Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2011-28

Identification of Nonseparable Models with Endogeneity and Discrete Instruments

X. D'HAULTFOEUILLE[1]
Ph. FÉVRIER[2]

[1] CREST. Email : xavier.dhaultfoeuille@ensae.fr

[2] CREST. Email : philippe.fevrier@ensae.fr

Identification of Nonseparable Models with Endogeneity and Discrete Instruments*

Xavier D'Haultfœuille † Philippe Février‡

May 2011

Abstract

We study the identification of a nonseparable function which relates a continuous outcome to a continuous endogenous variable. We suppose to have in hand a strongly exogenous instrument, and assume both a monotonicity and a rank similarity condition. We show that the combination of these restrictions has a large identifying power: full identification can be achieved even though the instrument is discrete. To prove our results, we rely on group and dynamical systems theories. The identification of the model depends on the properties of the orbits of a group generated by a well defined set of identified functions. Two cases are distinguished, depending on whether there exists a function in this group which admits a fixed point. In the first case, the univariate model if fully identified. In the second one, the univariate model is identified on a countable set with a binary instrument and fully identified in general when the instrument takes at least three values. We partially extend these results to multivariate endogenous variables.

Keywords: Nonparametric Identification, Discrete Instrument, Control Variable, Fixed Points, Group Theory.

JEL classification numbers: C14.

---

*We have benefited from discussions with Ivan Canay, Clément de Chaisemartin, Sylvain Chabé-Ferret, Andrew Chesher, James Heckman, Toru Kitagawa, Jean-Marc Robin, Susanne Schennach, Azeem Shaikh, Elie Tamer and Edward Vytlacil. We also thank participants of the CEMMAP, Northwestern and Chicago seminars.

†CREST (INSEE). E-mail address: xavier.dhaultfoeuille@ensae.fr.

‡CREST (ENSAE). E-mail address: fevrier@ensae.fr.

1 Introduction

While the issue of endogeneity and the idea of instrumental variables are at least as old as econometrics, it is only recently that thorough investigations on the nonparametric identification of models with endogenous explanatory variables and instruments have been done (see, e.g., Matzkin, 2007, for a survey). Such studies are yet important because usual assumptions such as linearity or the separability of the error terms are seldom justified by theory, and are also likely to fail in practice. At least two approaches have been taken to generalize results on the linear model. The first one is based on estimating equations (see, e.g., Newey and Powell, 2003 or Chernozhukov and Hansen, 2005). This approach relies on relatively weak exogeneity conditions, and can handle both continuous and discrete endogenous variables. On the other hand, it imposes restrictions such as additive separability or rank similarity on the outcome disturbance. Besides, it relies on various completeness conditions for which very few sufficient conditions have been obtained (see, e.g., Newey and Powell, 2003, and D'Haultfoeuille, 2011) and which are restrictive (see Severini and Tripathi, 2006, and D'Haultfoeuille, 2011, for counterexamples). The second one uses a control variable approach (see, e.g., Newey et al., 1999, Florens et al., 2008 and Imbens and Newey, 2009). This strategy requires stronger exogeneity conditions on the instrument, and is more suited to continuous endogenous variables. On the other hand, it has the advantage of imposing less conditions on the outcome disturbances and relying on more transparent dependence conditions between the endogenous variables and the instrument.

A common feature of both approaches is that when the endogenous variable is continuous, the instrument should be continuous to achieve identification. In this paper, we alleviate this latter restriction. This is convenient because in many cases, we have at our disposal only discrete instruments. A typical example is policy reforms, which may affect the endogenous variable but not directly the outcome. Similarly, in experiments, the randomization in the treatment or control group may affect a continuous endogenous variable but not directly the outcome.[1] To do so, we combine both approaches and consider a triangular system where the structural and the reduced form equations are nonseparable, and potential outcomes are strictly monotonic in a scalar disturbance. We essentially impose a strong exogeneity assumption identical to the one considered by Imbens and Newey

[1] For instance, de Mel et al. (2008) use randomized grants (equal to \$0, \$100 or \$200) to generate exogenous shocks to capital stock and evaluate the returns to capital in microentreprises.

(2009), and a monotonicity and rank similarity conditions, as Chernozhukov and Hansen (2005). We show that the combination of these restrictions has a huge identification power. Even though the endogenous variable is continuous, full identification of the model can be achieved in the univariate case when the instrument is binary, and is typically fulfilled with an instrument taking three values or more. These results are strikingly at odds with those of Imbens and Newey (2009), which not only rely on a continuous instrument but also require large support conditions. Using a stochastic polynomial restriction, Florens et al. (2008) remove the need for large support conditions but also require the instrument to be continuous. Compared to Chernozhukov and Hansen (2005), we remove the need for a completeness condition and rely on discrete instruments instead of continuous ones[2] by just slightly reinforcing the exogeneity condition on the instrument.

To establish our results, we rely on general theorems in group and dynamical system theories, in the same spirit as our previous paper on the identification of adverse selection with instruments (see D'Haultfoeuille and Février, 2007, 2010).[3] The idea is the following. We consider, thanks to the control function approach, a change in the endogenous variable X due to a change in the instrument but not to a modification of the control variable. Such a change is exogenous and the associated shift in X from x to $x' = g(x)$ is identified. Observing its effect on the outcome, one can also relate, under the monotonicity and rank similarity conditions, the structural function φ at x with itself at $x' = g(x)$. Considering all the functions g associated with any change in the value of the instrument, we show that the problem of identifying φ is closely related to the properties of the group action of G, the group generated by these functions g.

Studying the property of this group action, we distinguish whether X is univariate or not, and whether the group action is free (which means that there does not exist any function in G different from the identity function which admits a fixed point) or not. In the univariate case with free actions, we prove that full identification can be achieved under very mild conditions if and only if the instrument takes at least three values. Conversely, with a binary instrument, we prove that the model is identified only on a countable set. With nonfree actions, a slight restriction of nonfreeness is sufficient to obtain full identification.[4] The

[2]The completeness condition imposes, apart from a particular kind of dependence, that the instrument is continuous when the endogenous variable is continuous.

[3]Interestingly, Kocecki (2010) also relies on group theory to reconsider identification in parametric models, and applies his framework to simultaneous equation models.

[4]This slight restriction is that the function in G different from the identity function which admits a fixed point should actually have a finite number of fixed points.

underlying idea is to use fixed points to recover the entire function φ. In a recent and closely related paper, Torgovitsky (2011) also uses fixed points to achieve full identification of the model. The intersection conditions used in this paper can be interpreted as a particular case of our nonfree group action case. Interestingly, and contrary to the intuition behind the result of Imbens and Angrist (1994) in the case of a dummy endogenous variable, namely that monotonic instruments are important to derive causal effects, heterogeneity in responses to a binary instrument actually helps identifying the model in our framework.

The multivariate case is much more complicated than the univariate one and it seems difficult to obtain a full classification. Yet, previous results can be partially extended. We prove that when the first stage equation is a generalized location model, which is a particular case of a free group action, the model is fully identified under mild conditions, provided that the instrument takes at least $d+2$ values, where d denotes the dimension of X. With nonfree group actions, fixed points can be used to fully identify the structural function under a supplementary restriction on the relative positions of g and the identity function.

The paper is organized as follows. Section 2 presents the model. Section 3 describes our identification strategy and its link with group theory. Our main identification results, in the univariate case, are presented in Section 4. Section 5 considers the multivariate case. Section 6 concludes. All proofs are deferred to the appendix.

2 The model

Let $X \in \mathbb{R}^d$ be the endogenous variable and Z denote the instrument. We denote by Y_x the potential outcome corresponding to the situations where $X = x$. For the sake of simplicity, we do not introduce exogenous covariates hereafter, but our analysis holds with such covariates by simply conditioning on them. We consider the following triangular nonseparable model:

$$\begin{aligned} Y_x &= \varphi(x, U_x) \\ X &= \psi(Z, V). \end{aligned}$$

We observe X, Z and $Y \equiv Y_X$. Such a model is also considered by, e.g., Chernozhukov and Hansen (2005), Florens et al. (2008), Imbens and Newey (2009) or Torgovitsky (2011). We aim at recovering the function φ from the distribution of (Y, X, Z). Contrary to Imbens and Newey (2009) and Florens et al. (2008), we suppose to have at our disposal only a discrete instrument variable $Z \in \{1, ..., K\}$, $K \geq 2$. Our first assumption is the exogeneity of the instrument. Subsequently, we denote, for any random variables S and T, F_S and $F_{S|T}$ the cumulative distribution functions of S and of S conditional on T, respectively. Similarly, we let $\text{Supp}(S)$ and $\text{Supp}(S|T)$ denote the support of S and of S conditional on T, respectively. Finally, we denote by $\mathcal{X}$ the interior of $\text{Supp}(X)$.

Assumption 1 *(Strong exogeneity)* $Z \perp\!\!\!\perp (V, (U_x)_{x\in\mathcal{X}})$.

Assumption 1 is at the basis of the control variable approach also followed by Newey et al. (1999), Florens et al. (2008) and Imbens and Newey (2009). The reason is that under this condition and additional restrictions (implied by Assumptions 3-4 below), X is independent of $U \equiv U_X$ conditional on $F_{X|Z}(X|Z)$ (see, e.g., Theorem 1 of Imbens and Newey, 2009). $F_{X|Z}(X|Z)$ is thus a control variable in the sense that conditioning on it removes the endogeneity of X. An alternative to Assumption 1 is to suppose only that $Z \perp\!\!\!\perp (U_x)_{x\in\mathcal{X}}$, following, e.g., Chernozhukov and Hansen (2005) or Chesher (2010). In this case however, it is not possible to find a control variable in general.

Our second and third assumptions, also imposed by Chernozhukov and Hansen (2005), are a rank similarity and a monotonicity conditions, which puts some restriction on the potential outcome disturbances $(U_x)_{x\in\mathcal{X}}$ and on functions φ and ψ. We let $\psi_1, ..., \psi_d$ denote the different components of ψ.

Assumption 2 *(Rank similarity) The distribution of U_x conditional on V does not depend on x.*

Assumption 3 *(Strict monotonicity in the disturbances) $(U_x, V) \in \mathbb{R}^2$ and for all $(x, z, m) \in \mathcal{X} \times \{1, ..., K\} \times \{1, ...d\}$, $u \mapsto \varphi(x, u)$ and $v \mapsto \psi_m(z, v)$ are strictly increasing.*

Assumption 2 is satisfied for instance under rank invariance (see, e.g., Doksum, 1974) i.e. when $U_x = U$ for all x. In the returns to education example, rank invariance means that individuals would always be at the same rank of potential wages, for all possible values of schooling. Such an assumption is restrictive, since it is likely that different shocks affect individuals in all counterfactual situations. In this case, rank invariance does not hold in general, while rank similarity may still be satisfied. For instance it holds if $U_x = q(U, \nu_x)$, where U is a common term which can be anticipated by the individual (and is thus correlated with V) and ν_x are identically distributed unanticipated shocks, which are thus independent of V. See also Chernozhukov and Hansen (2005) for a detailed discussion on this condition.

Supposing the error term to be real is standard and also assumed by, e.g., Chernozhukov and Hansen (2005) or Chesher (2007, 2010), but is not imposed for instance by Hoderlein and Mammen (2007) or Imbens and Newey (2009). The monotonicity condition on $\varphi(x, .)$ encompasses in particular the additive model $\varphi(x, u) = g(x)+u$ and transformation models $\varphi(x, u) = h(g(x) + u)$, where h is strictly increasing (see, e.g., Chiappori et al., 2010, for a recent analysis of this model).

Finally, we impose the following technical restrictions.

Assumption 4 *(Technical restrictions) (i) V is continuously distributed, (ii) $(u, v) \mapsto F_{U_x|V=v}(u)$ is continuous and strictly increasing in u for all $v \in Supp(V)$, (iii) $Supp(X|Z = z) = \prod_{m=1}^{d}[\underline{x}_m, \overline{x}_m]$ with $-\infty \leq \underline{x}_m < \overline{x}_m \leq \infty$ independent of z, (iv) $\varphi(., .)$ and $\psi(z, .)$ are continuous on $\{(x, u) \in \mathcal{X} \times Supp(U_x)\}$ and $Supp(V)$ respectively.*

The first two conditions basically formalize the idea that Y and X are continuous. The third allows the support of X to be either bounded or unbounded. It however imposes that this support does not depend on Z. Our analysis below could be adapted without this restriction, but at the price of an additional complexity. The fourth, finally, excludes discontinuous effects of X on Y. It is important as we often obtain identification of $\varphi(., u)$ on dense subsets of $\mathcal{X}$. By continuity, this ensures that φ is identified everywhere. We only impose continuity of $\varphi(., u)$ on the interior $\mathcal{X}$ of the support of X and not on the whole support. This is important when $\varphi(., u)$ tends to infinity on the boundaries.

We show below that actually, Assumptions 1 to 4 together have a strong identification power, since they allow to achieve identification of φ with a discrete instrument, whereas X is continuous. This result is at odds with those of Chernozhukov and Hansen (2005), Florens et al. (2008) and Imbens and Newey (2009). They all require the instrument to be continuous, together with other restrictions: completeness in the first, functional restrictions in the second and a “large support” condition in the third. Our conclusion is closer, on the other hand, with the one of Torgovitsky (2011) who considers assumptions very similar to ours. We discuss in more details the relationship between our results and those of Torgovitsky (2011) later on.

3 The identification strategy

3.1 A reformulation of the identification problem

As mentioned previously, we suppose that we observe Y, X and Z, and we seek to recover the function φ. It is well known that normalizations are possible in Model (2). First, denoting $V = (V_1, ..., V_d)$, we can replace, for any $m \in \{1, ..., d\}$, V_m by any $h_m(V_m)$ (h_m being a continuous strictly increasing function) and modify ψ_m accordingly without changing the model. We can thus always suppose that V_m is uniformly distributed. Under this normalization, for any observation $V_m = v_m$ and $Z = j$, $\psi_m(.,.)$ satisfies $\psi_m(j, v) = F^{-1}_{X_m|Z=j}(v_m)$, which is identified. $\psi_m(j, v_m)$ is the v_m-th quantile of X_m conditional on $Z = j$. Then $\psi(j, v)$, and similarly $V = F_{X|Z}(X)$, are identified. Hence, the identification of φ can be studied supposing that we observe not only Y, X and Z, but also V. In the rest of the paper, we denote by $\psi^{-1}(j, .)$ the inverse of $\psi(j, .)$. This inverses exists by Assumption 3, and is defined on Supp(X) by Assumption 4.

For similar reasons, a normalization on U_x is also possible. Rather than imposing the distribution of U_{x_0} to be uniform (for a given $x_0 \in \mathcal{X}$), it is more convenient to suppose, still without loss of generality, that the distribution of U_{x_0} conditional on $X = x_0$ is uniform. In this case, $\varphi(x_0, .)$ is identified by $\varphi(x_0, u) = F^{-1}_{Y|X=x_0}(u)$.

The idea of the control function approach is that we can yield an exogenous change in X by moving Z while keeping V constant (since under the exogeneity condition, $U_x \perp\!\!\!\perp Z|V$). Observing the effect of such a change on Y, one can relate $\varphi(x, u)$ and $\varphi(x', u)$ for an appropriate value of x', using the monotonicity and rank similarity conditions. Formally, let us define $g_{ij}(.)$ for all $(i, j) \in \{1, ..., K\}^2$, by

$$g_{ij}(x) = \psi(j, \psi^{-1}(i, x)),$$

where $\circ$ denote the composition operator ($f \circ g(x) = f(g(x))$). $g_{ij}(x) - x$ is the shift in X when Z moves from i to j while V remains constant. Because ψ and ψ^{-1} are identified,

g_{ij} also is. Moreover, for all $x \in \mathcal{X}$,

$$\begin{aligned}
F_{Y|Z=i,V=\psi^{-1}(i,x)}(\varphi(x,u)) &= P(Y_x \leq \varphi(x,u)|Z=i, V=\psi^{-1}(i,x)) \\
&= P(U_x \leq u|Z=i, V=\psi^{-1}(i,x)) \\
&= P(U_x \leq u|V=\psi^{-1}(i,x)) \\
&= P(U_{g_{ij}(x)} \leq u|V=\psi^{-1}(i,x)) \\
&= P(U_{g_{ij}(x)} \leq u|Z=j, V=\psi^{-1}(i,x)) \\
&= P(Y_{g_{ij}(x)} \leq \varphi(g_{ij}(x),u)|Z=j, V=\psi^{-1}(i,x)) \\
&= F_{Y|Z=j,V=\psi^{-1}(i,x)}(\varphi(g_{ij}(x),u)).
\end{aligned}$$

The second equality is satisfied because $\varphi(x,.)$ is strictly increasing (Assumption 3). The third stems from the independence between Z and (U_x, V) (Assumption 1), which implies that U_x is independent of Z conditional on V. This step is crucial in our analysis and can be interpreted as a reduction of the dimensionality of the problem. Instead of working in a space of dimension 2 (Z and V), we can work in a space of dimension 1 (V only). The fourth equality, which stems from the rank similarity condition (Assumption 2), is also important, as it allows to relate the distribution of U_x with the one of $U_{g_{ij}(x)}$. As a result, we can also relate the distributions of the potential outcomes Y_x and $Y_{g_{ij}(x)}$. The last equalities apply the same reasoning as the first ones, but in a reverse way.

By Assumptions 3 and 4, $y \mapsto F_{Y|Z=i,V=v}(y)$ is strictly increasing for all v.[5] Hence, its inverse exists, and

$$\varphi(g_{ij}(x),u) = F^{-1}_{Y|Z=j,V=\psi^{-1}(i,x)} \circ F_{Y|Z=i,V=\psi^{-1}(i,x)}(\varphi(x,u)). \tag{3.1}$$

This equation shows that if $\varphi(x,.)$ is identified, for a given $x \in \mathcal{X}$, then $\varphi(g_{ij}(x),.)$ is also identified. We state this in the following lemma.

Lemma 3.1 *Suppose that $\varphi(x,.)$ is identified, for a given $x \in Supp(X)$. Then, for any $(i,j) \in \{1,...,K\}^2$, $\varphi(g_{ij}(x),.)$ is also identified.*

3.2 The link with group theory

Before showing the relationship between identification issues and group theory, let us recall some definitions and results on groups. A group G is a set endowed with a binary operator

[5] As before, $F_{Y|Z=i,V=v}(y) = F_{U_x|V=v}(\varphi^{-1}(x,y))$, where $\varphi^{-1}(x,.)$ denotes the inverse of $\varphi(x,.)$. The result follows since both $F_{U_x|V=v}(.)$ and $\varphi^{-1}(x,.)$ are strictly increasing.

$*$ which satisfies three properties. The first is associativity: for all $(a,b,c) \in G^3$, $(a*b)*c = a*(b*c)$. The second is the existence of an identity element $e \in G$ satisfying $a*e = e*a = a$ for all $a \in G$. The third is the existence of inverses. Every element $a \in G$ admits an element called its inverse and denoted a^{-1} which satisfies $a * a^{-1} = a^{-1} * a = e$. G is abelian if $*$ is commutative, i.e. if for all a, b, $a * b = b * a$. A subgroup H of G is a subset of G which is itself a group for $*$. If we let $(H_i)_{i\in\mathcal{I}}$ denote a family of subgroups of G, one can check that $\cap_{i\in\mathcal{I}} H_i$ is also a group. The group generated by a subset I of G is the intersection of all subgroups of G containing I. By definition, it is the smallest subgroup of G including I.

We also need to introduce the notion of group actions and orbits. For any set $\mathcal{A}$ and a group G, a group action . is a function from $G \times \mathcal{A}$ to $\mathcal{A}$ (denoted by $g.x$) satisfying, for every $(g,h) \in G^2$ and $x \in \mathcal{A}$, $(g*h).x = g.(h.x)$ and $e.x = x$. The orbit $\mathcal{O}_x$ of $x \in \mathcal{A}$ is then defined by

$$\mathcal{O}_x = \{g.x, g \in G\}.$$

A property which will prove useful is freeness. We say that the group action . is free if, for any $x \in \mathcal{A}$, $g.x = x$ implies $g = e$.

Finally, we recall definitions on equivalence relations. A relation $\mathcal{R}$ on a set $\mathcal{A}$ is an equivalence relation if it+ is reflexive ($x\mathcal{R}x$), symmetric ($x\mathcal{R}y$ implies $y\mathcal{R}x$) and transitive ($x\mathcal{R}y$ and $y\mathcal{R}z$ implies $x\mathcal{R}z$). The class of equivalence of x is then defined as the set $\{a \in \mathcal{A} : x\mathcal{R}a\}$.

Going back to our framework, we denote by G the group generated by the set of functions $(g_{ij})_{(i,j)\in\{1,\dots,K\}^2}$ and the composition operator $\circ$. This group includes for instance g_{12}, g_{12}^{-1}, $g_{12} \circ g_{12}$ (denoted g_{12}^2)[6] but also for instance, if $K \geq 3$, $g_{12}^2 \circ g_{31} \circ g_{12}$. We can properly define G because g_{ij} are bijections from $\mathcal{X}$ to itself.[7] Indeed, for all m, both $\psi_m(j,.)$ and $\psi_m^{-1}(i,.)$ are strictly increasing, and $\psi_m^{-1}(i,.)$ is onto $(0,1)$ while $\psi_m(j,.)$ is onto the interior of the support of X_m. Hence, we can safely define the inverse function of any g_{ij} and take their compositions. The function . from $G \times \mathcal{X}$ to $\mathcal{X}$ satisfying $g.x = g(x)$ is a group action, since $(g \circ h)(x) = g(h(x))$ and $\mathrm{Id}(x) = x$, where Id denotes the identity function. Hence, G acts on $\mathcal{X}$ through the composition operator. The orbit of $x \in \mathcal{X}$ is, by definition, $\mathcal{O}_x = \{g(x), g \in G\}$. Our first result, based on Lemma 3.1, is that $\varphi(x_0, u)$ is identified on $\overline{\mathcal{O}}_{x_0}$, the closure of $\mathcal{O}_{x_0}$.

Lemma 3.2 *Under Assumptions 1-4, $\varphi(.,.)$ is identified on $\overline{\mathcal{O}}_{x_0} \times (0,1)$.*

[6] More generally, we let $g^k(x)$ denote $g \circ \dots \circ g(x)$, if $k \geq 1$, x if $k = 0$ and $g^{-1} \circ \dots \circ g^{-1}(x)$, if $k \leq -1$.

[7] For technical reasons detailed below, we restrict elements of G to be defined on $\mathcal{X}$ instead of $\mathrm{Supp}(X)$.

Proof: fix $u \in (0,1)$ and $g \in G$. There exists $(i_1, j_1, ..., i_p, j_p) \in \{1, ..., K\}^2$ such that $g = g_{i_1 j_1} \circ ... \circ g_{i_p j_p}$. Because $\varphi(x_0, u) = F^{-1}_{Y|X=x_0}(u)$ is identified, we can identify, by Lemma 3.1, $\varphi(g_{i_p j_p}(x_0), u)$ and, by a straightforward induction, $\varphi(g(x_0), u)$. Thus, $\varphi(., .)$ is identified on $\mathcal{O}_{x_0} \times (0,1)$. By continuity of $\varphi(., u)$, $\varphi(., .)$ is identified on $\overline{\mathcal{O}}_{x_0} \times (0,1)$ $\square$

Let us define the relation $\mathcal{R}$ by $x\mathcal{R}x'$ if there exists $x_1 = x, ..., x_n = x'$ such that for all $i = 1, ..., n-1$,

$$\overline{\mathcal{O}}_{x_i} \cap \overline{\mathcal{O}}_{x_{i+1}} \cap \mathcal{X} \neq \emptyset.$$

It is easy to see that $\mathcal{R}$ is an equivalence relation. We denote by $\mathcal{X}_0$ the equivalence class of x_0. $\mathcal{X}_0$ includes $\overline{\mathcal{O}}_{x_0}$, but is, in general, larger. Theorem 3.1 below extends Lemma 3.1 to this set.

Theorem 3.1 *Under Assumptions 1-4, $\varphi(., .)$ is identified on $\mathcal{X}_0 \times (0,1)$.*

This theorem establishes that as soon as one orbit is “related” to the one of x_0, in the sense that the intersection of their closure is nonempty, we can identify $\varphi(., u)$ on this closure. Then, by induction, we can identify $\varphi(., u)$ on all orbits indirectly “related” to the one of x_0. To establish this result, we prove the nontrivial fact that if $x' \in \partial\mathcal{O}_x$ (the frontier of $\mathcal{O}_x$) and $\varphi(x', u)$ is identified, then we can actually identify $\varphi(., u)$ on $\overline{\mathcal{O}}_x$. To better understand this point, we will discuss it in details in the univariate setting.

An important consequence of Theorem 3.1 is that identification of $\varphi(x, .)$ can be rephrased in terms of determining the set $\mathcal{X}_0$. In turn, the nature of this set depends on the dimension of X and on how G acts on $\mathcal{X}$. In the following, we separate the univariate case $\mathcal{X} \subset \mathbb{R}$ from the multivariate one, and free actions from nonfree ones.

4 Results in the univariate case

We first consider the univariate case where $\mathcal{X} = (\underline{x}, \overline{x})$, for which the topology of $\mathcal{X}_0$ is more simple. It turns out that under mild restrictions, $\mathcal{X}_0$ is either equal to $\mathcal{X}$ or discrete, so that φ is either fully identified or identified on a sequence of points.

4.1 Free actions

We first suppose that the action of G on $\mathcal{X}$ is free: apart from the identity function, no $g \in G$ admits any fixed point.

Assumption 5 *(Free action) G acts freely on $\mathcal{X}$: if $g(x) = x$ for any $(g, x) \in G \times \mathcal{X}$, then g is the identity function.*

When the instrument is binary, G acts freely if and only if g_{12} admits no fixed point or is equal to the identity. The latter case corresponds to an instrument independent of X. Otherwise, we have either $\psi(1, .) > \psi(2, .)$ or $\psi(2, .) > \psi(1, .)$. Hence, this assumption may be seen as an extension, in a continuous setting, to the monotonicity condition considered by Imbens and Angrist (1994) in the case of a dummy endogenous variable. The important difference with their monotonicity condition, however, is that we can test it directly in the data, by checking that $F_{X|Z=1}$ stochastically dominates (or is dominated by) $F_{X|Z=2}$ at the first-order.

In this case of a binary instrument, the orbit of x_0 is discrete, and consists of the monotonic sequence $(x_k)_{k\in\mathbb{Z}}$ defined by $x_k = g_{12}^k(x_0)$. Figure 1 depicts the correspondence between this sequence and the identification of $\varphi(., u)$. The left graph shows how to build $(x_k)_{k\in\mathbb{Z}}$ by applying $g_{12}(.)$ successively. The right graph depicts in dash the true curve $x \mapsto \varphi(x, u)$ and the black points corresponding to the sequence where $\varphi(., u)$ is identified. By Lemma 3.1, we can indeed identify $\varphi(x_k, u)$ as soon as $\varphi(x_{k-1}, u)$ has been recovered. One may expect that $\varphi(., u)$ is identified only on this sequence. Theorem 4.1 formalizes this result, by proving that $\mathcal{X}_0 = \mathcal{O}_{x_0}$ and that no other point can be identified outside $\mathcal{X}_0$.[8]

Theorem 4.1 *If $K = 2$ and Assumptions 1-5 hold, $\varphi(x, u)$ is identified for any $u \in (0, 1)$ if and only if $x \in \{x_k : k \in \mathbb{Z}\}$.*

[8]When g_{12} is the identity function, we identify $\varphi(., u)$ only at x_0, which makes sense since this case corresponds to an instrument independent of X.

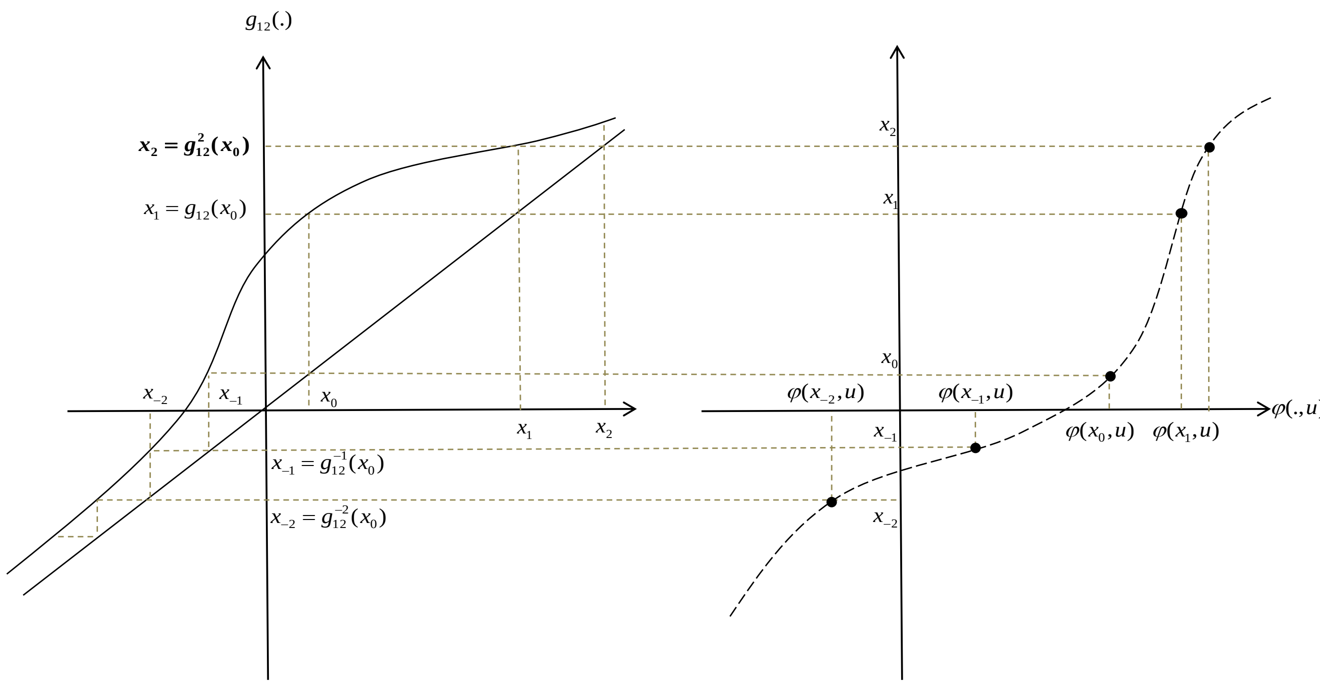

Figure 1: $\mathcal{X}_0$ in the free case, $K = 2$.

Even if we obtain identification only on a sequence of points, we can still identify several interesting parameters.[9] Consider the average treatment effect $\Delta_{jkl}^{ATE} = E\left(Y_{x_k} - Y_{x_j}|X = x_l\right)$ and the quantile treatment effect $\Delta_{jkl}^{QTE}(\tau) = F^{-1}_{Y_{x_k}|X=x_l}(\tau) - F^{-1}_{Y_{x_j}|X=x_l}(\tau)$. Δ_{jkl}^{ATE} is the average effect of moving X from x_j to x_k for people with $X = x_l$. $\Delta_{jkl}^{QTE}(\tau)$ is the effect of moving from x_j to x_k for people who remain on the τ-th quantile of potential outcomes and with $X = x_l$. Corollary 4.2 shows that both parameters can be identified.

Corollary 4.2 *If $K = 2$ and Assumptions 1-4 and 5 hold, Δ_{jkl}^{ATE} and $\Delta_{jkl}^{QTE}(\tau)$ are identified, for any $(j, k, l, \tau) \in \mathbb{Z}^3 \times (0, 1)$.*

A consequence of Corollary 4.2 is that we can identify in particular Δ_{jkk}^{ATE}, which corresponds to the average treatment (of going from x_j to x_k) for the treated. On the other hand, without further assumption, we cannot identify average treatment effect on the whole population or average marginal effects such as $E\left(\partial\varphi/\partial x(x, U_x)|X = x\right)$.

When $K \geq 3$, the situation is very different from the binary case, because the orbit of x_0 does not reduce to a monotonous sequence. To see this, consider Figure 2. As previously

[9] For $x \notin \mathcal{X}_0$, bounds on $\varphi(x, u)$ can also be achieved under additional restrictions on $\varphi(., u)$ such as monotonicity.

$x_1 = g_{12}(x_0) > x_0$ lies in $\mathcal{O}_{x_0}$. Suppose in this example that we apply g_{13}^{-1} to x_1, and then g_{12}. We obtain a point $x_2 \in \mathcal{O}_{x_0}$ that lies between x_0 and x_1. Contrary to the binary case, we are thus able to get some information on (x_0, x_1). Many points can be reached this way. Actually, we show below that under Assumption 2 and a mild restriction, $\mathcal{O}_{x_0}$ is dense in Supp(X). As a result, x_0 has only one equivalence class and φ is fully identified by Theorem 3.1.

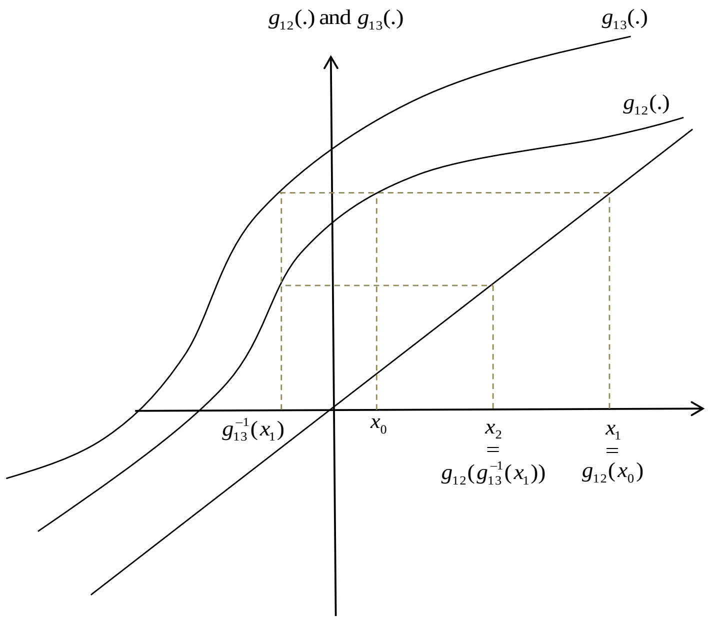

Figure 2: Some points in the orbit of x_0 in the free case, $K \geq 3$.

To get some intuition on this result, consider the generalized location model defined by

$$X = \lambda\left(\sum_{k=2}^{K} a_k \mathbb{1}\{Z = k\} + q(V)\right), \tag{4.1}$$

for continuous and strictly increasing functions q and λ (from $(0,1)$ to $\mathbb{R}$ and from $\mathcal{X}$ to itself, respectively). In this case, $g_{ij}(x) = \lambda(-a_i + a_j + \lambda^{-1}(x))$ (with $a_1 = 0$), so that any $g \in G$ satisfies

$$g(x) = \lambda\left(\sum_{k=2}^{K} c_k a_k + \lambda^{-1}(x)\right), \tag{4.2}$$

with $(c_2, ...c_K) \in \mathbb{Z}^{K-1}$. Hence, we have either $g(x) > x$, $g(x) < x$ or $g(x) = x$ for all $x \in$ Supp(X), depending on the sign of $\sum_{k=2}^{K} c_k a_k$. In other words, Assumption 5 is satisfied: the group action is free. Besides, the orbit of x_0 satisfies

$$\mathcal{O}_{x_0} = \left\{ \lambda\left(\sum_{k=2}^{K} c_k a_k + \lambda^{-1}(x_0)\right), (c_2, ..., c_K) \in \mathbb{Z}^{K-1} \right\}.$$

The set $\sum_{k=2}^{K} a_k\mathbb{Z}$ is an additive subgroup of $\mathbb{R}$. By a classical result on these additive subgroups (see, e.g., Stillwell, 1992, p.33), it is either discrete or dense. Density is achieved if and only if there exists $(i, j) \in \{1, ..., K\}^2$ such that $a_i/a_j \notin \mathbb{Q}$, a very mild condition since the Lebesgue measure of $\mathbb{Q}$ is zero. As a result, $\mathcal{O}_{x_0}$ is also either discrete or dense on Supp(X), depending on whether the ratios a_i/a_j are all rational or not.

Qualitatively, the same phenomenon arises in the general case where Assumption 5 holds but without necessarily a generalized location model. Provided that the functions g_{ij} are regular enough, The orbit $\mathcal{O}_{x_0}$ is either discrete or dense, the discrete case being the exception. Assumption 6 rules out this particular case.

Assumption 6 *(Regularity and non periodicity) There exists $(i, j, k) \in \{1, ..., K\}^3$ such that $\psi(i, .), \psi(j, .)$ and $\psi(k, .)$ are C^2 diffeomorphisms and for all $(m, n) \in \mathbb{Z}^2$, $(m, n) \neq (0, 0)$, $g_{ij}^m \neq g_{ik}^n$.*

Theorem 4.3 *If $K \geq 3$ and Assumptions 1-4 and 5-6 hold, $\mathcal{X}_0 = \mathcal{X}$, and φ is fully identified.*

The proof relies on Hölder and Denjoy theorems, two deep results in group and dynamical systems theories. Denjoy theorem, in particular, ensures that $\mathcal{O}_{x_0}$ is either discrete or dense in Supp(X), depending on whether a scalar called the *rotation number* (which corresponds to the ratio a_2/a_3 in the example above) is rational or not. Assumption 6 ensures that this number is irrational, establishing the density of $\mathcal{O}_{x_0}$. Interestingly, the proof shows that only three different values of Z are needed to achieve point identification of $\varphi(., .)$. If $K \geq 4$ and Assumption 6 holds for four indices or more, we can use different subsets of this set of indices to recover $\varphi(., .)$. If the different corresponding functions do not coincide, the model is rejected. In other words, the model is overidentified in general when $K \geq 4$.

Theorem 4.3 presents the identification result under what we believe is the most common case on ψ (namely, when Assumption 6 holds). If, on the contrary, for any $(i, j, k) \in \{1, ..., K\}^2$ there exists (m, n) such that $g_{ij}^m = g_{ik}^n$, one can show that $\varphi(., u)$ is identified only on a sequence of points as in the binary case. Moreover, when the $\psi(j, .)$ functions are not C^2, we cannot apply Denjoy theorem. The orbits may neither be discrete nor dense in this case, but related to Cantor sets (see, e.g., Ghys, 2001, Proposition 5.6).

One may wonder how to test for Assumption 5 in practice. Actually, by Hölder theorem, the group G is abelian in this case. Hence, if $K = 3$ for instance, we can test for the much simpler condition $g_{12} \circ g_{13} = g_{13} \circ g_{12}$.

4.2 Nonfree actions

We now turn to the situation where the action of G on $\mathcal{X}$ is not free. Recall that this means that there exists $x \in \mathcal{X}$ and $g \in G$ different from the identity function such that $g(x) = x$. Actually, we slightly reinforce this condition here by limiting the number of fixed points to be finite.[10]

Assumption 7 *(Non freeness) There exists $g \in G$ different from the identity function which admits a positive and finite number of fixed points.*

To provide intuitions on identification in the nonfreeness case, let us consider a function g that satisfies this condition. Suppose for simplicity that it has only one fixed point x_f, with $g(x) > x$ for $x < x_f$ and $g(x) < x$ otherwise (see Figure 3). Then for any value of x_0, $x_f = \lim_{k\to\infty} g^k(x_0)$, so that $x_f \in \overline{\mathcal{O}}_{x_0}$. We can identify $\varphi(., u)$ at $g^k(x_0)$ and also, by continuity, at x_f (see Figure 3). To go further, we rely on a property we have not exploited so far: by Theorem 3.1, we can identify $\varphi(x, u)$ for all x such that $x_f \in \overline{\mathcal{O}}_x$, because in this case $x_0 \mathcal{R} x$. As any $x \neq x_f$ satisfies this property, the model is fully identified.

To understand this last step, consider a point x which is not in $\overline{\mathcal{O}}_{x_0}$, and choose a value a for $\varphi(x, u)$ (the upper cross in Figure 3). By Equation (3.1), $\varphi(g^k(x), u)$ should then be equal to a function of a, say $h_k(a)$ (other crosses in Figure 3). Because the sequence $(g^k(x))_{k\in\mathbb{N}}$ converges to x_f, the sequence $(h_k(a))_{k\in\mathbb{N}}$ also converges (let $h_\infty(a)$ denote its limit). $h_\infty(a)$ should be equal to $\varphi(x_f, u)$, which is identified. This imposes constraints on the choice of a. The proof of Theorem 3.1 implies actually that there is a unique a (namely, the true value $\varphi(x, u)$) such that $h_\infty(a) = \varphi(x_f, u)$. Figure 3 illustrates this: for a given $a \neq \varphi(x, u)$, we obtain $h_\infty(a) \neq \varphi(x_f, u)$. This implies that $\varphi(x, u)$ is identified.

Hence, when g admits a unique fixed point, $\mathcal{X}_0 = \mathcal{X}$ and the model is fully identified. Theorem 4.4 extends this result to functions with a finite number of fixed points.

[10] Hence, our classification in the univariate case rules out pathological situation where all elements of G cross an infinite number of times the identity function, without being the identity functions.

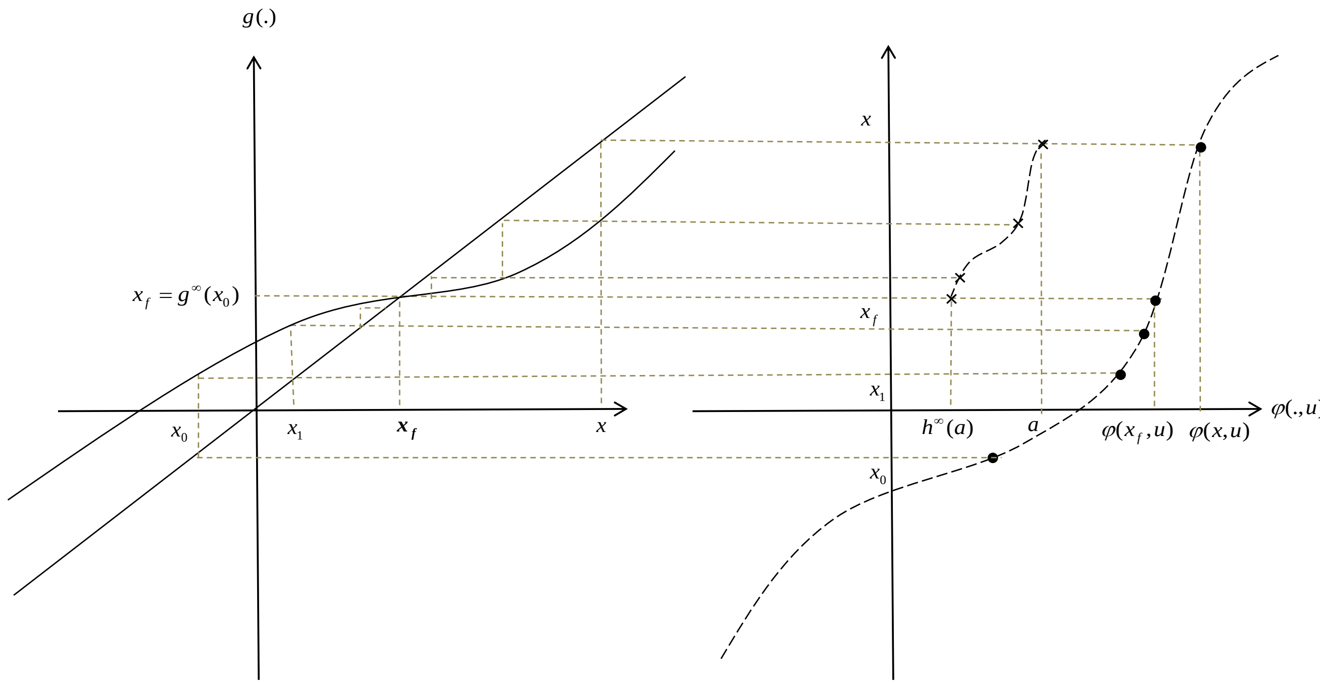

Figure 3: Illustration of $\mathcal{X}_0 = \mathcal{X}$ under Assumption 7.

Theorem 4.4 *If Assumptions 1-4 and 7 hold, $\mathcal{X}_0 = \mathcal{X}$. As a result, φ is fully identified.*

To better understand the nonfreeness property, suppose that $F_{X|Z=i}$ and $F_{X|Z=j}$ cross at least once and at most a finite number of times, for some $i \neq j \in \{1, ..., K\}^2$. Such a case is likely to hold when the effect of Z on X is heterogenous. Reconsider the example of returns to education, and suppose that individuals face two tuition fee policies. In the first case, tuition fees remain constant through years while in the second, they are larger than the first for first years of college, and then lower. If the assignment of the policy is random, we can use the tuition fee policy as an instrument (Z denoting the label of the policy). It is likely that $F_{X|Z=1}$ and $F_{X|Z=2}$ cross, because by the policies design, $F_{X|Z=1}(x) < F_{X|Z=2}(x)$ for small x, and $F_{X|Z=1}(x) > F_{X|Z=2}(x)$ for large x. In such a case, Assumption 7 holds, because g_{ij} will have as many fixed points as crossing points between $F_{X|Z=i}$ and $F_{X|Z=j}$.

Corollary 4.5 *Suppose that $F_{X|Z=i}$ and $F_{X|Z=j}$ cross at least once and at most a finite number of times on $\mathcal{X}$. Then φ is fully identified.*

This corollary implies in particular that the model is fully identified with a binary instrument when $F_{X|Z=1}$ and $F_{X|Z=2}$ cross, i.e. when the monotonicity condition discussed above fails to hold. The intuition conveyed by the result of Imbens and Angrist (1994) that monotonicity helps for identification is actually reversed here: it is better to have a nonmonotonic instrument.

The "crossing case" on $F_{X|Z}$ considered in Corollary 4.5 is studied by Torgovitsky (2011), who also shows, in a closely related paper, that the model is fully identified thanks to crossing points.[11] It is also related to the main result of Guerre et al. (2009), who shows identification of an auction model with (discrete) variations in the number of players, and its generalization by D'Haultfœuille and Février (2007) to other adverse selection models. All these papers rely on the fact that two curves cross to identify the model of interest.

In the case of a binary instrument, Assumption 7 is equivalent to the "crossing" condition stated above, and there is no need to introduce the group G to understand identification. However, when the instrument takes three values or more, reasoning on G is important because Assumption 7 is much weaker than the simple "crossing" condition. It may hold even if none of the functions $F_{X|Z=i}$ and $F_{X|Z=j}$ cross, for all $(i,j) \in \{1, ..., K\}^2$, so that none of the g_{ij}, $(i,j) \in \{1, ..., K\}^2$, admits a fixed point. To illustrate this claim, consider for instance that shifting Z from 1 to 2 has a pure location effect, so that $F_{X|Z=2}(x) = F_{X|Z=1}(x-m)$ for a given $m > 0$, and suppose that $F_{X|Z=3}(x) = F_{X|Z=1}(x-m-\exp(-x))$. The functions $F_{X|Z=1}$, $F_{X|Z=2}$ and $F_{X|Z=3}$ do not cross each other, because $F_{X|Z=1}(x) > F_{X|Z=2}(x) > F_{X|Z=3}(x)$ for all x. Now, $g_{21}(x) = x - m$, so that $g_{21}^2(x) = x - 2m$. Similarly,

$$g_{31}(x) = F^{-1}_{X|Z=1}(F_{X|Z=3}(x)) = x - m - \exp(-x).$$

Because the equation $x - 2m = x - m - \exp(-x)$ admits a unique solution in x (namely $-\ln(m)$), $g_{21}^2 \circ g_{13}$ admits a unique fixed point, and Assumption 7 is satisfied.

Up to now, we have supposed that the number of fixed points was finite. An interesting case where it fails to hold is when the instrument has a "local" effect on the endogenous variable. Consider for instance the case where $K = 2$ and suppose that $F_{X|Z=1}(x) = F_{X|Z=2}(x)$ for all $x \geq x_0$ and $F_{X|Z=1}(x) > F_{X|Z=2}(x)$ for $x < x_0$. Such a case is likely to hold when the instrument targets a very specific population. Reconsidering the returns to schooling example, for instance, it is plausible that a reform which raises the maximum compulsory age of school only has only an effect at the bottom of the distribution of years of schooling. In this case, $g_{12}(x) = x$ for $x \geq x_0$, so that Assumption 7 fails to hold. Still, $\mathcal{X}_0 = (\underline{x}, x_0]$,

[11] Torgovitsky (2011) strengthens Assumption 4 by supposing that the support of X is bounded in at least one direction and by imposing that $\varphi(., u)$ is continuous on the whole support of X. This allows him to use, at the limit, a crossing point at one boundary of the support.

so that we can point identify $\varphi(., u)$ on this whole interval.[12]

5 The multivariate case

In the multivariate case, the topology of $\mathcal{X}_0$ may be more complicated than before. As a result, it seems difficult to yield a full classification, contrary to the univariate case. Yet, Theorem 3.1 is still valid and previous ideas can be partially extended.

5.1 The free case

Let us suppose first that Assumption 5 holds. When $K = 2$, we get a similar result as in Theorem 4.1. $\mathcal{X}_0 = \{g_{12}^k(x_0), k \in \mathbb{Z}\}$ and we identify $\varphi(., u)$ on this sequence only. The case $K \geq 3$ is far more delicate. In particular, we cannot apply Hölder and Denjoy's theorems to yield a similar result as Theorem 4.3, because these theorems only apply to real, univariate functions. Still the generalized location model provides some interesting insights. Suppose, as in the univariate case, that

$$X = \lambda\left(\sum_{k=2}^{K} a_k \mathbb{1}\{Z = k\} + q(V)\right), \tag{5.1}$$

with $a_k = (a_{k1}, ...a_{kd}) \in \mathbb{R}^d$, $q = (q_1, ..., q_d)$, $\lambda = (\lambda_1, ..., \lambda_d)$ and where q_m and λ_m ($m = 1...d$) are strictly increasing continuous functions. As before, $g_{ij}(x) = \lambda(a_i - a_j + \lambda^{-1}(x))$. As a result,

$$\mathcal{O}_{x_0} = \left\{\lambda\left(\sum_{k=2}^{K} c_k a_k + \lambda^{-1}(x_0)\right), (c_2, ..., c_K) \in \mathbb{Z}^{K-1}\right\}.$$

Using a characterization of additive subgroups of $\mathbb{R}^d$, we obtain a necessary and sufficient condition for $\mathcal{X}_0 = \mathcal{X}$ to hold. Introducing $H = \sum_{k=2}^{K} \mathbb{Z}a_k$, we denote, for any $x \in \mathbb{R}^d$, $< x, H >= \sum_{k=2}^{K} < x, a_k > \mathbb{Z}$, where $< ., . >$ is the usual scalar product in $\mathbb{R}^d$.

Theorem 5.1 *Under Assumptions 1-4, if Equation (5.1) holds, $\mathcal{X}_0 = \mathcal{X}$ if and only if*

$$\{x \in \mathbb{R}^d :< x, H >\subset \mathbb{Z}\} = \{0\}. \tag{5.2}$$

A necessary condition for $\mathcal{X}_0 = \mathcal{X}$ is that $K \geq d + 2$.

[12] We could also prove, in the same spirit as in Theorem 4.1, that $\varphi(x, u)$ is not identified on $(x_0, \overline{x})$. Such an example is also considered by Torgovitsky (2011).

Theorem 5.1 provides interesting insights on how K should grow with d to yield $\mathcal{X}_0 = \mathcal{X}$. A necessary condition is $K \geq d+2$. Actually, it turns out that this condition is sufficient under mild additional restrictions. To see this, reconsider the univariate case $d = 1$ and suppose that $K = 3$. Then $< x, H > \subset \mathbb{Z}$ implies that $xa_2 \in \mathbb{Z}$ and $xa_3 \in \mathbb{Z}$. Hence, as discussed before, $\mathcal{X}_0 = \mathcal{X}$ as soon as $a_2/a_3 \notin \mathbb{Q}$. Similar (but more tedious) restrictions arise when $d > 1$. Another implication of Theorem 5.1 is that the usual intuition that we need one instrument per endogenous variable is misleading here. When $d = 4$ for instance, $K = 6$ is sufficient in general. One binary and one ternary instrument are thus sufficient to get full identification.

5.2 The nonfree case

Let us turn to the free case where Assumption 7 holds. We can still use fixed points to characterize $\mathcal{X}_0$, however another element comes into play, namely the attractiveness of these fixed points. To understand this, consider the bivariate case with $K = 2$, and let $x_f = (x_{1,f}, x_{2,f})$ denote a fixed point of $g_{12} = (g_{1,12}, g_{2,12})$.

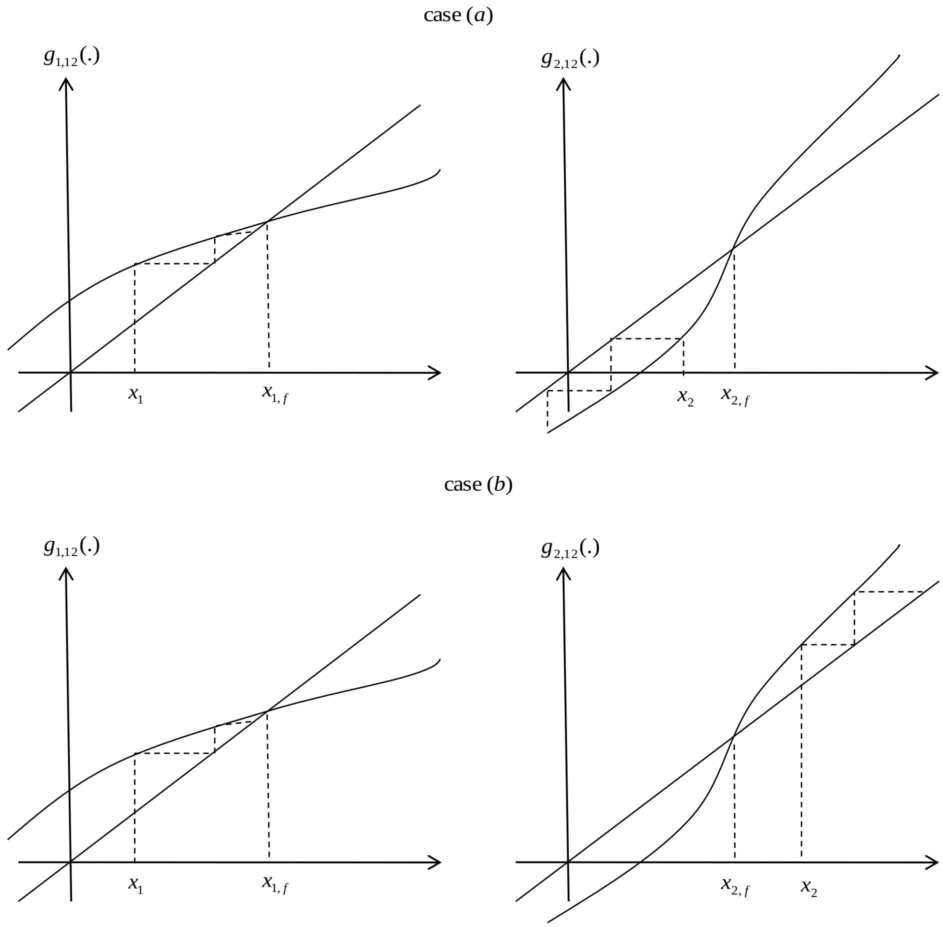

Figure 4: A case where $\mathcal{X}_0 \neq \mathcal{X}$ when $K = d = 2$ under Assumption 7.

Consider first a situation where $g_{1,12}(x_1) > x_1$ if and only if $x_1 < x_{1,f}$, while $g_{2,12}(x_2) < x_2$ if and only if $x_2 < x_{2,f}$ (see Figure 4). Then no sequence $(g_{12}^k(x))_{k\in\mathbb{N}}$ converges in $\mathcal{X}$. When $x = (x_1, x_2) \in (-\infty, x_{1,f}) \times (-\infty, x_{2,f})$, the sequence $(g_{1,12}^k(x_1))_{k\in\mathbb{N}}$ converges to $x_{1,f}$ but the sequence $(g_{2,12}^k(x_2))_{k\in\mathbb{N}}$ tends to $-\infty$, with $(x_{1,f}, -\infty) \notin \mathcal{X}$ (case (a)). Similarly, when $x = (x_1, x_2) \in (-\infty, x_{1,f}) \times (x_{2,f}, +\infty)$, the sequence $(g_{1,12}^k(x_1))_{k\in\mathbb{N}}$ converges to $x_{1,f}$ but the sequence $(g_{2,12}^k(x_2))_{k\in\mathbb{N}}$ tends to $+\infty$ (case (b)). In this case, one would like to apply simultaneously $g_{1,12}$ and $g_{2,12}^{-1}$, but $(g_{1,12}, g_{2,12}^{-1}) \notin G$. As a result, the fixed point does not allow us to relate different points together, and $\mathcal{X}_0$ corresponds to a sequence of points, as in the free case studied before. The model is partially identified.

On the other hand, suppose that $g_{m,12}(x_m) < x_m$ if and only if $x_m < x_{m,f}$, for $m \in \{1, 2\}$. Then, for any $x = (x_1, x_2)$, the sequence $(g_{12}^{-k}(x))_{k\in\mathbb{N}}$ converges to x_f. In Figure 5, we represent such a sequence for $x \in (-\infty, x_{1,f}) \times (-\infty, x_{2,f})$ (case (a)) and for $x \in (-\infty, x_{1,f}) \times (x_{2,f}, +\infty)$ (case (b)). As in the univariate case, $\mathcal{X}_0 = \mathcal{X}$ and the model is fully identified.

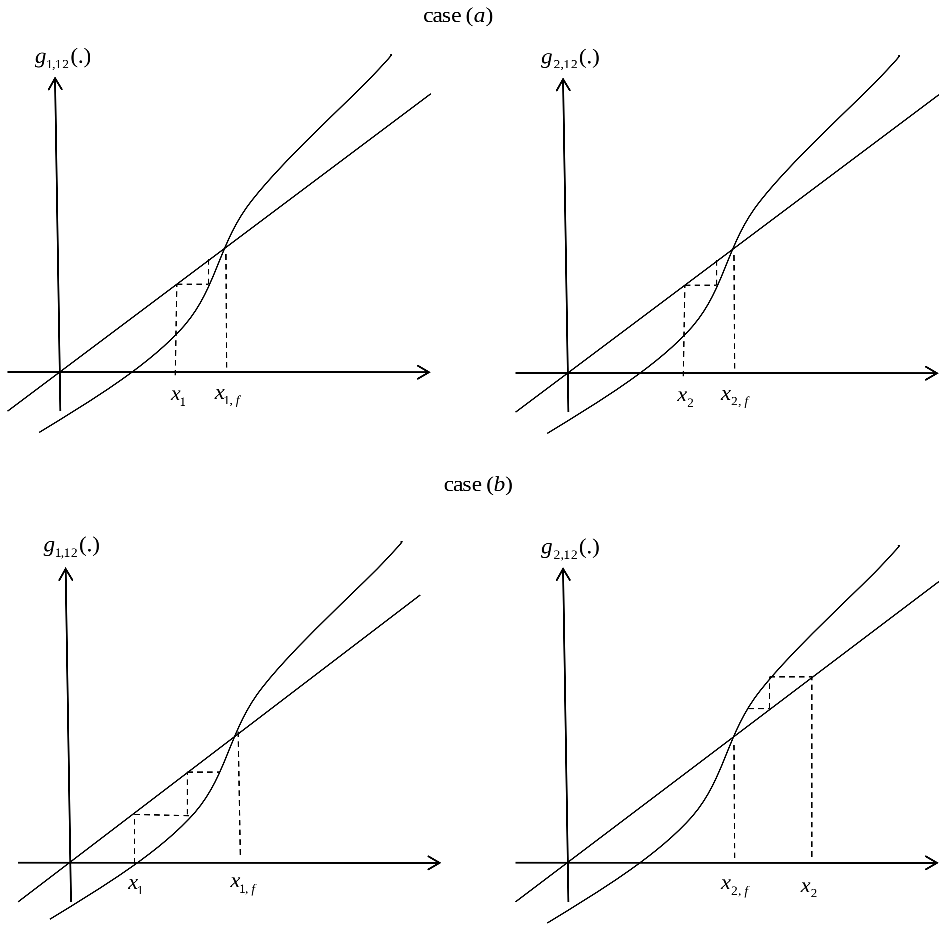

Figure 5: A case where $\mathcal{X}_0 = \mathcal{X}$ when $K = d = 2$ under Assumption 7.

In short, a condition on the position of the coordinates of g_{12} is necessary and sufficient to secure identification when $K = d = 2$. The sufficient part of this result actually extends to any K and d, as Theorem 5.2 shows.

Theorem 5.2 *Under Assumptions 1-4, if there exists $g = (g_1, ..., g_d) \in G$ with exactly one fixed point $x_f = (x_{1,f}, ..., x_{d,f})$ and such that $sgn\left[(g_m(x) - x)(x - x_{m,f})\right]$ does not depend on $m \in \{1, ..., d\}$, then $\mathcal{X}_0 = \mathcal{X}$.*

The restriction on signs and on the uniqueness of the fixed point may be unnecessary when $K \geq 3$, because we can use other functions in the group G to extend the set where $\varphi(., u)$ is identified. However, it seems difficult to find general sufficient conditions for getting $\mathcal{X}_0 = \mathcal{X}$ in this case.

6 Conclusion

In this paper, we study the identification, with a discrete instrument at hand, of a nonseparable function relating a continuous outcome to a continuous endogenous variable. Using group theory, we show that with one endogenous variable, the model is either partially or fully identified with a binary instrument, and fully identified under mild restrictions when the instrument takes three values or more. These results have implications for the design of experiments where the aim is to evaluate the effect of a continuous and endogenous treatment. We show that under very mild conditions, we can recover any treatment effects as soon as the instrument used in the experiment takes three values. With two values only, we can either identify some or all treatment effects, depending on whether the instrument acts monotonically or not.

We partially extend these results to the multivariate case. This latter case seems to raise delicate issues, and a future avenue of research is to obtain general sufficient conditions for full identification. Another issue we have not addressed here is estimation. In the free case, we could rely on an estimator defined by induction, as in the context of adverse selection models (see D'Haultfoeuille and Février, 2007). The adaptation to the current context of nonparametric IV models remains however to be done. Finally, we believe that the idea of using group theory for identification, which we also put forwards in our study of adverse selection models (see D'Haultfoeuille and Février, 2010), may be applied to other economic models.

Appendix: proofs

6.1 Proof of Theorem 3.1

First, by Lemma 3.2, we can identify $\varphi(.,.)$ on $\overline{\mathcal{O}}_{x_0} \times (0,1)$. Let x_1 be such that $\overline{\mathcal{O}}_{x_1} \cap \overline{\mathcal{O}}_{x_0} \cap \mathcal{X} \neq \emptyset$, and let us show that $\varphi(.,.)$ is identified on $\overline{\mathcal{O}}_{x_1} \times (0,1)$. Let $\widetilde{x} \in \overline{\mathcal{O}}_{x_0} \cap \overline{\mathcal{O}}_{x_1} \cap \mathcal{X}$. $\widetilde{x} \in \overline{\mathcal{O}}_{x_0}$ implies that $\varphi(\widetilde{x},.)$ is identified. Besides, by definition of the closure of $\mathcal{O}_{x_1}$ and because this orbit is identified, we can build an identified sequence $(x^n)_{n\in\mathbb{N}}$ in $\mathcal{O}_{x_1}$ such that $\lim_{n\to\infty} x^n = \widetilde{x}$. Fix y in the interior of $\text{Supp}(Y|X = x^0)$, and let u be such that $\varphi(x^0, u) = y$. We shall show that u is identified. Because y is arbitrary, this proves that the inverse of $\varphi(x^0,.)$, and thus $\varphi(x^0,.)$, is identified. Then, reasoning as with x_0, $\varphi(.,.)$ is identified on $\overline{\mathcal{O}}_{x^0} = \overline{\mathcal{O}}_{x_1}$.

By definition of the orbit of x_1, there exists $g_n \in G$ such that $x^n = g_n(x_1)$, for all $n \geq 0$. Hence, there exists $\widetilde{g}_n \in G$ (defined by $\widetilde{g}_n = g_n \circ g_0^{-1}$) such that $x^n = \widetilde{g}_n(x^0)$. Because $\widetilde{g}_n \in G$ and G is the group generated by the $(g_{ij})_{(i,j)\in\{1,...,K\}^2}$, there exists $(i_1, j_1, ..., i_k, j_k)$ such that $\widetilde{g}_n = g_{i_1 j_1} \circ ... \circ g_{i_k j_k}$. Using Equation (3.1) and a simple induction, there exists an identified function Q_n such that

$$\varphi(x^n, u) = Q_n \circ \varphi(x^0, u) = Q_n(y).$$

Because $\varphi(., u)$ is continuous, the left-hand side tends to $\varphi(\widetilde{x}, u)$. As a result, the right-hand side also admits a limit, and

$$u = \varphi^{-1}(\widetilde{x}, \lim_{n\to\infty} Q_n(y)).$$

where $\varphi^{-1}(x,.)$ denotes the inverse of $\varphi(x,.)$. The right-hand side is identified since $\varphi(\widetilde{x},.)$ and Q_n are identified. Hence u is identified, and $\varphi(.,.)$ also is on $\overline{\mathcal{O}}_{x_1}$.

Finally, let $x' \in \mathcal{X}_0$. By definition, there exists $x_1, x_2, ..., x_n = x'$ such that for all $i = 0, ..., n-1$,

$$\overline{\mathcal{O}}_{x_i} \cap \overline{\mathcal{O}}_{x_{i+1}} \cap \mathcal{X} \neq \emptyset.$$

Using what precedes, we identify by a straightforward induction φ on $\overline{\mathcal{O}}_{x_i}$ for $i = 0, ..., n$, and in particular at $x' = x_n$. ■

6.2 Proof of Theorem 4.1

Let us first prove the "if" part, by showing that $\mathcal{X}_0 = \{g_{12}^k(x_0) : k \in \mathbb{Z}\}$. It follows from the discussion above the theorem that $\mathcal{O}_{x_0} = \{x_k : k \in \mathbb{Z}\}$. The sequence $(x_k)_{k\in\mathbb{Z}}$ being strictly monotonic, its limit at $+\infty$ (and similarly at $-\infty$) is either $\inf \operatorname{Supp}(X)$ or $\sup \operatorname{Supp}(X)$. Because neither belongs to $\mathcal{X}$,

$$\overline{\mathcal{O}}_{x_0} \cap \mathcal{X} = \mathcal{O}_{x_0}.$$

Now consider $x \in \mathcal{X} \backslash \mathcal{O}_{x_0}$, so that there exists $k \in \mathbb{Z}$ satisfying $x_k < x < x_{k+1}$. As before, $\mathcal{O}_x = \{g_{12}^k(x) : k \in \mathbb{Z}\}$ and $\overline{\mathcal{O}}_x \cap \mathcal{X} = \mathcal{O}_x$. Besides, because g_{12} is strictly increasing, $x_{k+k'} < g_{12}^{k'}(x) < x_{k+k'+1}$ for all $k' \in \mathbb{Z}$, and

$$\overline{\mathcal{O}}_x \cap \overline{\mathcal{O}}_{x_0} \cap \mathcal{X} = \emptyset.$$

Hence, for all $x \in \mathcal{X} \backslash \mathcal{O}_{x_0}$, x does not belong to the equivalence class of x_0, and $\mathcal{X}_0 = \mathcal{O}_{x_0}$. The "if" part of Theorem 4.1 follows from Theorem 3.1.

Let us turn to the "only if" part, by showing that $\varphi(\widetilde{x}, .)$ is not identified for any given $\widetilde{x} \notin \mathcal{X}_0$. For that purpose, we show that for any arbitrary strictly increasing function γ, we can define $\widetilde{\varphi}$, $(\widetilde{Y}_x, \widetilde{U}_x)_{x \in \mathcal{X}}$ such that

(i) $\widetilde{\varphi}(x, u) = \varphi(x, u)$ for all $x \in \mathcal{O}_{x_0}$ and $\widetilde{\varphi}(\widetilde{x}, u) = \gamma(u)$;

(ii) $\widetilde{Y}_x = \widetilde{\varphi}(x, \widetilde{U}_x)$ for all $x \in \mathcal{X}$ and $F_{\widetilde{Y}|X,Z} = F_{Y|X,Z}$;

(iii) Assumptions 1-4 are satisfied for the model defined by $(\widetilde{Y}_x)_{x \in \mathcal{X}}, X, Z, V, (\widetilde{U}_x)_{x \in \mathcal{X}}$ and $\widetilde{\varphi}$.

We first define $\widetilde{\varphi}$ so that (i) holds. Let $k \in \mathbb{Z}$ be such that $x_k < \widetilde{x} < x_{k+1}$. We define $\widetilde{\varphi}(., .)$ by any continuous function on $[x_k, x_{k+1}) \times (0, 1)$ such that $\widetilde{\varphi}(x, .)$ is strictly increasing for all $x \in [x_k, x_{k+1})$, $\widetilde{\varphi}(x_k, u) = \varphi(x_k, u)$, $\widetilde{\varphi}(\widetilde{x}, u) = \gamma(u)$ and

$$\lim_{x \to x_{k+1}, u' \to u} \widetilde{\varphi}(x, u') = \varphi(x_{k+1}, u). \tag{6.1}$$

We then extend $\widetilde{\varphi}(., .)$ on $\mathcal{X} \times (0, 1)$ using inductively

$$\begin{aligned} \widetilde{\varphi}(g_{12}(x), u) &= F^{-1}_{Y|Z=j,V=\psi^{-1}(i,x)} \circ F_{Y|Z=i,V=\psi^{-1}(i,x)}(\widetilde{\varphi}(x, u)), \\ \widetilde{\varphi}(g_{12}^{-1}(x), u) &= F^{-1}_{Y|Z=i,V=\psi^{-1}(i,x)} F_{Y|Z=j,V=\psi^{-1}(i,x)}(\widetilde{\varphi}(x, u)). \end{aligned} \tag{6.2}$$

By a straightforward induction, $\widetilde{\varphi}(x, u) = \varphi(x, u)$ for all $x \in \mathcal{O}_{x_0}$, and (i) is satisfied.

We now define $(\widetilde{Y}_x, \widetilde{U}_x)_{x \in \mathcal{X}}$ such that (ii) holds. Consider a random variable $\widetilde{U}$ independent of X conditional on V and such that for all $(j, x, u) \in \{1, ..., K\} \times \mathcal{X} \times \operatorname{Supp}(U)$,

$$F_{\widetilde{U}|V=\psi^{-1}(j,x))}(u) = F_{U_{x_0}|V=\psi^{-1}(j,x)}(\varphi^{-1}(x, \widetilde{\varphi}(x, u))). \tag{6.3}$$

Letting for all $x \in \mathcal{X}$ $\widetilde{U}_x = \widetilde{U}$ and $\widetilde{Y}_x = \widetilde{\varphi}(x, \widetilde{U}_x)$, we have

$$\begin{aligned} P(\widetilde{Y} \leq y | X = x, Z = j) &= P(\widetilde{\varphi}(x, \widetilde{U}) \leq y | X = x, V = \psi^{-1}(j, x)) \\ &= P(\widetilde{U} \leq \widetilde{\varphi}^{-1}(x, y) | X = x, V = \psi^{-1}(j, x)) \\ &= P(\widetilde{U} \leq \widetilde{\varphi}^{-1}(x, y) | V = \psi^{-1}(j, x)) \\ &= F_{U_{x_0}|V=\psi^{-1}(j,x)}(\varphi^{-1}(x, \widetilde{\varphi}(x, \widetilde{\varphi}^{-1}(x, y)))) \\ &= F_{U_{x_0}|V=\psi^{-1}(i,x)}(\varphi^{-1}(x, y)) \\ &= F_{Y|X=x,Z=j}(y), \end{aligned}$$

and (ii) is satisfied.

Finally, let us prove (iii). First, by construction, $\widetilde{U} \perp\!\!\!\perp X|V$ or equivalently, $(\widetilde{U}_x)_{x\mathcal{X}} \perp\!\!\!\perp Z|V$. Because $V \perp\!\!\!\perp Z$, one can check easily that $((\widetilde{U}_x)_{x\mathcal{X}}, V) \perp\!\!\!\perp Z$, and Assumption 1 is satisfied. Second, the rank similarity condition trivially holds since $\widetilde{U}_x = \widetilde{U}$ for all x. Third, the function $\widetilde{\varphi}(x, .)$ is strictly increasing. Indeed, it is strictly increasing on $[x_k, x_{k+1})$ and $F^{-1}_{Y|Z=j,V=\psi^{-1}(i,x)} \circ F_{Y|Z=i,V=\psi^{-1}(i,x)}$ is strictly increasing. Thus Assumption 3 holds. Fourth, we have to verify that Assumption 4, (ii) and (iv) hold, (i) and (iii) being automatically satisfied. $\widetilde{\varphi}$ is continuous on $[x_k, x_{k+1}) \times (0, 1)$. We have $\widetilde{\varphi}(x_{k+1}, u) = \varphi(x_{k+1}, u)$, so that by (6.1), $\widetilde{\varphi}$ is continuous at (x_{k+1}, u) for all $u \in (0, 1)$ and when coming on the left with x. To prove that $\widetilde{\varphi}(., .)$ on $\mathcal{X} \times (0, 1)$, it suffices, by (6.2), to prove that $F^{-1}_{Y|Z=j,V=\psi^{-1}(i,x)} \circ F_{Y|Z=i,V=\psi^{-1}(i,x)}$ is continuous. First, reasoning as in Subsection 3.1,

$$F_{Y|Z=i,V=\psi^{-1}(i,x)}(y) = F_{U_{x_0}|V=\psi^{-1}(i,x)}(\varphi^{-1}(x, y)),$$

where $\varphi^{-1}(x, .)$ denotes the inverse of $\varphi(x, .)$. Thus, by Assumption 4, (i) and (iv), $(x, y) \mapsto F_{Y|Z=j,V=\psi^{-1}(i,x)}(y)$ is continuous. Similarly, $(x, \tau) \mapsto F^{-1}_{Y|Z=j,V=\psi^{-1}(i,x)}(\tau)$ is continuous. As a result, $\widetilde{\varphi}(., .)$ is continuous on $\mathcal{X} \times (0, 1)$ and Assumption 4, (iv) holds. Finally, the function $(u, v) \mapsto F_{U_{x_0}|V=v}(\varphi^{-1}(x, \widetilde{\varphi}(x, u)))$ is continuous and strictly increasing in u as a composition of continuous and strictly increasing functions. Thus, by (6.3), $F_{\widetilde{U}|V}$ satisfies Assumption 4, (ii). The result follows. ■

6.3 Proof of Corollary 4.2

It suffices to show that for all (k, l) the distribution of Y_{x_k} conditional on $X = x_l$ is identified. We prove actually more, i.e. that the distribution of Y_{x_k} conditional on $(X =$

$x_l, Z = z)$ is identified. Fixing $y \in \text{Supp}(Y|X = x, Z = 1)$, we have

$$\begin{aligned} P(Y_{x_k} \leq y|X = x_l, Z = 1) &= P(U_{x_k} \leq \varphi^{-1}(x_k, y)|V = v_l, Z = 1) \\ &= P(U_{x_k} \leq \varphi^{-1}(x_k, y)|V = v_l) \\ &= P(U_{x_l} \leq \varphi^{-1}(x_k, y)|V = v_l) \\ &= P(U_{x_l} \leq \varphi^{-1}(x_k, y)|V = v_l, Z = 1) \\ &= P(Y_{x_l} \leq \varphi(x_l, \varphi^{-1}(x_k, y))|X = x_l, Z = 1) \\ &= P(Y \leq \varphi(x_l, \varphi^{-1}(x_k, y))|X = x_l, Z = 1). \end{aligned}$$

The last term is identified, so that the distribution of Y_{x_k} conditional on $(X = x_l, Z = 1)$ is also identified. Similarly, the distribution of Y_{x_k} conditional on $(X = x_l, Z = 2)$ is identified. Hence, the distribution of Y_{x_k} conditional on $X = x_l$ is also identified. This implies that Δ^{ATE}_{jkl} and $\Delta^{QTE}_{jkl}(\tau)$ are identified. ■

6.4 Proof of Theorem 4.3

Before proving the theorem, let us introduce some notations. We let hereafter $t(x) = x + 1$ denote the translation on the real line and $\pi(x)$ denotes the fractional part of x or, equivalently, the projection on the unit circle $[0, 1)$.[13] A group generated by functions $f_1, ..., f_p$ and the composition operator is denoted $[f_1, ..., f_p]$. Because we consider several groups, we let, to avoid any confusion, $\mathcal{O}^{G'}_x$ denote the orbit of x generated by the group G', with the group action $g.x = g(x)$. To establish the result, we shall show in three steps that $\overline{\mathcal{O}}^G_{x_0} = \text{Supp}(X)$.

1. There exists a function f such that if $\overline{\mathcal{O}}^{[f,t]}_x$ is dense for all x, then $\overline{\mathcal{O}}^G_{x_0} = \text{Supp}(X)$.

First, without loss of generality, we set the indices (i, j, k) defined in Assumption 6 to $1, 2, 3$. g_{12} does not admit any fixed point. Suppose without loss of generality that $g_{12}(x) > x$ (otherwise it suffices to consider $x \mapsto x - 1$ instead of $t(.)$). By Assumption 6, $g_{12} = \psi(2, .) \circ \psi(1, .)^{-1}$ is a C^2 diffeomorphism on $(\underline{x}, \overline{x})$. We first prove that there exists an increasing C^2 diffeomorphism h from $\mathbb{R}$ to $(\underline{x}, \overline{x})$ such that $g_{12} = h \circ t \circ h^{-1}$. Let us consider an increasing C^2 diffeomorphism $\widetilde{h}$ defined on $[0, 1)$ such that $\widetilde{h}(0) > \underline{x}$, $\lim_{x \to 1} \widetilde{h}(x) = g_{12} \circ \widetilde{h}(0)$, $\lim_{x \to 1} \widetilde{h}'(x) = \left[g_{12} \circ \widetilde{h}\right]'(0)$ and $\lim_{x \to 1} \widetilde{h}''(x) = \left[g_{12} \circ \widetilde{h}\right]''(0)$. Such a $\widetilde{h}$ exists. Then define the function h by $h = \widetilde{h}$ on $[0, 1)$ and extend it on the real line, using $h(x+1) =$

[13] Formally, the unit circle corresponds to classes of equivalence for the equivalence relationship $\mathcal{R}'$ defined on $\mathbb{R}$ by $x\mathcal{R}'y \Leftrightarrow x - y \in \mathbb{Z}$, but this can be ignored here.

$g_{12} \circ h(x)$ or $h(x) = g_{12}^{-1} \circ h(x+1)$. By construction, h is strictly increasing and C^2. Hence, it admits a limit at $-\infty$ and $+\infty$. Suppose that $\lim_{x \to -\infty} h(x) = M > \underline{x}$. Then, because $h(x+1) = g_{12} \circ h(x)$, we would have $g_{12}(M) = M$, a contradiction. Thus, $\lim_{x \to -\infty} h(x) = \underline{x}$. Similarly, $\lim_{x \to +\infty} h(x) = \overline{x}$. Consequently, h is a C^2 diffeomorphism from $\mathbb{R}$ to $(\underline{x}, \overline{x})$.

Now, let $f = h^{-1} \circ g_{13} \circ h$. By a straightforward induction, $\mathcal{O}_{x_0}^{G} = h\left(\mathcal{O}_{h^{-1}(x_0)}^{H}\right)$. Thus, because h is continuous and onto $(\underline{x}, \overline{x})$, $\mathcal{O}_{x_0}^{G}$ is dense if $\overline{\mathcal{O}}_{x}^{[f,t]}$ is dense for all $x \in \mathbb{R}$.

2. There exists a function $\widetilde{f}$ on $[0,1)$ such that if $\overline{\mathcal{O}}_{\dot{x}}^{[\widetilde{f}]}$ is dense on the unit circle for all $\dot{x} \in [0,1)$, then $\overline{\mathcal{O}}_{x}^{[f,t]}$ is dense for all $x \in \mathbb{R}$.

By definition of f and by Assumption 5, the group action of $[f,t]$ is free. Thus, by a theorem of Hölder (see, e.g., Ghys, 2001, Theorem 6.10), $[f,t]$ is abelian. As a result, $f(x+1) = f \circ t(x) = t \circ f(x) = f(x) + 1$ for all $x \in \mathbb{R}$. This allows us to define $\widetilde{f}$ on the unit circle $[0,1)$ by $\widetilde{f} \circ \pi = \pi \circ f$. $\widetilde{f}$ is well defined because

$$\begin{aligned}
\pi(x) = \pi(y) \quad &\Leftrightarrow \quad \exists\, k \in \mathbb{Z} \;/\; x = y + k \\
&\Rightarrow \quad f(x) = f(y+k) = f(y) + k \\
&\Rightarrow \quad \pi \circ f(x) = \pi \circ f(y).
\end{aligned}$$

Fix $(x,y) \in \mathbb{R}^2$ and consider a neighbourhood $\mathcal{V}_y$ of y. By definition of the topology on the unit circle, $\pi(\mathcal{V}_y)$ is a neighbourhood of $\pi(y)$ in the unit circle. Because $\mathcal{O}_{\pi(x)}^{[\widetilde{f}]}$ is dense in $[0,1)$, there exists $n \in \mathbb{Z}$ such that $\widetilde{f}^n \circ \pi(h^{-1}(x_0)) \in \pi\left(\mathcal{V}_y\right)$. By definition of $\widetilde{f}$, $\widetilde{f}^2 \circ \pi = \widetilde{f} \circ (\pi \circ f) = \pi \circ f^2$, so that, by a direct induction,

$$\widetilde{f}^n \circ \pi = \pi \circ f^n, \quad \forall n \in \mathbb{Z} \tag{6.4}$$

Hence, $\pi \circ f^n(h^{-1}(x_0)) \in \pi\left(\mathcal{V}_y\right)$, and there exists $m \in \mathbb{Z}$ such that $t^m \circ f^n(x) \in \mathcal{V}_y$. This proves that if $\mathcal{O}_{\dot{x}}^{[\widetilde{f}]}$ is dense for all $\dot{x} \in [0,1)$, then $\mathcal{O}_{x}^{[f,t]}$ is dense on the real line for all $x \in \mathbb{R}$.

3. $\overline{\mathcal{O}}_{x}^{[\widetilde{f}]}$ is dense on the unit circle for all $\dot{x} \in [0,1)$.

Because g_{13} and h are increasing C^2 diffeomorphisms, so is f. Because $\widetilde{f} \circ \pi = \pi \circ f$, $\widetilde{f}$ is thus an orientation-preserving C^2 diffeomorphism on the unit circle. We can thus apply Denjoy's theorem (see, e.g., Navas, 2009, Theorem 3.1.1), and $\overline{\mathcal{O}}_{\dot{x}}^{[\widetilde{f}]}$ is either finite or dense. Suppose that it is finite. Then there exists $n \in \mathbb{Z}^*$ such that $\widetilde{f}^n(\dot{x}) = \dot{x}$. Let $x \in \mathbb{R}$ be such that $\pi(x) = \dot{x}$. Then, using (6.4), there exists $m \in \mathbb{Z}$ such that $f^n(x) = t^m(x)$. Hence, by definition of f and t, $g_{13}^n(x) = g_{12}^m(x)$ with $n \neq 0$, contradicting Assumption 6. We thus conclude that any orbit for the group generated by $\widetilde{f}$ is dense in $[0,1)$. ■

6.5 Proof of Theorem 4.4

Let $x^1 < ... < x^M$ denote the fixed points of g, $x^0 = \underline{x}$ and $x^{M+1} = \overline{x}$ and let i be such that $x_0 \in [x^i, x^{i+1})$. Suppose for instance that $g(x) > x$ for all $x \in (x^i, x^{i+1})$ (the proof is identical if $g(x) < x$). Suppose first that $x_0 > x^i$. A straightforward induction shows that the sequence $(g^n(x_0))_{n\in\mathbb{N}}$ is increasing and bounded by x^{i+1}. Thus, it converges to $l \in (x^i, x^{i+1}]$ which satisfies $g(l) = l$, and $l = x^{i+1}$. Hence $x^{i+1} \in \overline{O}_{x_0}$. Using similarly the sequence $(g^{-n}(x_0))_{n\in\mathbb{N}}$ shows that $x^i \in \overline{O}_{x_0}$.

Applying the same reasoning to any $x \in (x^i, x^{i+1})$ establishes that $x^i \in \overline{O}_x$. This proves that $x\mathcal{R}x_0$, and $[x^i, x^{i+1}] \subset \mathcal{X}_0$. In a similar way, $x^i \in \overline{O}_x$ for all $x \in (x^{i-1}, x^i)$ (if $i > 0$), and we get $x_0\mathcal{R}x$ for any $x \in (x^{i-1}, x^i)$. By induction, $(\underline{x}, \overline{x}) \subset \mathcal{X}_0$, and thus $\mathcal{X}_0 = \mathcal{X}$.

Finally, suppose that $x_0 = x^i$. Then, as previously, $x_0 \in \overline{O}_x$ for any $x \in (x^i, x^{i+1})$, and $x_0\mathcal{R}x$, for any $x \in [x^i, x^{i+1}]$. We conclude as before. ■

6.6 Proof of Theorem 5.1

Bourbaki (1974) shows (see paragraph 1, n°3) that H is dense in $\mathbb{R}^d$ if and only if Condition (5.2) holds. The first result follows because $\mathcal{O}_{x_0} = \lambda(H + \lambda^{-1}(x_0))$ and λ is continuous. To prove the necessary condition, suppose that $K = d + 1$ (the case where $K < d + 1$ is similar), and let A denote the $d \times d$-matrix of typical (i, j) element a_{ij}. If A is nonsingular, taking $x = A^{-1}f$, for any nonzero $f \in \mathbb{Z}^d$ implies that $Ax \in \mathbb{Z}^d$. As a result, $< x, H > \subset \mathbb{Z}$, implying that $\{x \in \mathbb{R}^d :< x, H > \subset \mathbb{Z}\} \neq \{0\}$. If A is singular, then any $x \neq 0$ in the kernel of A satisfies $Ax \in \mathbb{Z}^d$, and then again $\{x \in \mathbb{R}^d :< x, H > \subset \mathbb{Z}\} \neq \{0\}$. Thus, Condition (5.2) fails to hold, and $\mathcal{X}_0 \not\subset \mathcal{X}$. ■

6.7 Proof of Theorem 5.2

Suppose without loss of generality that $\text{sgn}\left[(g_m(x) - x)(x - x_{m,f})\right] = -1$ for all $m = 1...d$. To prove Theorem 5.2, it suffices to show that $x_f = \lim_{k\to\infty} g^k(x)$ for all $x = (x_1, ..., x_d) \in \mathcal{X}$, or, equivalently, that for all $m = 1...d$, $x_{m,f} = \lim_{k\to\infty} g_m^k(x_m)$.

If $x_m < x_{m,f}$, a straightforward induction shows that $(g_m^k(x))_{k\in\mathbb{N}}$ is increasing and bounded above by $x_{m,f}$. Because g has a unique fixed point, $x_{m,f} = \lim_{k\to\infty} g_m^k(x_m)$. Similarly, if $x_m > x_{m,f}$, $g_m^k(x)$ is decreasing and bounded below by $x_{m,f}$, and the sequence also converges to $x_{m,f}$. ■

References

Bourbaki, N. (1974), *Topologie générale, chap. VII*, Hermann, Paris.

Chernozhukov, V. and Hansen, C. (2005), 'An IV model of quantile treatment effects', *Econometrica* **73**, 245–261.

Chesher, A. (2007), 'Instrumental values', *Journal of Econometrics* **73**, 1525–1550.

Chesher, A. (2010), 'Instrumental variable models for discrete outcomes', *Econometrica* **78**, 575–601.

Chiappori, P. A., Komunjer, I. and Kristensen, D. (2010), Identification and estimation of a transformation model. Mimeo.

de Mel, S., McKenzie, D. and Woodruff, C. (2008), 'Returns to capital in microenterprises:evidence from a field experiment', *Quarterly Journal of Economics* **123**, 1329–1372.

D'Haultfoeuille, X. (2011), 'On the completeness condition in nonparametric instrumental regressions', *Econometric Theory, forthcoming* .

D'Haultfoeuille, X. and Février, P. (2007), Identification and estimation of incentive problems: Adverse selection. CREST working paper.

D'Haultfoeuille, X. and Février, P. (2010), Inference on a class of adverse selection models through contracts variation. Mimeo.

Doksum, K. (1974), 'Empirical probability plots and statistical inference for nonlinear models in the two-sample case', *The Annals of Statistics* **2**, 267–277.

Florens, J., Heckman, J. J., Meghir, C. and Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica* **76**, 1191–1206.

Ghys, E. (2001), 'Groups acting on the cirle', *L'enseignement mathématique* **47**, 329–407.

Guerre, E., Perrigne, I. and Vuong, Q. (2009), 'Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions', *Econometrica* **77**, 1193–1227.

Hoderlein, S. and Mammen, E. (2007), 'Identification of marginal effects in nonseparablle models without monotonicity', *Econometrica* **75**, 1513–1518.

Imbens, G. W. and Angrist, J. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–475.

Imbens, G. W. and Newey, W. K. (2009), 'Identification and estimation of triangular simultaneous equations models without additivity', *Econometrica* **77**, 1481–1512.

Kocecki, A. (2010), Algebraic theory of identification in parametric models. MPRA working paper No 26820.

Matzkin, R. L. (2007), Nonparametric identification, *in* J. J. Heckman and E. E. Lleamer, eds, 'Hanbook of Econometrics', Vol. 6B, Elsevier.

Navas, A. (2009), Groups of circle diffeomorphisms. Mimeo.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Newey, W. K., Powell, J. L. and Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

Severini, T. and Tripathi, G. (2006), 'Some identification issues in nonparametric linear models with endogenous regressors', *Econometric Theory* **22**, 258–278.

Stillwell, J. (1992), *Geometry of Surfaces*, Springer Verlag.

Torgovitsky, A. (2011), Identification of nonseparable models with general instruments. Mimeo.