijk}, which is observed from the data. The second equality states that when fixing D, R_Y, R_D (equivalent to fixing D, U) Z is independent of Y.[19] The third equation ensures that Z is marginally independent of (R_Y, R_D), whereas the forth restricts the space of joint distributions to those that have $ACE(D \to Y)$ equal to θ.

Note that the second restriction is quadratic so the whole problem is not a linear program. Quadratic restrictions might give rise to the use of semidefinite programming routines.

Although the nonlinear constraint causes significant computational difficulties, results in Balke and Pearl (1997) can be replicated to a reasonable precision (10^{-4}).

Example 4: Binary Choice Model with Zero Median Restriction

This subsection aims to capture the identification setup of binary choice model with discrete explanatory variables within the extended GH framework. Identification for

[18] $ACE(D \to Y) = Pr(Y = y_1 | D = d_1) - Pr(Y = y_1 | D = d_0) = Pr(R_Y = 1) + Pr(R_Y = 2) - (Pr(R_Y = 2) + Pr(R_Y = 3)) = Pr(R_Y = 1) - Pr(R_Y = 3)$

[19] Instrument Z only affects Y via D: $Pr(Y|D, Z, R_Y, R_D) = Pr(Y|D, R_Y, R_D)$ and this equation can be reformulated as $Pr(Y, D, Z, r_Y, r_D)Pr(D, R_Y, R_D) = Pr(Y, D, r_Y, r_D)Pr(D, Z, R_Y, R_D)$.

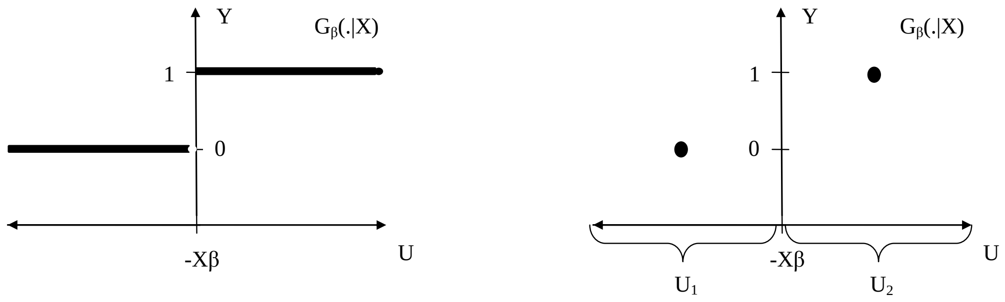


Figure 10: Left figure shows support restrictions and the figure on the right is a result of the naive discretization.

this type of problem was extensively studied in the recent work of Komarova (2009). It is well known that if all explanatory variables in binary choice model are discrete, parameters of the model are in general set rather than point identified. An identification strategy was outlined earlier (Manski and Thompson, 1986), in Komarova (2009) computationally attractive recursive procedure is outlined that determines sharp bounds on the identified set.

The problem that is studied takes the following form

- $Y = 1(X\beta + U \geq 0)$ (16)
- $Pr(U \leq 0|X = x) = 0.5 \quad \forall x \in \mathcal{X}$ (17)

where Y is the outcome variable, X is k-dimensional random variable with discrete support $\mathcal{X}$, β is k-dimensional parameter of interest and U is unobservable scalar vector variable. The only distributional assumption about U that is made is that median of U is zero conditional on X.

Discretization of Unobservables

Observed variables X is exogenous in this setup, so the analysis can be done conditional on a particular x. The identified set for β will therefore be an intersection of bounds created by conditioning on all values of X that have non-zero probability.[20] The only restriction put on the unobservable variable U is the zero median restriction, which has to be taken into account when finding a suitable discretization of U. Naive discretization is presented on figure 10 and does not allow the unobservables to meet the conditional zero median condition. When the discretization is done by virtue of Lemma 2, so the further distributional restrictions are taken into account as shown on figure 11, the discretization is rich enough to allow us to formulate the conditional zero median condition. Note that Lemma 2 proves that this discretization leaves the identified set unaffected.

[20] As with exogenous instruments, the marginal distribution of X does not have an identifying power.

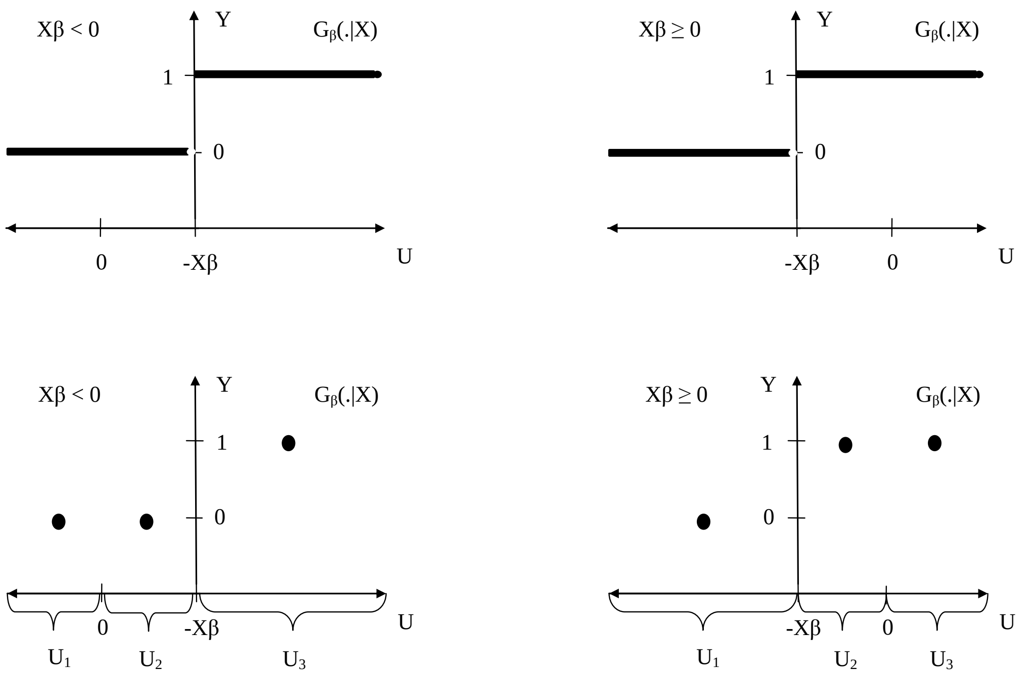


Figure 11: The two panes on the left (right) side stand for suitable discretization when $X\beta < 0$ ($X\beta \geq 0$), this discretization was obtained using Lemma 1.

Formulation in the Extended GH framework

Let $X = x$ be fixed and $p_i = Pr(Y = y_i|X = x)$, where $y_1 = 0$ and $y_2 = 1$. A penalty c_{ij},

$$c_{ij} = \begin{cases} 0, & \text{if } y_i = 1(x\beta + u_j \geq 0), \\ 1, & \text{otherwise,} \end{cases}$$

carries the information on support restrictions.

The problem can now be formulated as

$$\begin{array}{cc} \min_{(\pi)} \sum_{i,j} \pi_{ij} c_{ij} & \\ \text{s.t.} & \\ \sum_j \pi_{ij} = p_i, & \forall i \\ \sum_i \pi_{i1} = \sum_i \pi_{i2} + \sum_i \pi_{i3}, & \\ \pi_{ij} \geq 0, & \forall i, j. \end{array}$$

whenever $X\beta < 0$ and

$$\begin{array}{cc} \min_{(\pi)} \sum_{i,j} \pi_{ij} c_{ij} & \\ \text{s.t.} & \\ \sum_j \pi_{ij} = p_i, & \forall i \\ \sum_i \pi_{i1} + \sum_i \pi_{i2} = \sum_i \pi_{i3}, & \\ \pi_{ij} \geq 0, & \forall i, j. \end{array}$$

when $X\beta \geq 0$. The first set of equalities says that the joint distribution π is compatible with observed data p_i, the second equality restricts U to have zero median.[21] As in previous examples, parameter β is included in the identified set if the optimized value of the problem is equal to 0.

To simplify notation, all probabilities are implicitly conditioned on $X = x$. If $X\beta < 0$ one can immediately see that $Pr(U = u_3) = Pr(Y = 1) = p_2$ and $Pr(U = u_1) + Pr(U = u_2) = Pr(Y = 0) = p_1$. So $Pr(U = u_1) = Pr(U = u_2) + Pr(U = u_3) = 0.5$ implies that a proper distribution on U exists if and only if $Pr(U = u_3) = Pr(Y = 1) < 0.5$. On the other hand if $X\beta \geq 0$ then $Pr(U = u_2) + Pr(U = u_3) = Pr(Y = 1) = p_2$ and $Pr(U = u_1) = Pr(Y = 0) = p_1$ together with $Pr(U = u_1) + Pr(U = u_2) = Pr(U = u_3) = 0.5$ imply that $Pr(U = u_3) = Pr(Y = 1) < 0.5$, so we got precisely the same result as (25).

This example is simple, but shows how identification can be easily approached in a systematic manner.

3 Imperfect Instruments in Single Equation Endogenous Binary Response Model

As opposed to the previous section, this section shows how the extended GH framework works in a problem that has not been studied before. It is shown on the example with imperfect instruments how can the flexibility of adding extra distributional constraints help to access this problem. The extension plays a crucial role and the original GH framework can not be applied.

Identification based on instrumental variables has become a workhorse in applied research. The exogeneity of instruments cannot be tested in the just-identified case. It is of great interest to know the identifying power of this assumption. This information can serve as a sensitivity analysis, when relaxing this assumption one can see how the identified set grows. If the identified set is substantially larger if exogeneity is only slightly relaxed, more attention should be focused on the discussion about this assumption. One may then need to defend the assumption of exogeneity very well for the results to be credible. If on the other hand exogeneity of instruments is shown not to have large identifying power, the analysis could be said to be robust to some departures from the exogeneity.

In the literature different approaches have been employed to address the issue of imperfect instruments. Conley, Hansen, and Rossi (2012) parametrize the amount of instrument endogeneity and derive the identified set in the linear regression model. Hahn and Hausman (2005), rather than deriving the identified set, compare properties of OLS and TSLS estimators. Manski and Pepper (2000) made use of monotonicity of instrumental variables instead of an exogeneity assumption. Nevo and Rosen (2012) derive sharp bound on parameters under the assumption that the correlation between the instrument and an error term has the same sign as the correlation between the

[21]If $X\beta < 0$ equation (17) is equivalent to $Pr(U = u_1|X = x) = Pr(U = u_2|X = x) + Pr(U = u_3|X = x)$ and if $X\beta \geq 0$ equation (17) can be rewritten as $Pr(U = u_1|X = x) + Pr(U = u_2|X = x) = Pr(U = u_3|X = x)$. Note that this restriction can be rewritten as $\sum_{i,j} \pi_{ij}(1(U \leq 0) - 0.5) = 0$.

endogenous regressor and the error term, and the instrument is assumed to be less correlated with the error term than the endogenous regressor.

The example of the single equation endogenous binary response model from section 2 demonstrates how the extended Galichon and Henry setup can be used to trace the identified set if the strict exogeneity condition is relaxed. The way how this assumption is relaxed is the following: under the strict exogeneity restriction, $Pr(Z) \cdot Pr(U) = Pr(Z \cap U)$ for all pairs (Z, U). The distribution $Pr(Z) \cdot Pr(U)$ can be represented as a point in the $n_Z \times n_U$-dimensional unit simplex. Instead of restricting $Pr(Z \cap U)$ to be exactly equal to $Pr(Z) \cdot Pr(U)$ we will assume that the difference $Pr(Z \cap U) - Pr(Z) \cdot Pr(U)$ has to be less or equal $\alpha Pr(Z) \cdot Pr(U)$ in absolute value for some fixed $\alpha > 0$ and all (Z, U). The parameter α hence controls the amount of endogeneity of instruments. There are many ways how we may model the departure from exogeneity, however this somewhat ad hoc way of relaxing strict exogeneity is chosen so that the problem is still within linear programming framework and so that the discretization is possible.

The model under the study is (10) with support restrictions (11). In addition, instruments are not assumed to be strictly exogenous. The problem can be formulated within the extended GH framework in a following way:

$$
\begin{aligned}
&\min_{(\pi)} \textstyle\sum_{i,j,k,l} \pi_{ijkl} c_{ijkl} && \\
&\text{s.t.} && \\
&\textstyle\sum_l \pi_{ijkl} = p_{ijk}, && \forall i,j,k \\
&\textstyle\sum_{i,j,k} \pi_{ijkl} = \nu_l, && \forall l \\
&\textstyle\sum_{i,j} \pi_{ijkl} - \sum_{i,j} p_{ijk}\nu_l \leq \alpha \sum_{i,j} p_{ijk}\nu_l, && \forall k,l \\
&-\textstyle\sum_{i,j} \pi_{ijkl} + \sum_{i,j} p_{ijk}\nu_l \leq \alpha \sum_{i,j} p_{ijk}\nu_l, && \forall k,l \\
&\pi_{ijkl} \geq 0, && \forall i,j,k,l.
\end{aligned}
\tag{18}
$$

As in (2), probabilities of observed variables were generated according to (21), with Z having support on $\{-0.75, 0, 0.75\}$ with probabilities $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Results

Results of the illustration are presented on figures 12 and 13. We can see how the identified set gets larger as the departure from strict exogeneity increases.

4 Conclusion

A new method to obtain identified set as a simple extension of Galichon and Henry identification strategy was proposed so that a broader class of problems can be solved. A considerable advantage of the new method is its algorithmic structure, sharp bounds of the identified set need not be derived from case to case, but efficient algorithms can be employed to trace the identified set independently of the structure of the problem.

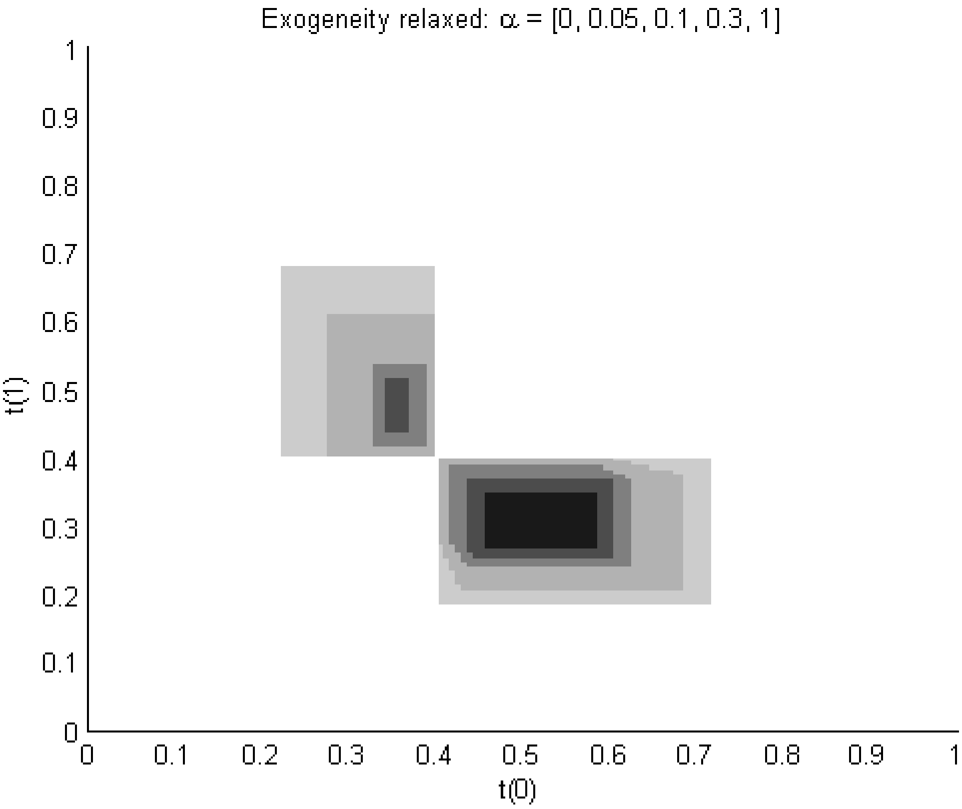


Figure 12: Identified sets corresponding to different values of parameter α, the case with a strong instrument. Darker areas stands for stronger exogenenity. Note that observed probabilities together with the assumption of uniform U and support restrictions given by the economic model do have some identifying power even if the instrument is completely endogenous.

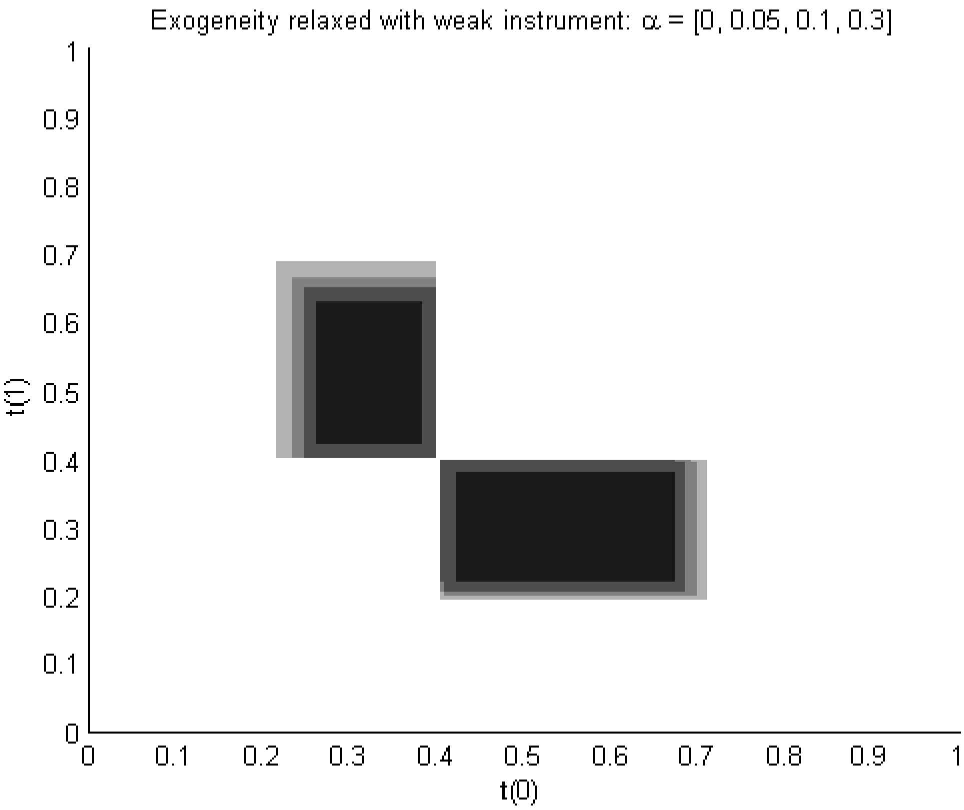


Figure 13: Identified sets corresponding to different values of parameter α, case with weak instrument. Darker areas stands for stronger exogenenity.

Some existing identification results were replicated in a fairly straightforward manner. Moreover, the new method allowed us to study the impact of relaxing the strict exogeneity in non-linear model with discrete variables.

The main message is that if observed variables are discrete identification can be attacked at the lowest level: by searching in the space of joint distribution functions of observed and unobserved variables. This delivers great flexibility in studying the identifying power of different sets of assumptions.

The proposed method also allows us to access *local* identification if weaker local restrictions are made. The approach is no different than in the study of global identification.

How to make this method operational in a continuous case, e.g. an analog of condition (4) and how to do statistical inference, remain open questions. The iterative subsampling scheme of Romano and Shaikh (2010) appears to be helpful. Further research is warranted.

References

Andrews, D. W. and X. Shi (2012): "Inference Based on Conditional Moment Inequalities," *Econometrica*, to appear.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46, 313–324.

Balke, A. and J. Pearl (1994): "Counterfactual Probabilities: Computational Methods, Bounds, and Applications," in *Uncertainty in Artificial Intelligence 10*, ed. by L. R. de Mantaras and D. Poole, Morgan Kaufmann, 46–54.

——— (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 439, 1172–1176.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

——— (2012): "Partial identification using random set theory," *Journal of Econometrics*, 166, 17 – 32.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Brock, W. A. and S. N. Durlauf (2001): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235–260.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Finite sample inference for quantile regression models," *Journal of Econometrics*, 152, 93–103.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models1," *Econometrica*, 75, 1243–1284.

CHERNOZHUKOV, V., S. S. LEE, AND A. ROSEN (2012): "Intersection Bounds: estimation and inference," *Econometrica*, to appear.

CHESHER, A. (2009): "Single equation endogenous binary reponse models," CeMMAP working papers CWP23/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

——— (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575–601.

CHESHER, A., A. ROSEN, AND K. SMOLINSKI (2011): "An instrumental variable model of multiple discrete choice," CeMMAP working papers CWP06/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

CONLEY, T. G., C. B. HANSEN, AND P. E. ROSSI (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272.

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

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

——— (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies, Vol. 78, No. 4, pp. 1264-1298, 2011*.

HAHN, J. AND J. HAUSMAN (2005): "Estimation with Valid and Invalid Instruments," *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, pp. 25–57.

HENRY, M., R. MEANGO, AND M. QUEYRANNE (2011): "Combinatorial Bootstrap Inference in Partially Identified Incomplete Structural Models," *SSRN eLibrary*.

IMBENS, G. W. AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

KOMAROVA, T. (2009): "Binary Choice Models with Discrete Regressors: Identification and Misspecification," .

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 319–23.

——— (1995): *Identification Problems in the Social Sciences*, Cambridge, Harvard University Press.

——— (2003): *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

——— (2008): "partial identification in econometrics," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Basingstoke: Palgrave Macmillan.

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables, with an Application to the Returns to Schooling," *Econometrica*, 68, 997–1012.

MANSKI, C. F. AND T. S. THOMPSON (1986): "Operational characteristics of maximum score estimation," *Journal of Econometrics*, 32, 85–108.

MOURIFIÉ, I. AND M. HENRY (2012): "Set inference in latent variables models," *The Econometrics Journal*, to appear.

NEVO, A. AND A. M. ROSEN (2012): "Identification with Imperfect Instruments," *Review of Economics and Statistics*, 93, 127–137.

ROMANO, J. P. AND A. M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

ROSEN, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146, 107–117.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

SHAIKH, A. M. AND E. J. VYTLACIL (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955.

TAMER, E. T. (2010): "Partial Identification in Econometrics," *Annual Review of Economics, Vol. 2, pp. 167-195, 2010.*

A Proofs

A.1 Proof of Lemma 1

Proof. We have to show that there exists $\pi_1 : \mathcal{Y} \times \mathcal{U} \mapsto [0,1]$ satisfying

$$\sum_{i=1}^{n} \int_{u \in \mathcal{U}} \pi_1(y_i, u) 1(y_i \in G(u)) du = 1, \quad \text{(C1)}$$

$$\forall i = 1, ..., n: \quad \int_{u \in \mathcal{U}} \pi_1(y_i, u) du = p_i, \quad \text{(C2)}$$

$$\forall u \in \mathcal{U}: \quad \sum_{i=1}^{n} \pi_1(y_i, u) = \nu(u), \quad \text{(C3)}$$

$$\forall I \in \mathbb{I}; \forall u \in \mathcal{U}: \quad \left| \sum_{i \in I} \pi_1(y_i, u) - \sum_{i \in I} p_i \nu(u) \right| \leq \alpha \sum_{i \in I} p_i \nu(u). \quad \text{(C4)}$$

$$\forall i = 1, ..., n; \forall u \in \mathcal{U}: \quad \pi_1(y_i, u) \geq 0 \quad \text{(C5)}$$

if and only if there exists $\pi_2 : \mathcal{Y} \times \mathcal{U}^* \mapsto [0,1]$ satisfying

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_2(y_i, u_j^*) 1(y_i \in G(u_j^*)) du = 1, \quad \text{(D1)}$$

$$\forall i = 1, ..., n: \quad \sum_{j=1}^{m} \pi_2(y_i, u_j^*) = p_i, \quad \text{(D2)}$$

$$\forall j = 1, ..., m: \quad \sum_{i=1}^{n} \pi_2(y_i, u_j^*) = \nu^*(u_j^*), \quad \text{(D3)}$$

$$\forall I \in \mathbb{I}; \forall j = 1, ..., m: \quad \left| \sum_{i \in I} \pi_2(y_i, u_j^*) - \sum_{i \in I} p_i \nu^*(u_j^*) \right| \leq \alpha \sum_{i \in I} p_i \nu^*(u_j^*). \quad \text{(D4)}$$

$$\forall i = 1, ..., n; \forall j = 1, ..., m: \quad \pi_2(y_i, u_j^*) \geq 0 \quad \text{(D5)}$$

"($\Rightarrow$)" - Given π_1, we will construct π_2 according to

$$\forall i = 1, ..., n; \forall j = 1, ..., m: \quad \pi_2(y_i, u_j^*) = \int_{\Delta_j} \pi_1(y_i, u) du, \quad (\Pi_2)$$

and this will ensure that {(C1),(C2),(C3M),(C4M),(C5)} implies {(D1),(D2),(D3M),(D4M),(D5)} as shown below:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\pi_2(y_i,u_j^*)1(y_i\in G(u_j^*)) \overset{(\Pi_2)}{=} \sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)du1(y_i\in G(u_j^*)) \overset{(\text{PartU1})}{=}$$

$$\overset{(\text{PartU1})}{=} \sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)1(y_i\in G(u))du = \sum_{i=1}^{n}\int_{u\in\mathcal{U}}\pi_1(y_i,u)1(y_i\in G(u))du \overset{(C1)}{=} 1,$$

$$\forall i: \qquad \sum_{j=1}^{m}\pi_2(y_i,u_j^*) \overset{(\Pi_2)}{=} \sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)du = \int_{u\in\mathcal{U}}\pi_1(y_i,u)du \overset{(C2)}{=} p_i,$$

$$\forall j: \quad \sum_{i=1}^{n}\pi_2(y_i,u_j^*) \overset{(\Pi_2)}{=} \sum_{i=1}^{n}\int_{\Delta_j}\pi_1(y_i,u)du = \int_{\Delta_j}\sum_{i=1}^{n}\pi_1(y_i,u)du \overset{(C3)}{=} \int_{\Delta_j}\nu(u)du \overset{(P)}{=} \nu^*(u_j^*),$$

$$\forall j,\forall I: \qquad \left|\sum_{i\in I}\pi_2(y_i,u_j^*) - \sum_{i\in I}p_i\nu^*(u_j^*)\right| \overset{(\Pi_2),(P)}{=} \left|\sum_{i\in I}\int_{\Delta_j}\pi_1(y_i,u)du - \sum_{i\in I}p_i\int_{\Delta_j}\nu(u)du\right| =$$

$$= \left|\int_{\Delta_j}\left(\sum_{i\in I}\pi_1(y_i,u) - \sum_{i\in I}p_i\nu(u)\right)du\right| \overset{(C4)}{\leq} \left|\int_{\Delta_j}\alpha\sum_{i\in I}p_i\nu(u)du\right| = \alpha\sum_{i\in I}p_i\nu^*(u_j^*).$$

$$\forall i,\forall j: \qquad \pi_2(y_i,u_j^*) \overset{(\Pi_2)}{=} \int_{\Delta_j}\pi_1(y_i,u)du \overset{(C5)}{\geq} \int_{\Delta_j}0du = 0,$$

"$(\Leftarrow)$" - If we know π_2 we obtain π_1 using

$$\forall i=1,...,n;\forall j=1,...,m;\forall u\in\Delta_j: \quad \pi_1(y_i,u) = \pi_2(y_i,u_j^*)\frac{\nu(u)}{\nu^*(u_j^*)}, \qquad (\Pi_1)$$

(note that (Π_1) implies (Π_2)) and we now show that {(D1),(D2),(D3M),(D4M),(D5)} implies {(C1),(C2),(C3M),(C4M),(C5)}:

$$\sum_{i=1}^{n} \int_{u \in \mathcal{U}} \pi_1(y_i, u) 1(y_i \in G(u)) du = \sum_{i=1}^{n} \sum_{j=1}^{m} \int_{\Delta_j} \pi_1(y_i, u) 1(y_i \in G(u)) du \overset{\text{(PartU1)}}{=}$$

$$\overset{\text{(PartU1)}}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} \int_{\Delta_j} \pi_1(y_i, u) du 1(y_i \in G(u_j^*)) \overset{(\Pi_1)}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_2(y_i, u_j^*) 1(y_i \in G(u_j^*)) \overset{\text{(D1)}}{=} 1,$$

$$\forall i: \quad \int_{u \in \mathcal{U}} \pi_1(y_i, u) du = \sum_{j=1}^{m} \int_{\Delta_j} \pi_1(y_i, u) du \overset{(\Pi_1)}{=} \sum_{j=1}^{m} \pi_2(y_i, u_j^*) \overset{\text{(D2)}}{=} p_i,$$

$$\forall j, \quad \forall u \in \Delta_j: \sum_{i=1}^{n} \pi_1(y_i, u) \overset{(\Pi_1)}{=} \sum_{i=1}^{n} \pi_2(y_i, u_j^*) \frac{\nu(u)}{\nu^*(u_j^*)} \overset{\text{(D3)}}{=} \nu(u),$$

$$\forall j, \forall I, \quad \forall u \in \Delta_j: \left| \sum_{i \in I} \pi_1(y_i, u) - \sum_{i \in I} p_i \nu(u) \right| \overset{(\Pi_1)}{=} \left| \sum_{i \in I} \pi_2(y_i, u_j^*) \frac{\nu(u)}{\nu^*(u_j^*)} - \sum_{i \in I} p_i \frac{\nu(u)}{\nu^*(u_j^*)} \nu^*(u_j^*) \right| =$$

$$= \left| \frac{\nu(u)}{\nu^*(u_j^*)} \left(\sum_{i \in I} \pi_2(y_i, u_j^*) - \sum_{i \in I} p_i \nu_j^* \right) \right| \overset{\text{(D4)}}{\leq} \left| \alpha \sum_{i \in I} p_i \nu(u) \right| = \alpha \sum_{i \in I} p_i \nu(u),$$

$$\forall i, \forall j, \quad \forall u \in \Delta_j: \pi_1(y_i, u_j) \overset{(\Pi_1)}{=} \pi_2(y_i, u_j^*) \frac{\nu(u)}{\nu^*(u_j^*)} \overset{\text{(D5)}}{\geq} 0,$$

□

A.2 Proof of Lemma 2

Proof. Similarly to the proof of Lemma 1, we have to show that there exists $\pi_1 : \mathcal{Y} \times \mathcal{U} \mapsto [0,1]$ satisfying (C1),(C2),(C5) and

$$\sum_{i=1}^{n} \int_{u \in \mathcal{U}} \pi_1(y_i, u) \phi(u) du = 0, \tag{C3M}$$

$$\forall I \in \mathbb{I}; \forall u \in \mathcal{U}: \quad \left| \sum_{i \in I} \pi_1(y_i, u) - \sum_{i \in I} p_i \sum_{i=1}^{n} \pi_1(y_i, u) \right| \leq \alpha \sum_{i \in I} p_i \sum_{i=1}^{n} \pi_1(y_i, u). \tag{C4M}$$

if and only if there exists $\pi_2 : \mathcal{Y} \times \mathcal{U}^* \mapsto [0,1]$ satisfying (D1),(D2),(D5) and

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_2(y_i, u_j^*) \phi(u_j^*) du = 0, \tag{D3M}$$

$$\forall I \in \mathbb{I}; \forall j = 1, ..., m: \quad \left| \sum_{i \in I} \pi_2(y_i, u_j^*) - \sum_{i \in I} p_i \sum_{i=1}^{n} \pi_2(y_i, u_j^*) \right| \leq$$

$$\leq \alpha \sum_{i \in I} p_i \sum_{i=1}^{n} \pi_2(y_i, u_j^*). \tag{D4M}$$

"(⇒)" - Given π_1, will construct π_2 according to

$$\forall i = 1, ..., n; \forall j = 1, ..., m: \quad \pi_2(y_i, u_j^*) = \int\limits_{\Delta_j} \pi_1(y_i, u) du, \tag{Π_2}$$

and this will ensure that {(C1),(C2),(C3M),(C4M),(C5)} imply {(D1),(D2),(D3M),(D4M),(D5)}. Because the partitioning of the $\mathcal{U}$ space using (PartU2) is finer than the one of (PartU1) we get that {(C1),(C2),(C5)} imply {(D1),(D2),(D5)} immediately using the proof of Lemma 1. It is therefore sufficient to show that {(C3M),(C4M)} imply {(D3M),(D4M)}:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \pi_2(y_i, u_j^*)\phi(u_j^*) \overset{(\Pi_2)}{=} \sum_{i=1}^{n}\sum_{j=1}^{m} \int\limits_{\Delta_j} \pi_1(y_i, u) du\ \phi(u_j^*) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \int\limits_{\Delta_j} \pi_1(y_i, u)\phi(u^*) du \overset{(\text{PartU2})}{=} \sum_{i=1}^{n}\sum_{j=1}^{m} \int\limits_{\Delta_j} \pi_1(y_i, u)\phi(u) du =$$

$$= \sum_{i=1}^{n} \int\limits_{u\in\mathcal{U}} \pi_1(y_i, u)\phi(u) du \overset{(\text{C3M})}{=} 0,$$

$$\forall I \in \mathbb{I}; \forall j = 1, ..., m: \quad \left|\sum_{i\in I} \pi_2(y_i, u_j^*) - \sum_{i\in I} p_i \sum_{i=1}^{n} \pi_2(y_i, u_j^*)\right| \overset{(\Pi_2)}{=}$$

$$\overset{(\Pi_2)}{=} \left|\sum_{i\in I} \int\limits_{\Delta_j} \pi_1(y_i, u) du - \sum_{i\in I} p_i \sum_{i=1}^{n} \int\limits_{\Delta_j} \pi_1(y_i, u) du\right| =$$

$$= \left|\int\limits_{\Delta_j} \left(\sum_{i\in I} \pi_1(y_i, u) - \sum_{i\in I} p_i \sum_{i=1}^{n} \pi_1(y_i, u)\right) du\right| \overset{(\text{C4M}),(\Pi_2)}{\leq}$$

$$\overset{(\text{C4M}),(\Pi_2)}{\leq} \left|\alpha \sum_{i\in I} p_i \sum_{i=1}^{n} \pi_2(y_i, u_j^*)\right| = \alpha \sum_{i\in I} p_i \sum_{i=1}^{n} \pi_2(y_i, u_j^*).$$

"(⇐)" - Knowing π_2 we obtain π_1 using

$$\forall i = 1, ..., n; \forall j = 1, ..., m; \forall u \in \Delta_j: \quad \pi_1(y_i, u) = \pi_2(y_i, u_j^*) \frac{\gamma(u)}{\int\limits_{u\in\Delta_j} \gamma(u) du}, \tag{Π_1}$$

where γ is an arbitrary strictly positive probability density function. It is now sufficient to show that {(D3M),(D4M) (D5)} imply {(C3M),(C4M),(C5)} because the proof of Lemma 1 reveals that {(C1),(C2)} imply {(D1),(D2)} and (PartU2) provides a finer discretization of $\mathcal{U}$ than (PartU1) does:

$$\sum_{i=1}^{n}\int_{u\in\mathcal{U}}\pi_1(y_i,u)\phi(u)du=\sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)\phi(u)du\overset{(\text{PartU2})}{=}$$
$$\overset{(\text{PartU2})}{=}\sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)\phi(u_j^*)du=\sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\Delta_j}\pi_1(y_i,u)du\;\phi(u_j^*)\overset{(\Pi_1)}{=}$$
$$\overset{(\Pi_1)}{=}\sum_{i=1}^{n}\sum_{j=1}^{m}\pi_2(y_i,u_j^*)\phi(u_j^*)\overset{(\text{D3M})}{=}0,$$
$$\forall j,\forall I,\qquad \forall u\in\Delta_j:\left|\sum_{i\in I}\pi_1(y_i,u)-\sum_{i\in I}p_i\sum_{i=1}^{n}\pi_1(y_i,u)\right|\overset{(\Pi_1)}{=}$$
$$\overset{(\Pi_1)}{=}\left|\sum_{i\in I}\pi_2(y_i,u_j^*)\frac{\gamma(u)}{\int\limits_{u\in\Delta_j}\gamma(u)du}-\sum_{i\in I}p_i\sum_{i=1}^{n}\pi_2(y_i,u_j^*)\frac{\gamma(u)}{\int\limits_{u\in\Delta_j}\gamma(u)du}\right|=$$
$$=\left|\frac{\gamma(u)}{\int\limits_{u\in\Delta_j}\gamma(u)du}\left(\sum_{i\in I}\pi_2(y_i,u_j^*)-\sum_{i\in I}p_i\sum_{i=1}^{n}\pi_2(y_i,u_j^*)\right)\right|\overset{(\text{D4}),(\Pi_1)}{\leq}$$
$$\overset{(\text{D4}),(\Pi_1)}{\leq}\left|\alpha\sum_{i\in I}p_i\sum_{i=1}^{n}\pi_1(y_i,u)\right|=\alpha\sum_{i\in I}p_i\sum_{i=1}^{n}\pi_1(y_i,u),$$
$$\forall i,\forall j,\qquad \forall u\in\Delta_j:\pi_1(y_i,u)\overset{(\Pi_1)}{=}\pi_2(y_i,u_j^*)\frac{\gamma(u)}{\int\limits_{u\in\Delta_j}\gamma(u)du}\overset{(\text{D5})}{\geq}0,$$

□

B Technical Details on Presented Examples

B.1 Example 1

B.1.1 Chesher's approach

In order to present the identification result from Chesher (2009), the basic definitions have to be introduced. The notation that is being used is different to the one of Galichon and Henry that is employed in this paper.

- A **model** $\mathcal{M}$ is defined as (10) with $U \sim Unif(0,1)$ and $U \perp\!\!\!\perp Z$ for all $Z \in \mathcal{Z}$.
- A **structure** $S \equiv \{t, F_{UX|Z}\}$ is a pair of a threshold-crossing function t and a cumulative distribution function of the conditional distribution of U and X given Z.
- A structure S is said to be **admitted by a model** $\mathcal{M}$ if $F_{UX|Z}$ respects the independence property, that is $F_U(u|z) \equiv F_{UX|Z}(u,\bar{x}|z) = u$ for all $u \in (0,1)$ and all $z \in \mathcal{Z}$, where $\bar{x}$ is the upper bound of X.

- A structure S **generates** the joint distribution of Y and X given Z if $F_{YX|Z}(0,x|z) = F_{UX|Z}(t(x),x|z)$.
- Two structures $S^* \equiv \{t^*, F^*_{UX|Z}\}$ and $S^0 \equiv \{t^0, F^0_{UX|Z}\}$ are said to be **observationally equivalent** if they generate the same distribution of Y and X given Z for all $z \in \mathcal{Z}$, that is if $F^*_{YX|Z}(0,x|z) \equiv F^*_{UX|Z}(t^*(x),x|z) = F^0_{YX|Z}(0,x|z) \equiv F^0_{UX|Z}(t^0(x),x|z)$ for all $z \in \mathcal{Z}$ and for all $x \in \mathcal{X}$.

Theorem 1 from Chesher (2009) states that having a structure S_0 admitted by the model $\mathcal{M}$ that generates the conditional distribution of Y and X given Z with cumulative distribution function $F^0_{YX|Z}$ and if this threshold-crossing function t is in structure S admitted by model $\mathcal{M}$ observationally equivalent to S^0, then t satisfies

$$\begin{aligned} c_{0l}(u,z;p) &= Pr_0[Y=0 \cap t(X) < u | Z=z] < u, \forall u \in (0,1),\ \forall z \in \mathcal{Z} && (19)\\ c_{0u}(u,z;p) &= 1 - Pr_0[Y=1 \cap u \leq t(X)|Z=z] \geq u, \forall u \in (0,1),\ \forall z \in \mathcal{Z}. && (20) \end{aligned}$$

where Pr_0 states that probabilities were calculated using the measure that was generated by S^0, that is using $F^0_{YX|Z}$ and l and u stand for lower and upper bound respectively.

Under continuity of X the converse is also true, this is equal to say that the set of all functions p satisfying above set of inequalities is *sharply defined identified set*. In Chesher (2010) this theorem is proven even for a more general setup. It is important to note that the proof is constructive, so that for a given threshold-crossing function t, suitable distribution function $F_{UX|Z}$ is constructed such that $\{t, F_{UX|Z}\}$ is admitted by the model $\mathcal{M}$ and generates $F_{YX|Z}$ that is observed in the data. This highlights the link to GH setup since there the aim is to find the joint probability distribution that satisfies the independence restriction, has correct marginals and places all the probability on those combinations of variables that are compatible with the data.

B.1.2 Illustration: Discrete endogenous variable

Construction of true data-generating process

The following example is taken from Chesher (2010), suppose that both Y and X are binary: $Y \equiv 1(Y^* \geq 0)$ and $X \equiv 1(X^* \geq 0)$, where Y^* and X^* were generated in the following way:

$$\begin{gathered} Y^* = \theta_0 + \theta_1 X + W, \quad X^* = b_0 + b_1 Z + V \\ \begin{bmatrix} W \\ V \end{bmatrix} \perp\!\!\!\perp Z, \quad \begin{bmatrix} W \\ V \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \right) \end{gathered} \tag{21}$$

with parameters

$$(\theta_0, \theta_1, b_0, b_1, r) = (0, 0.5, 0, 1, -0.25). \tag{22}$$

and the instrument Z takes values in $\mathcal{Z} = \{-0.75, 1, 0.75\}$.

The econometrician however does not know how the data were generated. She only assumes (10) and $U \perp\!\!\!\perp Z$, $U \sim Unif(0,1)$ and $t(X) = \Phi(-\theta_0 - \theta_1 X)$ and observes the distribution of observable variables p_{ijk}.[22] Even though it is impossible to recover

[22]Observed probabilities p_{ijk} were obtain using Matlab function `mvtnorm`.

the true value of $\theta = (0, 0.5)$ exactly, it is possible to at least create informative bounds on it.

Since X threshold-crossing function t attains only two values $t(0) = \Phi(-\theta_0) = 0.5$ and $t(1) = \Phi(-\theta_0 - \theta_1) = 0.308$.

B.1.3 Illustration: Continuous endogenous variable

Construction of true data-generating process

Suppose that the economic model is described by (10) and the data generating process by (21) with the following set of parameters

$$(\theta_0, \theta_1, b_0, b_1, s_{wv}, s_{vv}) = (0, -1, 0, 0.3, 0.5, 1). \tag{23}$$

as before with the only difference that X is not binary anymore ($X = X^*$).

The distribution of observable variables $(Y^*, X|Z = z)$ (Y^* and X given $Z = z$) is given by $N(\mu(z), \Sigma)$, where

$$\mu(z) = \begin{bmatrix} \theta_0 + \theta_1 b_0 + \theta_1 b_1 z \\ b_0 + b_1 z \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 + 2\theta_1 s_{wv} + \theta_1^2 s_{vv} & s_{wv} + \theta_1 s_{vv} \\ s + wv + \theta_1 s_{vv} & s_{vv} \end{bmatrix}$$

Details of the simulations are provided here. Because of the continuity of X the unobservable U was discretized as equidistant point masses on $[0, 1]$. The distribution of observables is given by

$$p_{ijk} = Pr(Y = y_i \cap X = x_j \cap Z = z_k) = Pr(Y = y_i \cap X = x_j | Z = z_k) Pr(Z = z_k).$$

It is known that $(Y^*, X|Z) \sim N(\mu(z), \Sigma)$ and a suitable discretization of X is needed. It is easy to show that the density of $(Y^*|X = x, Z = z)$ is

$$N\left(\mu(z)_1 + \frac{\Sigma_{21}}{\Sigma_{22}}(x - \mu(z)_2), \left(1 - \sqrt{\frac{\Sigma_{21}^2}{\Sigma_{11}\Sigma_{22}}}\right)\Sigma_{11}\right).$$

Integrating corresponding probability density function at $(-\infty, 0)$ gives us $Pr(Y = 0|X = x, Z = z)$. The distribution of X given $Z = z$ is $N(b_0 + b_1 z, s_{vv})$, but now the question is how to discretize the support of X which is $\mathbb{R}$. If the number of nods is n_x then one suggestion would be to set the z to its mean value, that is to 0, and set values of discretized support of X to n_x equidistant quantiles.[23] Eventhough this discretization seems natural it brings some degree of arbitrariness to the problem.

Finally, taking together all the pieces yields

$$p_{ijk} = Pr(Y = y_i | X = x_j, Z = z_k) Pr(X = x_j | Z = z_k) Pr(Z = z_k),$$

where all quantities on the right-hand side are known.

[23] 0% and 100% quantiles are excluded.

B.2 Example 2

B.2.1 Illustration

True data-generating process

For the illustration (ϵ_1, ϵ_2) are assumed to be $N(0, I_2)$. This assumption together with (13) and (14) generate distribution of Y and D given X and Z. Support of Z is assumed to be $\{-1, 1\}$ and support of X is either $\{0\}$ or $\{-2, -1, 0, 1, 2\}$, (X, Z) are assumed to be uniformly distributed.[24]

B.3 Example 3

B.3.1 Balke and Pearl's Approach

Balke and Pearl (1997) made use of the fact that these restrictions impose a following decomposition on the joint distribution of (Y, D, Z, U):

$$Pr(Y, D, Z, U) = Pr(Y|D, U)Pr(D|Z, U)Pr(Z)Pr(U). \tag{24}$$

There exist four different functions from Z to D and four different functions from D to Y hence 16 different *types* of individuals that we can consider. Hence one can think of U having a discrete support with 16 points, each point represents a pair of functions one from Z to D and second from D to Y. For instance one type u may be persons who always accept treatment and who do not have positive outcome irrespective of the treatment. Bounds on (15) are found using linear program searching through the space of distributions of the types (U) subject to the joint distribution to be compatible with observed data $Pr(y, d|z)$. Full setup with discussions can be found in Balke and Pearl (1997, 1994).

B.4 Example 4

B.4.1 Komarova's Approach

Following Manski and Thompson (1986)

$$Pr(Y = 1|X = x) = 1 - Pr(U < -x\beta|X = x)$$

together with zero median restriction (17) implies

$$Pr(Y = 1|X = x) \geq 0.5 \;\Leftrightarrow\; x\beta \geq 0. \tag{25}$$

Therefore bounds on parameter vector β are obtained as an intersection of linear half spaces. In Komarova (2009) recursive procedure is proposed that translates this set of linear inequalities into bounds on parameters.

[24] As in example 1, the distribution of exogenous variables per se does not have any identifying power. It is added purely for the simplicity of the exposition.

C Implementation issues

C.1 Extended GH framework

Following routines were used and compared in order to solve linear program (12).

- `linprog`[25] - Matlab built in function from Optimization Toolbox. Interior point method was superior to simplex method because of the computational time. Since the objective value is not minimized to exact zeros, certain threshold had to be employed. Natural choice is the tolerance level of the optimization routine (10^{-8} for $n_x = n_u = 40$ was used). Results for the two approaches were identical.
- `GNU Linear Programming Kit (GLPK)` - Modified simplex method from Matlab MEX interface for the GLPK library[26]. Significantly faster than `linprog` with similar results.

Linear program is an old and well understood problem however if the discretization of X and U is large then the matrix that encodes the restrictions for the joint distribution[27] can reach the limits of Matlab's largest array that can be created. For instance if the sizes of supports are $n_x = n_u = 40$ together with $n_y = 2$ and $n_z = 10$, then the joint probability π_{ijkl} has 32000 elements. So that the matrix that carries the information about restrictions on π_{ijkl} will have 32000 columns.

[25] http://www.mathworks.com/help/toolbox/optim/ug/linprog.html
[26] http://glpkmex.sourceforge.net/
[27] this is a 4-dimensional array π_{ijkl} stacked into a vector

Issued in the series Discussion Papers 2012

2012

01/12 February, **Ola Honningdal Grytten**, “The Protestant Ethic and the Spirit of Capitalism the Haugian Way”.

02/12 February**, Alexander W. Cappelen**, Rune Jansen Hagen, **Erik Ø. Sørensen**, and **Bertil Tungodden**, «Do non-enforceable contracts matter? Evidence from an international lab experiment”.

03/12 February, **Alexander W. Cappelen** and **Bertil Tungodden**, “Tax policy and fair inequality”.

04/12 March, Mette Ejrnæs and **Astrid Kunze**, «Work and Wage Dynamics around Childbirth”.

05/12 March, **Lars Mathiesen**, “Price patterns resulting from different producer behavior in spatial equilibrium”.

06/12 March, **Kurt R. Brekke**, Luigi Siciliani, and Odd Rune Straume, “Hospital competition with soft budgets”.

07/12 March, **Alexander W. Cappelen** and **Bertil Tungodden**, “Heterogeneity in fairness views - a challenge to the mutualistic approach?”

08/12 March, Tore Ellingsen and **Eirik Gaard Kristiansen**, “Paying for Staying: Managerial Contracts and the Retention Motive”.

09/12 March, **Kurt R. Brekke**, Luigi Siciliani, and Odd Rune Straume, “Can competition reduce quality?”

10/12 April, Espen Bratberg, **Øivind Anti Nilsen**, and Kjell Vaage, “Is Recipiency of Disability Pension Hereditary?”

11/12 May, **Lars Mathiesen**, **Øivind Anti Nilsen**, and **Lars Sørgard**, “A Note on Upward Pricing Pressure: The possibility of false positives”.

12/12 May, **Bjørn L. Basberg**, “Amateur or professional? A new look at 19th century patentees in Norway”.

13/12 May, Sandra E. Black, Paul J. Devereux, Katrine V. Løken, and **Kjell G. Salvanes**, “Care or Cash? The Effect of Child Care Subsidies on Student Performance”.

14/12 July, **Alexander W. Cappelen**, Ulrik H. Nielsen, **Erik Ø. Sørensen**, **Bertil Tungodden**, and Jean-Robert Tyran, " Give and Take in Dictator Games".

15/12 August, **Kai Liu**, "Explaining the Gender Wage Gap: Estimates from a Dynamic Model of Job Changes and Hours Changes".

16/12 August, **Kai Liu**, **Kjell G. Salvanes**, and **Erik Ø. Sørensen**, «Good Skills in Bad Times: Cyclical Skill Mismatch and the Long-term Effects of Graduating in a Recession".

17/12 August, **Alexander W. Cappelen**, **Erik Ø. Sørensen**, and **Bertil Tungodden**, «When do we lie?».

18/12 September, **Kjetil Bjorvatn** and Tina Søreide, «Corruption and competition for resources".

19/12 September, **Alexander W. Cappelen** and Runa Urheim, "Pension Funds, Sovereign-wealth Funds and Intergenerational Justice"

20/12 October, **Ingvild Almås** and **Erik Ø. Sørensen**, "Global Income Inequality and Cost-of-Living Adjustment: The Geary–Allen World Accounts".

21/12 November, **Ingvild Almås** and Åshild Auglænd Johnsen, "The cost of living in China: Implications for inequality and poverty".

22/12 December, **Alexander W. Cappelen, Tom Eichele, Kenneth Hugdahl, Karsten Specht, Erik Ø. Sørensen**, and **Bertil Tungodden**, "Fair inequality: a neureconomic study".

23/12 December, **Ingvild Almås, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen,** and **Bertil Tungodden**, «Willingness to compete: family matters».

24/12 December, **Ingvild Almås, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen**, and **Bertil Tungodden**, «Willingness to compete in a gender equal society».

25/12 December, Ari Hyytinen, **Frode Steen**, and Otto Toivanen, "Anatomy of Cartel Contracts".

26/12 December, **Sandra E. Black**, Paul J. Devereux, and **Kjell G. Salvanes**, "Losing Heart? The Effect of Job Displacement on Health".

27/12 December, **Frode Steen** and **Lars Sørgard**, "Disadvantageous Semicollusion: Price competition in the Norwegian airline industry"

2013

01/13 January, **Lukáš Lafférs**, “Identification in Models with Discrete Variables”.

NHH

Norges Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no