Identification of Nonseparable Models with Endogeneity and Discrete Instruments*

Xavier D'Haultfœuille † Philippe Février‡

May 2012

Abstract

We study the identification of a nonseparable function that relates a continuous outcome to a continuous endogenous variable. We suppose to have in hand an instrument, and assume monotonicity in both the first stage and the outcome equation. We show that the combination of these restrictions has a large identifying power: full identification can be achieved even though the instrument is discrete. This conclusion highlights the importance of justifying monotonicity restrictions in economic applications. 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 is 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, Monotonicity, Fixed Points, Group Theory.

JEL classification numbers: C14.

*We thank James Stock and five anonymous referees for their remarks. 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 (ENSAE). E-mail address: xavier.dhaultfoeuille@ensae.fr.

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

1 Introduction

In this paper, we study the identification of a nonseparable model with a continuous endogenous variable when an instrument is available. This issue is important because usual assumptions such as linearity or separability of the error terms are seldom justified by theory, and are also likely to fail in practice. To investigate this question, two main approaches have been taken (see, e.g., Matzkin, 2007, for a more complete discussion). The first is based on estimating equations (see, e.g., Newey and Powell, 2003 or Chernozhukov and Hansen, 2005), the second on control variable approaches (see, e.g., Newey et al., 1999, Florens et al., 2008, Imbens and Newey, 2009 and Hoderlein and Sasaki, 2011*a*). A common feature of both approaches is that when the endogenous variable is continuous, the instrument should be continuous to achieve identification. We alleviate this restriction here. This can be useful because in many cases, we have at our disposal only discrete instruments. Typical examples are policy reforms or randomizations in the treatment or control group in experiments. More precisely, we show that imposing a monotonicity condition on the outcome equation, as Chernozhukov and Hansen (2005), but also on the first stage equation, as Imbens and Newey (2009), is sufficient to get partial or full identification when the instrument is discrete.

The ideas behind these results are the following. Using monotonicity in the first stage, and following the control function approach, we consider 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 $x' = s(x)$ is identified. Observing its effect on the outcome, one can relate, under monotonicity in the outcome equation, the structural function g at x with itself at $x' = s(x)$. Considering all the functions s associated with any change in the value of the instrument, we show that the problem of identifying g is closely related to the properties of $\mathcal{S}$, the group generated by all compositions of these functions s. Basically, a function in $\mathcal{S}$ can be interpreted as the effect of a new, binary, instrument. If the group is "rich" enough, any possible exogenous shift can be approximated by an element of this group, as if we had at hand a continuous instrument.

Studying the properties of this group, we distinguish whether X is univariate or not, and whether a freeness condition is satisfied or not. This latter property means that no function in $\mathcal{S}$ different from the identity function admits a fixed point. With a binary instrument, it is equivalent to monotonicity of the instrument, a condition close to the one of Imbens and Angrist (1994) but directly testable here. When X is univariate and the instrument is binary, we show that the model is identified on a countable set only under freeness, but fully identified otherwise. Hence, the intuition that monotonicity helps for identification, which may be conveyed by the result of Imbens and Angrist (1994) with dummy endogenous variables, is reversed here. On the other hand, when the instrument takes three values or more, the model is fully identified in general, whether freeness holds or not. Finally, when X is multivariate, the properties of the group become more complicated and it seems difficult to obtain a full classification. We show that full identification can still be achieved, but under additional restrictions.

Overall, our results emphasize the important identification power of monotonicity conditions. Without monotonicity on the first-stage and on the outcome equation, results of Chesher et al. (2011) suggest that we generally end up with partial identification. Hoderlein and Sasaki (2011*a*) also prove that an index monotonicity on the first stage equation is necessary for some quantities to identify local partial effects. With one monotonicity restriction, either on the first stage or on the outcome equation, point identification is achieved only with a continuous instrument (see Chernozhukov and Hansen, 2005, Chernozhukov et al., 2007, Imbens and Newey, 2009 and Hoderlein and Sasaki, 2011*a*).[1] With monotonicity on both equations, we show that full identification can be achieved with a discrete instrument. In a related paper, Torgovitsky (2011) obtains a similar result under an intersection condition. His condition can be interpreted as a particular case of the non-freeness property. All these results show that monotonicity conditions are not innocuous, as already pointed out by Imbens (2007). They either imply that heterogeneity is univariate or that it can be aggregated in a single dimension. Such restrictions can be appealing for some applications, but also rule out important frameworks such as random coefficient models or simultaneous equations.

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.

[1] The results of Hoderlein and Sasaki (2011*b*) imply that some marginal effects can be bounded with a discrete instrument under monotonicity on the first stage.

2 The model

Let Y denote a real outcome, $X \in \mathbb{R}^d$ be the endogenous variable and Z be the instrument. 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:

$$\begin{aligned} Y &= g(X, \varepsilon) \\ X &= h(Z, \eta). \end{aligned} \tag{2.1}$$

Such a model is also considered by, e.g., Chernozhukov and Hansen (2005), Florens et al. (2008), Imbens and Newey (2009) and Torgovitsky (2011).[2] We aim at recovering the function g from the distribution of (X, Y, Z). We suppose to have at our disposal a discrete instrument variable $Z \in \{1, ..., K\}$, $K \geq 2$.[3]

Our first assumption is the exogeneity of the instrument. Such an assumption is standard and also posit by Florens et al. (2008), Imbens and Newey (2009) or Hoderlein and Sasaki (2011*a*). It is however stronger than the "weak" exogeneity condition $Z \perp\!\!\!\perp \varepsilon$ supposed by Chernozhukov and Hansen (2005) or Chesher (2010).

Assumption 1 *(Exogeneity)* $Z \perp\!\!\!\perp (\varepsilon, \eta)$.

The crucial assumption is the following dual monotonicity condition. Subsequently, we denote by $\mathcal{X}$ the interior of the support of X.

Assumption 2 *(Dual strict monotonicity)* $\varepsilon \in \mathbb{R}$, $\eta = (\eta_1, ..., \eta_d) \in \mathbb{R}^d$, $h(Z, \eta) = (h_1(Z, \eta_1), ..., h_d(Z, \eta_d))$ *and for all* $(x, z, m) \in \mathcal{X} \times \{1, ..., K\} \times \{1, ...d\}$, $u \mapsto g(x, u)$ *and* $v \mapsto h_m(z, v)$ *are strictly increasing.*

This assumption gathers together the monotonicity on the outcome equation supposed by Chernozhukov and Hansen (2005), and monotonicity on the first stage used by Imbens and Newey (2009). Monotonicity in the outcome equation is typically satisfied in the additive model $g(X, \varepsilon) = \nu(X) + \varepsilon$ or in transformation models $g(X, \varepsilon) = \mu(\nu(X) + \varepsilon)$, where μ is

[2] As Chernozhukov and Hansen (2005), we could consider a more general model with potential outcomes $Y_x = g(x, \varepsilon_x)$ under a rank similarity condition on the residuals. All our results below apply to this more general setting.

[3] Our analysis also covers continuous instruments. It suffices to pick K values of such instruments. In such cases however, identification can be achieved under fewer restrictions on g or h. See e.g. Chernozhukov and Hansen (2005), Imbens and Newey (2009) and Hoderlein and Sasaki (2011*a*).

strictly increasing. More generally, dual monotonicity imposes two things. First, the error terms should be scalar. Second, it defines a one-to-one mapping between (X, Y) and (ε, η) for a given value of Z. This condition can be reasonable for some applications, but are not satisfied in all cases. Random coefficient models are a first counterexample. It also rules out Chesher (2003)'s setting, where η affects directly the outcome equation. Finally, it cannot handle simultaneity problems, as discussed by Imbens (2007).

We also impose the following common support condition.

Assumption 3 *(Common support) Support*$(X|Z = z) = \prod_{m=1}^{d}[\underline{x}_m, \overline{x}_m]$ *with* $-\infty \leq \underline{x}_m < \overline{x}_m \leq \infty$ *independent of* z.

This assumption allows the support of X conditional on $Z = z$ to be either bounded or unbounded. On the other hand, it should not depend on z. As shown in our supplementary material, much of the analysis below can be adapted without this restriction, at the price of an additional complexity.

Finally, we impose the following continuity conditions.

Assumption 4 *(Continuity conditions) (i)* η *is continuously distributed, (ii)* $(u, v) \mapsto F_{\varepsilon|\eta=v}(u)$ *is continuous and strictly increasing in* u *for all* $v \in$ *Support*(η)*, (iii)* $g(., .)$ *and* $h(z, .)$ *are continuous on* $\mathcal{X} \times$ *Support*(ε) *and Support*(η) *respectively.*

Conditions (i) and (ii) ensure that X and Y conditional on (X, Z) are continuously distributed. Condition (iii) excludes discontinuous effects of X on Y. It is important as we often obtain identification of $g(., u)$ on dense subsets of $\mathcal{X}$. By continuity, this ensures that g is identified everywhere. We finally impose continuity of $g(., u)$ on the interior $\mathcal{X}$ of the support of X, but not on the whole support. This is important to encompass linear models with unbounded X, and, more generally, models where $g(., u)$ tends to infinity at the boundaries.

3 The identification strategy

3.1 Reformulation of the identification problem

As mentioned previously, we suppose to observe Y, X and Z, and seek to recover the function g. First, applying the idea of the control function approach, we identify exogenous changes in X by moving Z while keeping η constant. Formally, for all $(x,i,j)\in\mathcal{X}\times\{1,...,K\}^2$, let

$$s_{ij}(x)=h(j,h^{-1}(i,x)),$$

where $h^{-1}(i,x)=\left(h_1^{-1}(i,x_1),...,h_d^{-1}(i,x_d)\right)$ and $h_m^{-1}(i,.)$ denotes the inverse of $h_m(i,.)$. $s_{ij}(x)-x$ is the shift in X when Z moves from i to j while η remains constant, equal to $h^{-1}(i,x)=h^{-1}(j,s_{ij}(x))$. Let similarly $s_{ijm}(x_m)=h_m(j,h_m^{-1}(i,x_m))$ denote the shift in X_m when Z moves from i to j. By strict monotonicity and independence,

$$\begin{aligned}
P(X_m\leq x_m|Z=i) &= P(\eta_m\leq h_m^{-1}(i,x_m)|Z=i)\\
&= P(\eta_m\leq h_m^{-1}(j,s_{ijm}(x_m))|Z=i)\\
&= P(\eta_m\leq h_m^{-1}(j,s_{ijm}(x_m))|Z=j)\\
&= P(X_m\leq s_{ijm}(x_m)|Z=j).
\end{aligned}$$

The first equality is satisfied by the strict monotonicity of h_m. The second equality holds by definition of s_{ijm}. The third equality stems from the independence between Z and η (Assumption 1). Hence s_{ijm} is identified by $s_{ijm}=F^{-1}_{X_m|Z=j}\circ F_{X_m|Z=i}$, where $\circ$ denote the composition operator. Because $s_{ij}(x)=(s_{ij1}(x_1),...,s_{ijd}(x_d))$, s_{ij} is also identified.

Observing the effect of such a change on Y, one can relate $g(x,u)$ and $g(s_{ij}(x),u)$, using the dual monotonicity condition. Indeed,

$$\begin{aligned}
F_{Y|X=x,Z=i}(g(x,u)) &= P(Y\leq g(x,u)|\eta=h^{-1}(i,x),Z=i)\\
&= P(\varepsilon\leq u|\eta=h^{-1}(i,x),Z=i)\\
&= P(\varepsilon\leq u|\eta=h^{-1}(i,x),Z=j)\\
&= P(Y\leq g(s_{ij}(x),u)|X=s_{ij}(x),Z=j)\\
&= F_{Y|X=s_{ij}(x),Z=j}(g(s_{ij}(x),u)).
\end{aligned}$$

The first equality is satisfied because $(X=x,Z=i)$ is equivalent, by strict monotonicity of h, to $(\eta=h^{-1}(i,x),Z=i)$. The second equality holds because $g(x,.)$ is strictly increasing (Assumption 2). Hence, the dual monotonicity ensures that, conditional on $Z=i$, there is a one-to-one mapping between the distribution of Y conditional on X and that of ε conditional on η. The third equality stems from the independence between Z and (ε,η) (Assumption 1), which implies that ε is independent of Z conditional on η. The last equalities apply the same reasoning as the first ones, but the other way around.

By Assumptions 2 and 4, $y\mapsto F_{Y|X=x,Z=i}(y)$ is strictly increasing for all (x,i).[4] Hence, its

[4] As before, $F_{Y|X=x,Z=i}(y)=F_{\varepsilon|\eta=h^{-1}(i,x)}(g^{-1}(x,y))$, where $g^{-1}(x,.)$ denotes the inverse of $g(x,.)$. $F_{Y|X=x,Z=i}(.)$ is thus the composition of two strictly increasing functions.

inverse exists, and

$$g(s_{ij}(x), u) = F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}(g(x, u)). \tag{3.1}$$

Hence, if $g(x, .)$ is identified for a given $x \in \mathcal{X}$, then $g(s_{ij}(x), .)$ is also identified. We state this in the following lemma.

Lemma 3.1 *If, for a given $x \in \mathcal{X}$, $g(x, .)$ is identified, then $g(s_{ij}(x), .)$ is also identified for all $(i, j) \in \{1, ..., K\}^2$.*

This lemma is at the basis of our identification results. It does not hold in Chernozhukov and Hansen (2005) or Imbens and Newey (2009) where only one monotonicity condition is imposed. It does however apply to the framework of Chesher (2003) where a dual monotonicity condition is also assumed.[5] Nevertheless, our induction technique presented below fails to hold in his model.

3.2 Link with group theory

It is well known that a normalization on g and ε is possible. More precisely, for any strictly increasing function f, g and ε are observationally equivalent to $\widetilde{g}$ and $\widetilde{\varepsilon}$, with $g(X, \varepsilon) = \widetilde{g}(X, f^{-1}(\varepsilon))$ and $\varepsilon = f(\widetilde{\varepsilon})$. The usual choice of f is $F_{\widetilde{\varepsilon}}$, which amounts to supposing that ε is uniformly distributed. To derive our results, it is more convenient to choose $f = F_{\widetilde{\varepsilon}|X=x_0}$ for a given $x_0 \in \mathcal{X}$, so that the distribution of ε conditional on $X = x_0$ is uniform. This normalization implies that $g(x_0, .)$ is identified by $g(x_0, u) = F^{-1}_{Y|X=x_0}(u)$. We then identify, by Lemma 3.1, $g(s_{ij}(x_0), u)$, but also $g(s_{ij} \circ s_{kl}(x_0), u)$ or $g(s_{ij} \circ s_{kl} \circ s_{mn}(x_0), u)$. By induction, $g(., u)$ is actually identified on all compositions of the $(s_{ij})_{(i,j)\in\{1,...,K\}^2}$ taken at x_0. Let $\mathcal{S}$ denote this set of functions. Because, by Assumptions 2 and 3, the functions s_{ij} are bijections onto $\mathcal{X}$, $\mathcal{S}$ has a group structure (see Appendix A for definitions related to groups). A function $s = s_{i_1 j_1} \circ ... \circ s_{i_p j_p}$ in $\mathcal{S}$ can be interpreted as the effect of a new binary instrument Z_s , which corresponds to successive shifts of Z (from i_1 to j_1 first, then from i_2 to j_2 etc).

$g(., u)$ is thus identified on the set

$$\mathcal{O}_{x_0} = \{s(x_0) : s \in \mathcal{S}\},$$

[5] More precisely, in his setting, $Y = g(X, \varepsilon, \eta)$, and one can prove that $g(s_{ij}(x), u, h^{-1}(i, x)) = F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}(g(x, u, h^{-1}(i, x)))$.

which is called the orbit of x_0. By continuity of $g(.,u)$, $g(.,u)$ is actually identified on the closure $\overline{\mathcal{O}}_{x_0}$ of $\mathcal{O}_{x_0}$.

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

This result shows that $g(.,u)$ is fully identified on $\mathcal{X}$ when $\mathcal{X} \subset \overline{\mathcal{O}}_{x_0}$. Otherwise, we may still achieve full identification by "connecting" orbits.

Lemma 3.3 *For any x, if Assumptions 1-4 hold and $\overline{\mathcal{O}}_{x_0} \cap \overline{\mathcal{O}}_x \cap \mathcal{X} \neq \emptyset$, $g(.,.)$ is identified on $(\overline{\mathcal{O}}_x \cap \mathcal{X}) \times (0,1)$.*

The proof of the lemma, and all subsequent proofs, are deferred to Appendix B. The intuition behind is that by Lemma 3.2, we can identify $g(.,u)$ at $x^* \in \overline{\mathcal{O}}_{x_0} \cap \overline{\mathcal{O}}_x \cap \mathcal{X}$. By "inverting" our induction method, we then recover $g(x,u)$ using $g(x^*,u)$.

An important consequence is that studying the identification of g amounts to determining the orbits and their connections. In turn, these orbits and connections are related to the following freeness and nonfreenes properties.[6]

Definition 3.1
Freeness property: *there exists no $s \in \mathcal{S}$ different from the identity function that admits a fixed point.*
Nonfreeness property: *there exists $s \in \mathcal{S}$ different from the identity function which admits a positive and finite number of fixed points.*

Whether freeness or nonfreeness holds depends on the way the instrument affects X. With a binary instrument ($K = 2$), $\mathcal{S}$ is generated by a unique function s_{12}. Freeness is then equivalent to s_{12} admitting no fixed point, or s_{12} being the identity function. The latter case corresponds to an instrument independent of X, which is not informative. Otherwise, we either have $h(1,.) > h(2,.)$ or $h(2,.) > h(1,.)$. This can be interpreted as an homogeneity of the instrument: all individuals either react positively or negatively to the instrument. It may also be seen as an extension, in a continuous setting, of the monotonicity condition considered by Imbens and Angrist (1994) for dummy endogenous variables. The important difference with their condition, however, is that we can test it directly in the data, by checking whether $F_{X|Z=1}$ stochastically dominates (or is dominated by) $F_{X|Z=2}$ at the first-order.

[6] Nonfreeness is not, with this definition, the strict opposite of freeness. We rule out situations where all elements of $\mathcal{S}$ cross an infinite number of times the identity function, without being the identity functions.

As an illustration, suppose that we are interested in measuring the effect of unemployment duration X on an health index Y, using a policy change on unemployment benefits as an instrument Z. Suppose that the hazard rate of X conditional on $Z = z$ satisfies a Cox model $\lambda_z(t) = \lambda_0(t)\exp(-cb_z(t))$, where λ_0 is the baseline hazard and $b_z(t)$ denotes unemployment benefits at date t under policy status z.[7] It is easy to see that if $b_1(t) < b_2(t)$ for all t, $F_{X|Z=2}$ stochastically dominates $F_{X|Z=1}$. The freeness property holds because all unemployed people have less incentives to find a job. Suppose on the other hand that before a reform, unemployment benefits were constant over time, $b_1(t) = b_1$, while after, they decrease over time, so that $b_2(t) = b_{21}\mathbb{1}\{t \leq t_0\} + b_{22}\mathbb{1}\{t > t_0\}$ for a given threshold t_0, with $b_{21} > b_1 > b_{22}$. The new policy is thus more generous for short periods of unemployment, and less generous for longer ones. Because of this pattern, $F_{X|Z=1}$ and $F_{X|Z=2}$ generally cross.[8]

When $K \geq 3$, freeness can still be interpreted as an homogenous effect of the instrument Z on X. Any instrument Z_s corresponding to the exogenous shift s should either have a strictly positive, negative or null effect on all individuals. It is thus impossible to yield an heterogenous effect on X. Generalized location models are examples of first stage equations satisfying this restriction.

Theorem 3.2 *Suppose that*

$$h(Z, \eta) = \mu(\nu(Z) + \eta), \tag{3.2}$$

where $Z \perp\!\!\!\perp \eta$, *Support*$(\eta) = \mathbb{R}$ *and* μ *is a strictly increasing function from* $\mathbb{R}$ *to* $\mathcal{X}$. *The freeness property holds.*

Model (3.2) holds for instance in the previous example of unemployment duration when $b_z(t) = a_z + b(t)$, namely when unemployment benefits under the different policies differ by the same constant over time. With $K \geq 3$, nonfreeness does not reduce to a crossing condition on the $(s_{ij})_{i \neq j}$, because all functions in $\mathcal{S}$ should be considered. We develop in the supplementary material an exemple, still with unemployment benefits, where the $(s_{ij})_{i \neq j}$ do not cross, but $s_{31} \circ s_{12}^2$ admits one fixed point.

[7] This toy model is useful to discuss the economic contents of our assumptions but does not pretend to be fully realistic.

[8] For a proof when $\lambda_0(.)$ is constant, see the supplementary material.

4 Results in the univariate case

We first consider the univariate case where $\mathcal{X} = (\underline{x}, \overline{x})$, for which the topology of the orbits is more simple.

4.1 Identification under freeness

With a binary instrument ($K = 2$), $\mathcal{S}$ is generated by a unique function s_{12}. The orbit of x_0 is discrete, and consists of the monotonic sequence $(x_k)_{k \in \mathbb{Z}}$ defined by $x_k = s_{12}^k(x_0)$. Figure 1 depicts the correspondence between this sequence and the identification of $g(., u)$. The left graph shows how to build $(x_k)_{k \in \mathbb{Z}}$ by applying $s_{12}(.)$ successively. The dashed curve in the right graph depicts the true function $x \mapsto g(x, u)$ whereas the black points correspond to the sequence where $g(., u)$ is identified. By Lemma 3.1, we can indeed identify $g(x_k, u)$ as soon as $g(x_{k-1}, u)$ has been recovered. One may expect that $g(., u)$ is identified only on this sequence. Theorem 4.1 formalizes this result.[9]

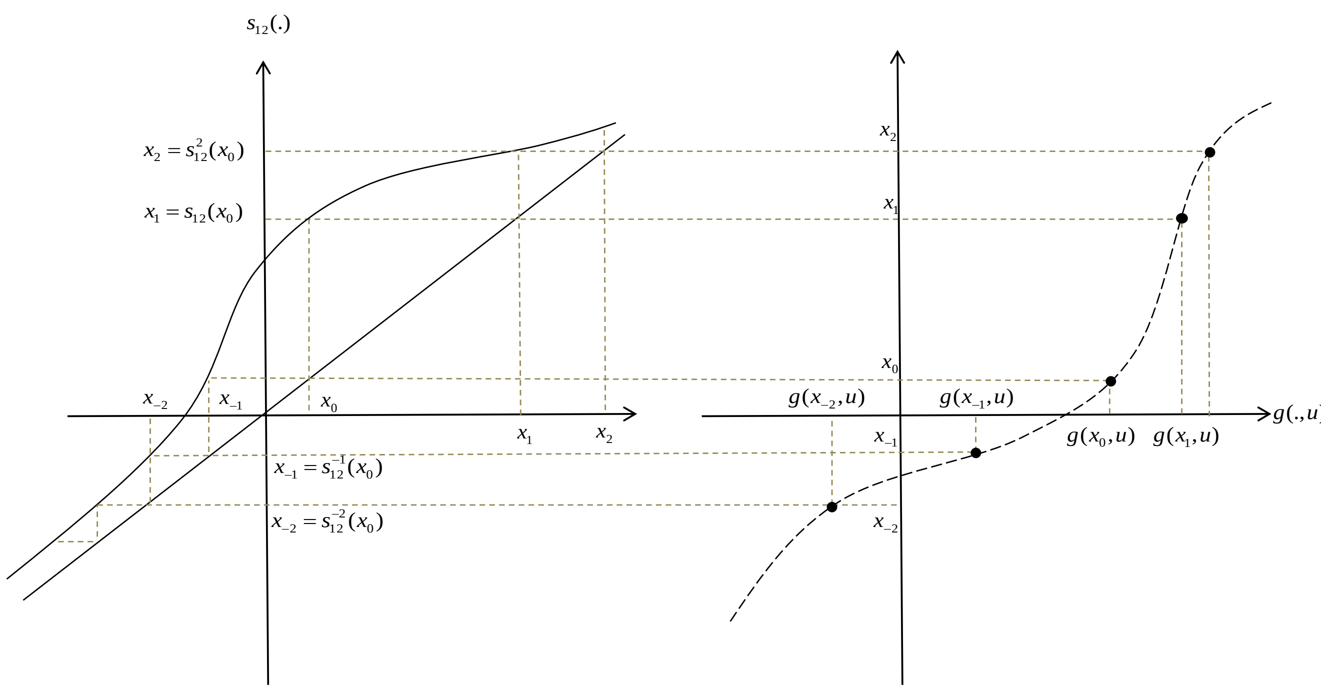

Figure 1: Identification under the freeness property, $K = 2$.

[9] When s_{12} is the identity function, we identify $g(., u)$ only at x_0. We thus get no information, which makes sense since the instrument is independent of X.

Theorem 4.1 *If $K = 2$ and if Assumptions 1-4 and the freeness property hold, $g(x, u)$ is identified for any $u \in (0, 1)$ if and only if $x \in \{x_k : k \in \mathbb{Z}\}$.*

A consequence of Theorem 4.1 is that g is fully identified and even overidentified in general when g belongs to a parametric family. Similarly, if one imposes shape restrictions such as monotonicity or concavity, g would be point identified on $\{x_k : k \in \mathbb{Z}\}$ and partially identified elsewhere. On the other hand, without further restriction, $g(x, u)$ can be chosen freely outside of $\{x_k : k \in \mathbb{Z}\}$. To fully identify g, one has to fix entirely $g(., u)$ on the interval $[x_0, x_1)$. We illustrate this nonidentification part of Theorem 4.1 in the supplementary material, through the example of an additively separable model.

Choosing a different starting point x_0 leads to a different sequence where $g(., u)$ is identified.[10] One may thus be worried about the choice of x_0. It is however important to understand that this is not an issue when considering policy relevant parameters. Theorem 4.2 shows indeed that some average and quantile treatment effects on the subpopulation defined by $X = x$ are identified, whether x belongs or not to $\{x_k : k \in \mathbb{Z}\}$.

Theorem 4.2 *If $K = 2$ and if Assumptions 1-4 and the freeness property hold, the average treatment effect $\Delta_{ij}^{ATE}(x) = E\left(g(s_{12}^j(x), \varepsilon) - g(s_{12}^i(x), \varepsilon)|X = x\right)$ and the quantile treatment effect $\Delta_{ij}^{QTE}(\tau) = F^{-1}_{g(s_{12}^j(x),\varepsilon)|X=x}(\tau) - F^{-1}_{g(s_{12}^i(x),\varepsilon)|X=x}(\tau)$ are identified for all $x \in \mathcal{X}$ and all $\tau \in (0, 1)$.*

We can identify treatment effects corresponding to exogenous shifts identical to those produced by one or multiple changes in Z. On the other hand, we cannot identify average marginal effects $E\left(\partial g/\partial x(x, \varepsilon)|X = x\right)$, because, basically, the binary instrument is not able to reproduce an exogenous infinitesimal change in X.

When $K \geq 3$, the situation is quite different from the binary case, because the orbit of x_0 does not reduce to a monotonous sequence. Figure 2 illustrates this point. As previously $x_1 = s_{12}(x_0) > x_0$ lies in $\mathcal{O}_{x_0}$. Suppose also that we apply s_{13}^{-1} to x_1, and then s_{12}. We obtain in this illustration a point $x_2 \in \mathcal{O}_{x_0}$ that lies between x_0 and x_1. We are thus able to get some information inside (x_0, x_1). The description of all the points that can be reached this way depends on the following assumption.

[10] This does not contradict Theorem 4.1. The normalization allows one to fix $g(., u)$ at one value only. Fixing g at two different values amounts to imposing restrictions on the function, even though such restrictions cannot be rejected by the data.

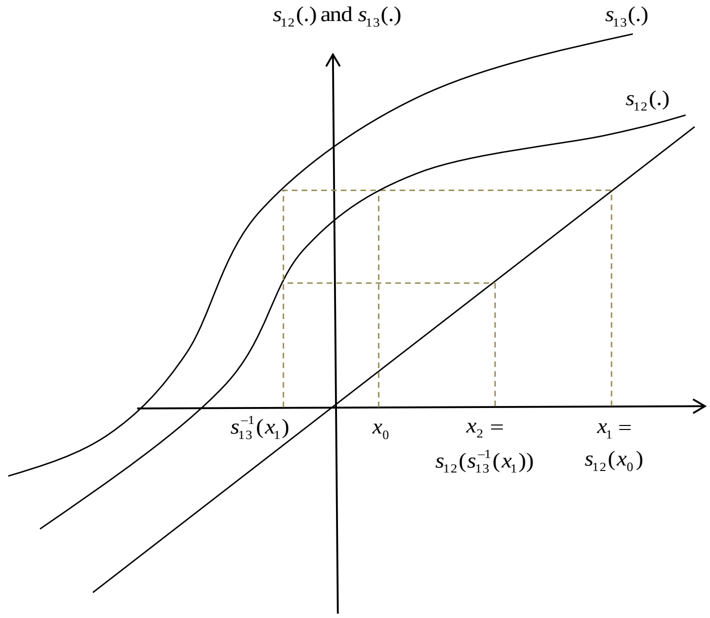

Figure 2: Some points in the orbit of x_0 under the freeness property, $K \geq 3$.

Assumption 5 *(Regularity and non-periodicity) There exists* $(i,j,k) \in \{1,...,K\}^3$ *such that* $h(i,.), h(j,.)$ *and* $h(k,.)$ *are* C^2 *diffeomorphisms and for all* $(m,n) \in \mathbb{Z}^2$, $(m,n) \neq (0,0)$, $s_{ij}^m \neq s_{ik}^n$.

The regularity condition is necessary to avoid complications due to irregularities in the derivatives of $h(z,.)$.[11] The non-periodicity assumption, on the other hand, is a rank condition. It states that the effect of moving from $Z = i$ to $Z = j$ is "truly" different from the effect of a shift from $Z = i$ to $Z = k$. If non-periodicity fails to hold, so that for any $(i,j,k) \in \{1,...,K\}^2$ there exists (m,n) such that $s_{ij}^m = s_{ik}^n$, we are actually back to the binary case and g is identified on a sequence of points only. It can be proved indeed that the group $\mathcal{S}$ is generated by a unique s, which can be interpreted as the effect of a (new) binary instrument. In the generalized location models considered above where we set, without loss of generality, $\nu(1) = 0$, non-periodicity holds if there exists $j \in \{2,...,K\}$ such that $\nu(j)/\nu(2) \notin \mathbb{Q}$.

As previously mentioned, a function s in $\mathcal{S}$ can be interpreted as the effect of a new instrument Z_s. Assumption 5 ensures that we will be able to build instruments that have

[11]The orbits may neither be discrete nor dense, but related to Cantor sets (see, e.g., Ghys, 2001, Proposition 5.6).

infinitesimal effect in X. It is thus as if we were able to construct a continuous instrument from the discrete observed ones, which will insure full identification of the model. Formally, Theorem 4.3 below shows that if Assumption 5 holds, the orbit $\mathcal{O}_{x_0}$ is dense in the support of X. g is then fully identified by Lemma 3.2.

Theorem 4.3 *If $K \geq 3$ and if Assumptions 1-5 and the freeness property hold, g is fully identified on $\mathcal{X}$.*

The proof of Theorem 4.3 relies on Hölder's and Denjoy's theorems, two deep results in group and dynamical systems theories. By Denjoy's theorem, in particular, $\mathcal{O}_{x_0}$ is either discrete or dense in the support of X, depending on whether a scalar called the *rotation number* is rational or not. Assumption 5 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 $g(.,.)$. If $K \geq 4$ and Assumption 5 holds for four indices or more (say $1,2,3,4$), we can identify $g(.,.)$ by using different subsets of this set ($1,2,3$ and $2,3,4$ for instance). If the model is not true, the functions that we recover using these different subsets do not coincide in general. This means that the model is overidentified (testable) in general when $K \geq 4$.

While in the binary case, freeness is equivalent to a stochastic dominance condition, there is no such simple characterization when $K \geq 3$. To test for it, we can rely on the fact that, by Hölder's theorem, every elements of $\mathcal{S}$ commute. Hence, if $K = 3$ for instance, we can test for the simpler condition $s_{12} \circ s_{13} = s_{13} \circ s_{12}$.

4.2 Identification under nonfreeness

Studying, as under the freeness property, the topology of the orbit $\mathcal{O}_{x_0}$ can be interesting. In particular, when $K \geq 3$, we can combine the functions s_{ij} to create "new" instruments. As soon as these instruments induce infinitesimal changes in X, we will obtain full identification. There is however a more direct approach under nonfreeness because infinitesimal changes are observed around the fixed points and because all points can be linked with these fixed points. This allows us to connect different orbits together and to obtain full identification directly, even when $K = 2$.

To illustrate this idea, let consider a function s that has only one fixed point x_f, with $s(x) > x$ for $x < x_f$ and $s(x) < x$ otherwise (see Figure 3). For any value of x, $x_f = \lim_{k\to\infty} s^k(x)$, and $x_f \in \overline{\mathcal{O}}_x$. Given x_0, $g(., u)$ is identified at $s^k(x_0)$ and, by continuity, at

x_f. Using Lemma 3.3, we can identify $g(x, u)$ for all x because $x_f \in \overline{\mathcal{O}}_x$. The model is thus fully identified.

Theorem 4.4 *If Assumptions 1-4 and the nonfreeness property hold, g is fully identified on $\mathcal{X}$.*

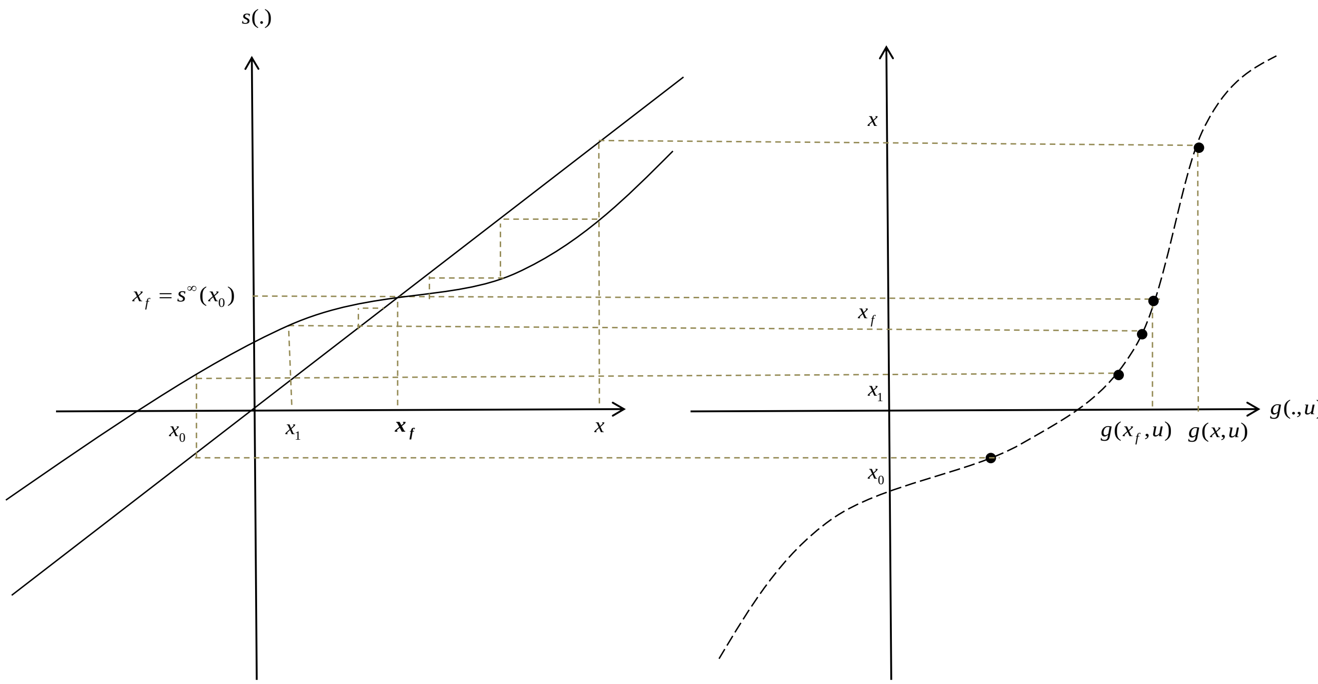

Figure 3: Full identification under nonfreeness.

This theorem 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 homogenous effect of Z discussed above fails to hold. The intuition conveyed by the result of Imbens and Angrist (1994) that an homogenous effect helps for identification is actually reversed. Our result is, on the other hand, in lines with Theorem 2 of Hoderlein and Sasaki (2011b). Within Model (2.1) and under monotonicity of h but not g, they show that some marginal effects are point identified when $F_{X|Z=i}$ and $F_{X|Z=j}$ cross for some (i, j), while they are only partially identified otherwise.

Corollary 4.5 *Suppose that there exists $(i, j) \in \{1, ..., K\}^2$ such 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 g is fully identified on $\mathcal{X}$.*

This "crossing case" is studied by Torgovitsky (2011), who also shows, in a closely related paper, that the model is fully identified thanks to crossing points.[12] 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.

If, in the case of a binary instrument, the "crossing" condition is equivalent to nonfreeness, this is not the case anymore when $K \geq 3$. As previously explained, reasoning on $\mathcal{S}$ is then important because non freeness 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, meaning that none of the s_{ij} $(i \neq j)$ admits a fixed point (see the supplementary material for an example).

5 The multivariate case

In the multivariate case, the topology of the orbits is more complicated than before and a full classification is difficult to obtain. Yet, Lemmas 3.2 and 3.3 are still valid and previous ideas can be partially extended.

5.1 Identification under freeness

Let us suppose first that the freeness property holds. When $K = 2$, we get a similar result as in Theorem 4.1: $g(., u)$ is identified on the sequence $\{s_{12}^k(x_0), k \in \mathbb{Z}\}$ only.

The case $K \geq 3$ is far more delicate. In particular, the powerful tools that we used in the univariate case, namely Hölder's and Denjoy's theorems, do not apply anymore. Hölder's theorem states that if freeness holds for a group of functions on the real line, all functions of this group commute. Thanks to this property, we can reduce our study to the unit circle. This result does not hold however for functions of several variables.[13] Moreover, in the multivariate case, even if we were able to come back to the unit circle on each coordinate, Denjoy's theorem would only prove density on each of these coordinates but not on the cartesian product of these unit circles, which would be necessary to establish full identification.

[12]Torgovitsky (2011) strengthens Assumption 4 by supposing that the support of X is bounded in at least one direction and by imposing that $g(., u)$ 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. This assumption however rules out important models such as the linear one.

[13]We provide a counterexample in the supplementary material.

Still the generalized location model provides some interesting insights. Suppose, as in the univariate case, that

$$X_m = \mu_m \left(\nu_m(Z) + \eta_m\right), \; m = 1...d \tag{5.1}$$

where μ_m is strictly increasing and continuous and, without loss of generality, $\nu_1(1) = ... = \nu_d(1) = 0$. We let hereafter A denote the $K-1 \times d$-matrix of typical $(k-1, m)$ element $\nu_m(k)$, for $k = 2, ..., K$, and A_k the kth line of A. We make the following assumption.

Assumption 6 *(Rank and non-periodicity condition) (i) The matrix A has rank d and (ii) supposing without loss of generality that $(A_1, ..., A_d)$ are linearly independent, there exists $i > d$ such that $A_i = \sum_{k=1}^d \lambda_k A_k$ and for all $(c_1, ..., c_d) \in \mathbb{Z}^d$, $(c_1, ..., c_d) \neq (0, ..., 0)$, $\sum_{k=1}^d \lambda_k c_k \notin \mathbb{Z}$.*

Condition (i) is similar to the standard rank condition in linear IV models, and actually identical when $\mu_1 = ... = \mu_d = \text{Id}$, the identity function. Condition (ii) is similar to the non-periodicity condition imposed in Assumption 5 in the univariate case, and can be interpreted as another rank condition. It basically states that using a value i of the instrument, we can yield a binary instrument Z_i whose effect is truly distinct from those we can produce using the first $d+1$ values of Z. A necessary condition for Assumption 5 to hold is that $K \geq d+2$, which is logical since full identification was obtained in the univariate case with $K \geq 3$. Theorem 5.1 shows that the model is fully identified under this condition. Its proof relies on a characterization of additive subgroups of $\mathbb{R}^d$, which can be found for instance in Bourbaki (1974).

Theorem 5.1 *If Equation (5.1) and Assumptions 1-4 and 6 hold, g is fully identified on $\mathcal{X}$.*

5.2 Identification under nonfreeness

Without freeness, we can still use fixed points to achieve identification. However another element comes into play, namely the attractiveness of these fixed points. Attractiveness is not an issue in the univariate case since the functions are strictly increasing. Any fixed point of s can be reached by applying several times either s or s^{-1} and g is thus identified at the fixed point.

This is not true anymore in a multidimensional setting, as illustrated in Figure 4. Consider the bivariate case with $K = 2$, and let $x_f = (x_{1,f}, x_{2,f})$ denote a fixed point of $s_{12} =$

$(s_{1,12}, s_{2,12})$. Suppose first that $s_{1,12}(x_1) > x_1$ if and only if $x_1 < x_{1,f}$, while $s_{2,12}(x_2) < x_2$ if and only if $x_2 < x_{2,f}$ (see Figure 4, case (a)). No sequence $(s_{12}^k(x))_{k\in\mathbb{N}}$ converges in $\mathcal{X}$. When $x = (x_1, x_2) \in (-\infty, x_{1,f}) \times (-\infty, x_{2,f})$, for instance, the sequence $(s_{1,12}^k(x_1))_{k\in\mathbb{N}}$ converges to $x_{1,f}$ but the sequence $(s_{2,12}^k(x_2))_{k\in\mathbb{N}}$ tends to $-\infty$, with $(x_{1,f}, -\infty) \notin \mathcal{X}$. On the other hand, suppose that $s_{m,12}(x_m) < x_m$ if and only if $x_m < x_{m,f}$, for $m \in \{1, 2\}$ (Figure 4 case (b)). For any $x = (x_1, x_2)$, the sequence $(s_{12}^{-k}(x))_{k\in\mathbb{N}}$ converges to x_f.

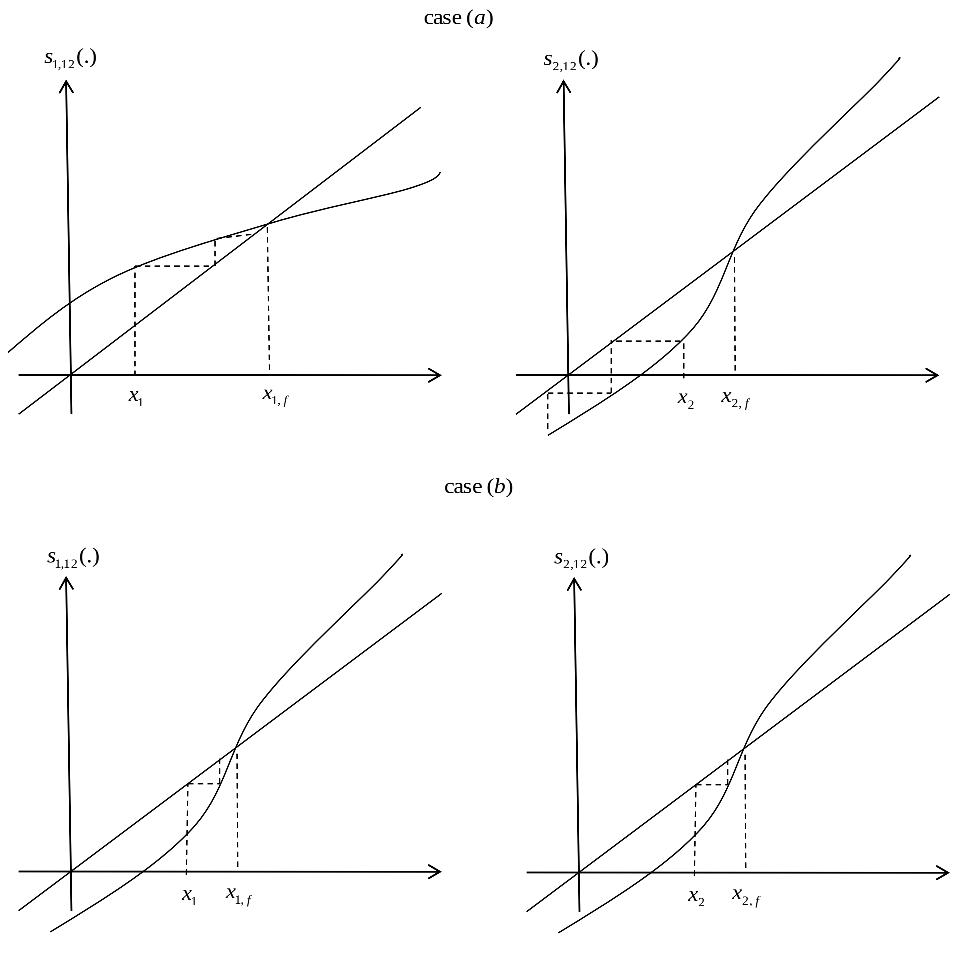

Figure 4: Illustration of the attractiveness issue under nonfreeness.

In short, a condition on the position of the coordinates of s_{12} is necessary and sufficient to secure identification when $K = d = 2$. The sufficiency 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* $s = (s_1, ..., s_d) \in \mathcal{S}$ *with exactly one fixed point* $x_f = (x_{1,f}, ..., x_{d,f})$ *and such that for all* $x = (x_1, ..., x_d)$*,* $sgn\left[(s_m(x) - x_m)(x_m - x_{m,f})\right]$ *does not depend on* $m \in \{1, ..., d\}$*, then* g *is fully identified on* $\mathcal{X}$*.*

Even if the attractiveness condition may seem restrictive, it is important to note that only one function in the group has to satisfy this condition. Hence, it may hold even when no function s_{ij} admits an attractive fixed point, because we also have at hand all the compositions of the s_{ij}. To illustrate this idea, consider the generalized location-scale models of the form

$$X_m = \mu_m \left(\nu_m(Z) + \sigma_m(Z)\eta_m\right), \tag{5.2}$$

with $\sigma_m(Z) > 0$ and μ_m a strictly increasing and continuous function. Without loss of generality, we set $\sigma_1(1) = ... = \sigma_d(1) = 1$. Unless $\sigma_m(.)$ is constant for some m, all the functions s_{ij} admit a unique fixed point, which is not attractive in general. Nevertheless, under a simple rank condition, the model is fully identified because one can always construct a function $s \in \mathcal{S}$ with an attractive fixed point.

Theorem 5.3 *If Assumptions 1-4 and Equation (5.2) hold, and the rank of the matrix of typical (i, j) element $\ln \sigma_i(j+1)$ is d, there exists $s \in \mathcal{S}$ that admits a unique and attractive fixed point. Thus, g is fully identified on $\mathcal{X}$.*

6 Concluding remarks

While the previous results show that full or partial identification can be achieved with a discrete instrument, estimation has not been addressed. Studying inference in this setting is beyond the scope of this paper, but we indicate a possible estimation method for a binary instrument, with a sample $(X_i, Y_i, Z_i)_{i=1...n}$ of independent copies of (X, Y, Z) at hand. The estimators that we propose follow our constructive identification strategy.

As the identification results differ, testing the freeness property is the first issue to solve. With a binary instrument, nonfreeness is equivalent to the existence of fixed points for s_{12}. Let $\widehat{s}_{12}(x) = \widehat{F}^{-1}_{X|Z=2} \circ \widehat{F}_{X|Z=1}(x)$, with $\widehat{F}_{X|Z=z}$ (resp. $\widehat{F}^{-1}_{X|Z=z}$) the empirical cdf (resp. quantile) of X on the subsample for which $Z = z$. A possible testing procedure is to accept nonfreeness if and only if the equation $\widehat{s}_{12}(x) = x$ admits at least a solution on a range $[\widehat{F}^{-1}_{X|Z=1}(q_{1n}), \widehat{F}^{-1}_{X|Z=1}(q_{2n})]$, with $q_{1n} \to 0$ and $q_{2n} \to 1$. This range restriction avoids picking fixed points due to the lack of accuracy of $\widehat{s}_{12}$ at the tails. We expect such a test of nonfreeness to be consistent under mild restrictions such as the fact that under the null, the sign of $x \mapsto s_{12}(x) - x$ is nonconstant.

If we accept nonfreeness, we can actually recover the whole function g. Let us suppose for simplicity that s_{12} has only one fixed point, and let x_0 denote this fixed point. We also suppose, without loss of generality that the fixed point is reached by applying s_{12}. Our normalization implies that $g(x_0,u) = F^{-1}_{Y|X=x_0}(u)$. Moreover, using Equation (3.1) and a direct induction, g writes as :

$$g(x,u) = \left[\lim_{k\to\infty} \gamma_{kx}\right]^{-1} \circ F^{-1}_{Y|X=x_0}(u).$$

where $\gamma_x = F^{-1}_{Y|X=s_{12}(x),Z=2} \circ F_{Y|X=x,Z=1}$ and $\gamma_{kx} = \gamma_{s_{12}^{k-1}(x)} \circ ... \circ \gamma_x$.[14]

We estimate γ_{kx} by

$$\widehat{\gamma}_{kx}(y) = \widehat{\gamma}_{\widehat{s}_{12}^{k-1}(x)} \circ \widehat{\gamma}_{\widehat{s}_{12}^{k-2}(x)} \circ ... \circ \widehat{\gamma}_x(y),$$

where we let

$$\widehat{\gamma}_x(y) = \widehat{F}^{-1}_{Y|X=\widehat{s}_{12}(x),Z=2} \circ \widehat{F}_{Y|X=x,Z=1}(y),$$

with $\widehat{F}_{Y|X=x,Z=z}$ (resp. $\widehat{F}^{-1}_{Y|X=x,Z=z}$) a nonparametric estimator of the conditional cdf (resp. quantile) of Y.[15] We suggest to estimate $[\lim_{k\to\infty}\gamma_{kx}]^{-1}$ by $\widehat{\gamma}^{-1}_{k_n x}$, with k_n tending to infinity at an appropriate rate. This rate should, as usually, achieve the best balance between variance (which is large for a large k_n) and bias (which is large when k_n is small). An estimator of g is finally given by

$$\widehat{g}(x,u) = \widehat{\gamma}^{-1}_{k_n x} \circ \widehat{F}^{-1}_{Y|X=\widehat{x}_0}(u),$$

where $\widehat{x}_0$ is an estimator of x_0. Possible choices include the fixed point of $\widehat{s}_{12}$ or $\widehat{s}_{12}^{k_n}(x)$, which tend to x_0 as $n\to\infty$.

If we reject the test, we adopt the freeness framework and identify g only on $(s_{12}^k(x_0))_{k\in\mathbb{Z}}$. In this case, it seems appropriate to estimate directly the average and quantile treatment effects $\Delta_{ij}^{ATE}(x)$ and $\Delta_{ij}^{QTE}(x)$.[16] Using $\Delta_{ij}^{ATE}(x) = E\left[\gamma_{jx}(Y) - \gamma_{ix}(Y)|X=x\right]$ and $\Delta_{ij}^{QTE}(x,\tau) = F^{-1}_{\gamma_{jx}(Y)|X=x}(\tau) - F^{-1}_{\gamma_{ix}(Y)|X=x}(\tau)$, we can estimate these quantities by

$$\begin{aligned}
\widehat{\Delta}_{ij}^{ATE}(x) &= \widehat{E}\left[\widehat{\gamma}_{jx}(Y) - \widehat{\gamma}_{ix}(Y)|X=x\right], \\
\widehat{\Delta}_{ij}^{QTE}(x,\tau) &= \widehat{F}^{-1}_{\widehat{\gamma}_{jx}(Y)|X=x}(\tau) - \widehat{F}^{-1}_{\widehat{\gamma}_{ix}(Y)|X=x}(\tau),
\end{aligned}$$

where $\widehat{\gamma}_{jx}$ is as before and $\widehat{E}(U|X=x)$ denotes a nonparametric estimator of $E(U|X=x)$, for any random variable U.

[14] See Equations (6.2) and (6.6) in Appendix B for a formal proof.

[15] Nonparametric quantile estimators are considered for instance by Chandhuri (1991) or Matzkin (2003).

[16] An estimator of $g(s_{12}^k(x_0),u)$ can be developed along the same lines.

Appendix A: definitions related to group theory

In this appendix, we recall some useful definitions on group theory.

A group $\mathcal{S}$ is a set endowed with a binary operator $*$ which satisfies three properties. The first is associativity: for all $(s_1, s_2, s_3) \in \mathcal{S}^3$, $(s_1 * s_2) * s_3 = s_1 * (s_2 * s_3)$. The second is the existence of an identity element $e \in \mathcal{S}$ satisfying $s * e = e * s = s$ for all $s \in \mathcal{S}$. The third is the existence of inverses. Every element $s \in \mathcal{S}$ admits an element called its inverse and denoted s^{-1} which satisfies $s * s^{-1} = s^{-1} * s = e$. The set $\mathcal{B}$ of all bijections onto $\mathcal{X}$, endowed with the composition operator, is an example of a group.

A subgroup $\mathcal{T}$ of $\mathcal{S}$ is a subset of $\mathcal{S}$ which is itself a group for $*$. If we let $(\mathcal{T}_i)_{i\in\mathcal{I}}$ denote a family of subgroups of $\mathcal{S}$, one can check that $\cap_{i\in\mathcal{I}}\mathcal{T}_i$ is also a group. The group generated by a subset I of $\mathcal{S}$ is the intersection of all subgroups of $\mathcal{S}$ containing I. By definition, it is the smallest subgroup of $\mathcal{S}$ including I. In the paper, $\mathcal{S}$ is the subgroup of $\mathcal{B}$ generated by the functions $(s_{ij})_{(i,j)\in\{1,\dots,K\}^2}$.

We also define the notion of group actions and orbits. For any set $\mathcal{A}$ and a group $\mathcal{S}$, a group action . is a function from $\mathcal{S} \times \mathcal{A}$ to $\mathcal{A}$ (denoted by $s.x$) satisfying, for every $(s_1, t) \in \mathcal{S}^2$ and $x \in \mathcal{A}$, $(s_1 * t).x = s_1.(t.x)$ and $e.x = x$. The orbit $\mathcal{O}_x$ of $x \in \mathcal{A}$ is then defined by

$$\mathcal{O}_x = \{s.x, s \in \mathcal{S}\}.$$

In the paper, the group action is $s.x = s(x)$ and the orbit of x is the set $\{s(x), s \in \mathcal{S}\}$.

Finally, a group action . is free if $s.x = x$ for some $x \in \mathcal{A}$ implies that $s = e$. This definition coincides, in the setting of the paper, with the freeness property.

Appendix B: proofs

6.1 Proof of Lemma 3.3

Fix $u \in (0,1)$. First, by Lemma 3.2, we can identify $g(.,u)$ at $x^* \in \overline{\mathcal{O}}_{x_0} \cap \overline{\mathcal{O}}_x \cap \mathcal{X}$. Second, we show by "inverting" our induction method that we can identify $g(x,u)$ using $g(x^*,u)$. To see this, note that by definition of x^*, there exists $(s_n)_{n\in\mathbb{N}}$ in the group such that $x^* = \lim_{n\to\infty} s_n(x)$. Using Equation (3.1) and a direct induction, there is an identified function Q_{nx} such that for all $n \in \mathbb{N}$,

$$g(s_n(x), u) = Q_{nx} \circ g(x, u). \tag{6.1}$$

Hence,

$$Q_{nx}(y) = g(s_n(x), g^{-1}(x, y)),$$

where $g^{-1}(x, .)$ denotes the inverse of $g(x, .)$. This proves that Q_{nx} converges to a limit $Q_{\infty x}$, which is strictly increasing as the composition of two strictly increasing function. Making (6.1) tend to infinity and composing by $Q_{\infty x}^{-1}$, we finally obtain

$$g(x, u) = Q_{\infty x}^{-1} \circ g(x^*, u). \tag{6.2}$$

Thus, $g(x,u)$ is identified. The result follows by applying once more Lemma 3.2. ■

6.2 Proof of Theorem 3.2

We have $h^{-1}(i,x) = -\nu(i) + \mu^{-1}(x)$. As a result, $s_{ij}(x) = \mu(\nu(j) - \nu(i) + \mu^{-1}(x))$. For any $s \in \mathcal{S}$, there exists $(i_1, j_1, ..., i_p, j_p) \in \{1, ..., K\}^{2p}$ such that $s = s_{i_1 j_1} \circ ... \circ s_{i_p j_p}$. By a straightforward induction, $s(x) = \mu\left(\sum_{i=1}^K \nu(i)n(i) + \mu^{-1}(x)\right)$, where $n(i) = \sum_{l=1}^p \mathbb{1}\{j_l = i\} - \mathbb{1}\{i_l = i\}$. $s(x) = x$ thus implies that $\sum_{i=1}^K \nu(i)n(i) = 0$, implying in turn that s is the identity function. The result follows. ■

6.3 Proof of Theorem 4.1

The "if" part follows directly from Lemma 3.2, because $\mathcal{O}_{x_0} = \{x_k : k \in \mathbb{Z}\}$. Now let us turn to the "only if" part, by showing that $g(\widetilde{x}, .)$ is not identified for any given $\widetilde{x} \notin \mathcal{O}_{x_0}$. For that purpose, we show that for any arbitrary strictly increasing function q, we can define $\widetilde{g}$, $\widetilde{Y}$ and $\widetilde{\varepsilon}$ such that

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

(ii) $\widetilde{Y} = \widetilde{g}(x,\widetilde{\varepsilon})$ 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 $\widetilde{Y}, X, Z, \eta, \widetilde{\varepsilon}$ and $\widetilde{g}$.

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

$$\lim_{x\to x_{k+1}, u'\to u} \widetilde{g}(x,u') = g(x_{k+1},u). \tag{6.3}$$

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

$$\begin{aligned}\widetilde{g}(s_{12}(x),u) &= F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}(\widetilde{g}(x,u)), \\ \widetilde{g}(s_{12}^{-1}(x),u) &= F^{-1}_{Y|X=x,Z=i} \circ F_{Y|X=s_{ij}(x),Z=j}(\widetilde{g}(x,u)).\end{aligned} \tag{6.4}$$

By a straightforward induction, $\widetilde{g}(x,u) = g(x,u)$ for all $x \in \mathcal{O}_{x_0}$, and (i) is satisfied. Moreover, the function $\widetilde{g}(x,.)$ is strictly increasing. Indeed, it is strictly increasing on $[x_k, x_{k+1})$ and $F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}$ is strictly increasing.

We now define $(\widetilde{Y},\widetilde{\varepsilon})$ such that (ii) holds. Consider a random variable $\widetilde{\varepsilon}$ independent of Z conditional on η and such that for all $(j,x,u) \in \{1,...,K\} \times \mathcal{X} \times \text{Support}(\varepsilon)$,

$$F_{\widetilde{\varepsilon}|\eta=h^{-1}(j,x)}(u) = F_{\varepsilon|\eta=h^{-1}(j,x)}(g^{-1}(x,\widetilde{g}(x,u))). \tag{6.5}$$

Letting for all $x \in \mathcal{X}$, $\widetilde{Y} = \widetilde{g}(x,\widetilde{\varepsilon})$, we have

$$\begin{aligned} P(\widetilde{Y} \le y|X=x,Z=j) &= P(\widetilde{g}(x,\widetilde{\varepsilon}) \le y|\eta = h^{-1}(j,x), Z=j) \\ &= P(\widetilde{\varepsilon} \le \widetilde{g}^{-1}(x,y)|\eta = h^{-1}(j,x), Z=j) \\ &= P(\widetilde{\varepsilon} \le \widetilde{g}^{-1}(x,y)|\eta = h^{-1}(j,x)) \\ &= F_{\varepsilon|\eta=h^{-1}(j,x)}(g^{-1}(x,\widetilde{g}(x,\widetilde{g}^{-1}(x,y)))) \\ &= F_{\varepsilon|\eta=h^{-1}(j,x),Z=j}(g^{-1}(x,y)) \\ &= F_{Y|X=x,Z=j}(y), \end{aligned}$$

and (ii) is satisfied.

Finally, let us prove (iii). First, by construction, $\widetilde{\varepsilon} \perp\!\!\!\perp Z|\eta$. Because $\eta \perp\!\!\!\perp Z$, one can check easily that $(\widetilde{\varepsilon},\eta) \perp\!\!\!\perp Z$, and Assumption 1 is satisfied. Second, we already saw that the function $\widetilde{g}(x,.)$ is strictly increasing. Thus Assumption 2 holds. Third, we have to verify that Assumption 4, (ii) and (iv) hold, (i) and (iii) being automatically satisfied. $\widetilde{g}$ is continuous on $[x_k, x_{k+1}) \times (0,1)$. We have $\widetilde{g}(x_{k+1}, u) = g(x_{k+1}, u)$, so that by (6.3), $\widetilde{g}$ is continuous at (x_{k+1}, u) for all $u \in (0,1)$ and when coming on the left with x. To prove that $\widetilde{g}(.,.)$ is continuous on $\mathcal{X} \times (0,1)$, it suffices, by (6.4), to prove that $F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}$ is continuous. First, reasoning as in Subsection 3.1,

$$F_{Y|X=x,Z=i}(y) = F_{\varepsilon|\eta=h^{-1}(i,x)}(g^{-1}(x,y)),$$

where $g^{-1}(x,.)$ denotes the inverse of $g(x,.)$. Thus, by Assumption 4, (i) and (iv), $(x,y) \mapsto F_{Y|X=x,Z=i}(y)$ is continuous. Similarly, $(x,\tau) \mapsto F^{-1}_{Y|X=s_{ij}(x),Z=j}(\tau)$ is continuous. As a result, $\widetilde{g}(.,.)$ is continuous on $\mathcal{X} \times (0,1)$ and Assumption 4, (iv) holds. Finally, the function $(u,v) \mapsto F_{\varepsilon|\eta=v}(g^{-1}(x, \widetilde{g}(x,u)))$ is continuous and strictly increasing in u as a composition of continuous and strictly increasing functions. Thus, by (6.5), $F_{\widetilde{\varepsilon}|\eta}$ satisfies Assumption 4, (ii). The result follows. ■

6.4 Proof of Theorem 4.2

Let $\gamma_x = F^{-1}_{Y|X=s_{12}(x),Z=2} \circ F_{Y|X=x,Z=1}$ and

$$\gamma_{jx} = \gamma_{s_{12}^{j-1}(x)} \circ ... \circ \gamma_x \tag{6.6}$$

for all $j > 0$. γ_{jx} with $j < 0$ is defined similarly by replacing s_{12} by s_{21}, and γ_{0x} =Id. The functions γ_{jx} are identified. By a repeated use of Equation (3.1), we get $g(s_{12}^j(x), \varepsilon) = \gamma_{jx}(g(x,\varepsilon))$. Hence, for all x, the treatment effects are identified by

$$\begin{aligned}
\Delta_{ij}^{ATE}(x) &= E\left[\gamma_{jx}(Y) - \gamma_{ix}(Y)|X=x\right] && (6.7)\\
\Delta_{ij}^{QTE}(x,\tau) &= F^{-1}_{\gamma_{jx}(Y)|X=x}(\tau) - F^{-1}_{\gamma_{ix}(Y)|X=x}(\tau) \quad ■ && (6.8)
\end{aligned}$$

6.5 Proof of Theorem 4.3

Let us first provide an informal outline of the proof. Without loss of generality, we set the indices (i,j,k) defined in Assumption 5 to $1,2,3$. We show the result by proving in four steps that the orbit $\mathcal{O}'_{x_0}$ of x_0 relative to the group generated by s_{12} and s_{13} is dense.[17] The first two steps consist of transforming the problem in order to use Denjoy's theorem, which applies to a single mapping on the unit circle instead of two functions on $\mathcal{X}$. First,

[17]The orbit $\mathcal{O}'_{x_0}$ is included in the orbit $\mathcal{O}_{x_0}$ because we consider the subgroup generated by s_{12} and s_{13} only and not $\mathcal{S}$.

we show that s_{12} can be "transformed" into the translation $t(x) = x+1$ on $\mathbb{R}$, which means that there is an increasing smooth bijection r from $\mathbb{R}$ to $\mathcal{X}$ such that $s_{12} = r \circ t \circ r^{-1}$. We then consider $f = r^{-1} \circ s_{13} \circ r$ instead of s_{13}. In the second step, we prove that we can define a transformation of f, $\widetilde{f}$, on the unit circle $[0, 1)$. In this step we use the fact that s_{12} and s_{13} commute, by Hölder's theorem. In the third step, we show that we can use Denjoy's theorem on $\widetilde{f}$, implying that orbits of $\widetilde{f}$ on the unit circle are dense. Finally, in the fourth step, we show the density of $\mathcal{O}'_{x_0}$ by, basically, "unrolling" the unit circle through successive applications of the translation.

1. s_{12} can be "transformed" into the translation t.

s_{12} does not admit any fixed point. Suppose without loss of generality that $s_{12}(x) > x$ (otherwise it suffices to consider $x \mapsto x - 1$ instead of $t(.)$). By Assumption 5, $s_{12} = r(2, .) \circ r(1, .)^{-1}$ is a C^2 diffeomorphism on $(\underline{x}, \overline{x})$. We prove that there exists an increasing C^2 diffeomorphism r from $\mathbb{R}$ to $(\underline{x}, \overline{x})$ such that $s_{12} = r \circ t \circ r^{-1}$. Let us consider an increasing C^2 diffeomorphism $\widetilde{r}$ defined on $[0, 1)$ such that $\widetilde{r}(0) > \underline{x}$, $\lim_{x\to 1} \widetilde{r}(x) = s_{12}\circ\widetilde{r}(0)$, $\lim_{x\to 1} \widetilde{r}'(x) = [s_{12} \circ \widetilde{r}]'(0)$ and $\lim_{x\to 1} \widetilde{r}''(x) = [s_{12} \circ \widetilde{r}]''(0)$. Such a $\widetilde{r}$ exists. Then define the function r by $r = \widetilde{r}$ on $[0, 1)$ and extend it on the real line, using $r(x + 1) = s_{12} \circ r(x)$ or $r(x) = s_{12}^{-1} \circ r(x + 1)$. By construction, r is strictly increasing and C^2. Hence, it admits a limit at $-\infty$ and $+\infty$. Suppose that $\lim_{x\to-\infty} r(x) = M > \underline{x}$. Because $r(x+1) = s_{12}\circ r(x)$, we would have $s_{12}(M) = M$, a contradiction. Thus, $\lim_{x\to-\infty} r(x) = \underline{x}$. Similarly, $\lim_{x\to+\infty} r(x) = \overline{x}$. Consequently, r is a C^2 diffeomorphism from $\mathbb{R}$ to $(\underline{x}, \overline{x})$.

2. We can define a transformation $\widetilde{f}$ of $f = r^{-1} \circ s_{13} \circ r$ on the unit circle.

Because freeness holds, by a theorem of Hölder (see, e.g., Ghys, 2001, Theorem 6.10), s_{12} and s_{13} commute. This implies that for all $x \in \mathbb{R}$,

$$f(x+1) = f\circ t(x) = r^{-1}\circ s_{13}\circ r\circ r^{-1}\circ s_{12}\circ r = r^{-1}\circ s_{12}\circ r\circ r^{-1}\circ s_{13}\circ r = t\circ f(x) = f(x)+1.$$

As a result, letting $\pi(x)$ denote the fractional part of x,

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

This implies that there exists a function $\widetilde{f}$ on the unit circle $[0, 1)$ defined by $\widetilde{f} \circ \pi = \pi \circ f$. We also obtain that $\widetilde{f}^2 \circ \pi = \widetilde{f} \circ \pi \circ f = \pi \circ f^2$, so that, by a direct induction,

$$\widetilde{f}^n \circ \pi = \pi \circ f^n, \quad \forall n \in \mathbb{Z} \tag{6.9}$$

3. Orbits of $\widetilde{f}$ on the unit circle are dense.

Because s_{13} and r are increasing C^2 diffeomorphisms, so is f. Thus, $\widetilde{f}$ is an orientation-preserving C^2 diffeomorphism on the unit circle.[18] We can thus apply Denjoy's theorem (see, e.g., Navas, 2009, Theorem 3.1.1), and the orbits of any element $\dot{x} \in [0,1)$ of the group generated by $\widetilde{f}$ are either all finite or all dense. Suppose that they are finite. Then there exists $n \in \mathbb{Z}^*$ such that $\widetilde{f}^n(\dot{x}) = \dot{x}$ for all $\dot{x}$ on the unit circle. Let $x \in \mathbb{R}$ be such that $\pi(x) = \dot{x}$. Then, using (6.9), there exists $m \in \mathbb{Z}$ such that $f^n(x) = t^m(x)$. Hence, by definition of f and t, $s_{13}^n(x) = s_{12}^m(x)$ with $n \neq 0$, contradicting Assumption 5. We conclude that any orbit for the group generated by $\widetilde{f}$ is dense in $[0,1)$.

4. $\mathcal{O}'_{x_0}$ is dense.
First, $\mathcal{O}'_{x_0} = r\left(\mathcal{O}_{r^{-1}(x_0)}\right)$, where $\mathcal{O}_{r^{-1}(x_0)}$ denotes the orbit of $r^{-1}(x_0)$ for the group generated by f and t. Because r is continuous, it suffices to show that $\mathcal{O}_{r^{-1}(x_0)}$ is dense. For that purpose, we basically "unroll" the unit circle by successive applications of t.

Fix $y \in \mathbb{R}$ and consider a neighborhood $\mathcal{V}_y$ of y. By definition of the topology on the unit circle, $\pi(\mathcal{V}_y)$ is a neighborhood of $\pi(y)$ in the unit circle. Because the orbit of $\pi(r^{-1}(x_0))$ through $\widetilde{f}$ is dense in $[0,1)$, there exists $n \in \mathbb{Z}$ such that $\widetilde{f}^n \circ \pi(r^{-1}(x_0)) \in \pi\left(\mathcal{V}_y\right)$. Hence, using Equation 6.9, $\pi \circ f^n(r^{-1}(x_0)) \in \pi\left(\mathcal{V}_y\right)$, and there exists $m \in \mathbb{Z}$ such that $t^m \circ f^n(r^{-1}(x_0)) \in \mathcal{V}_y$. This proves that $\mathcal{O}_{r^{-1}(x_0)}$ is dense on the real line. ∎

6.6 Proof of Theorem 4.4

Let $x^1 < ... < x^M$ denote the fixed points of s, $x^0 = \underline{x}$ and $x^{M+1} = \overline{x}$ and let i be such that $x_0 \in [x^i, x^{i+1})$. Suppose for instance that $s(x) > x$ for all $x \in (x^i, x^{i+1})$ (the proof is identical if $s(x) < x$). Suppose first that $x_0 > x^i$. A straightforward induction shows that the sequence $(s^n(x_0))_{n\in\mathbb{N}}$ is increasing and bounded by x^{i+1}. Thus, it converges to $l \in (x^i, x^{i+1}]$ which satisfies $s(l) = l$, and $l = x^{i+1}$. Hence $x^{i+1} \in \overline{\mathcal{O}}_{x_0}$. Using similarly the sequence $(s^{-n}(x_0))_{n\in\mathbb{N}}$ shows that $x^i \in \overline{\mathcal{O}}_{x_0}$. Thus, by Lemma 3.2, $g(.,u)$ is identified at x^i and x^{i+1}. Applying the same reasoning to any $x \in (x^i, x^{i+1})$ establishes that $x^i \in \overline{\mathcal{O}}_x$. By Lemma 3.3, $g(.,u)$ is identified at x. In a similar way, $x^i \in \overline{\mathcal{O}}_x$ for all $x \in (x^{i-1}, x^i)$ (if $i > 0$). Hence, $g(.,u)$ is identified on $(x^{i-1}, x^i]$. By a straightforward induction, it is identified on $(\underline{x}, \overline{x})$.

Finally, suppose that $x_0 = x^i$. Then, as previously, $x_0 \in \overline{\mathcal{O}}_x$ for any $x \in (x^i, x^{i+1})$, so that

[18] A map q on the unit circle is orientation-preserving if there exists an increasing function Q on the real line such that $q \circ \pi = \pi \circ Q$ and $Q(x+1) = Q(x) + 1$.

by Lemma 3.3 once more we can identify $g(x,u)$ for any $x \in [x^i, x^{i+1}]$. We conclude as before ■

6.7 Proof of Theorem 5.1

First, as in the univariate case, functions $s \in \mathcal{S}$ take the form

$$s(x_1, ..., x_d) = \left(\mu_1 \left[\sum_{k=2}^{K} n(k)\nu_1(k) + \mu_1^{-1}(x_1) \right], ..., \mu_d \left[\sum_{k=2}^{K} n(k)\nu_d(k) + \mu_d^{-1}(x_d) \right] \right),$$

for some $n = (n(2), ..., n(K)) \in \mathbb{Z}^{K-1}$. Moreover, any $n \in \mathbb{Z}^{K-1}$ corresponds to a function $s \in \mathcal{S}$. We thus have

$$\begin{aligned} \mathcal{O}_{x_0} \quad = \quad & \left\{ \left(\mu_1 \left[\sum_{k=2}^{K} n(k)\nu_1(k) + \mu_1^{-1}(x_{01}) \right], ..., \mu_d \left[\sum_{k=2}^{K} n(k)\nu_d(k) + \mu_d^{-1}(x_{0d}) \right] \right), \right. \\ & \left. (n(2), ..., n(k)) \in \mathbb{Z}^{K-1} \right\}. \end{aligned}$$

By continuity of $\mu_1, ..., \mu_d$, it suffices to show that $H = \{\sum_{k=2}^{K} n(k)A'_{k-1},\ n(k) \in \mathbb{Z}\}$ is dense in $\mathbb{R}^d$. Because H is an additive subgroup of $\mathbb{R}^d$, it suffices to show (see, e.g., Bourbaki, 1974, paragraph 1, n°3) that

$$< H, x > \subset \mathbb{Z} \Longrightarrow\ x = 0, \tag{6.10}$$

where for any $x \in \mathbb{R}^d$,

$$< H, x >= \{h'x,\ h \in H\} = \left\{ \sum_{k=2}^{K} n(k)A_{k-1}x,\ n(k) \in \mathbb{Z} \right\}.$$

Suppose that $< x, H > \subset \mathbb{Z}$ for some $x \in \mathbb{R}^d$. Then $A_k x \in \mathbb{Z}$ for all $k = 1, ..., d$. Choosing $i > d+1$ as in Assumption 6, we also have $A_i x \in \mathbb{Z}$. This implies that $\sum_{k=1}^{d} \lambda_k(A_k x) \in \mathbb{Z}$. Because $A_k x \in \mathbb{Z}$, $A_k x = 0$ for $k = 1, ..., d$ by Assumption 6. Because $(A_1, ..., A_d)$ are linearly independent, $x = 0$, implying (6.10) ■

6.8 Proof of Theorem 5.2

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

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

6.9 Proof of Theorem 5.3

First, some algebra shows that functions $s \in \mathcal{S}$ take the form

$$s(x_1, ..., x_d) = \left(\mu_1\left[\alpha_1 + \left(\prod_{k=2}^{K}\sigma_1(k)^{e_k}\right)\mu_1^{-1}(x_1)\right], ..., \mu_d\left[\alpha_d + \left(\prod_{k=2}^{K}\sigma_d(k)^{e_k}\right)\mu_d^{-1}(x_d)\right]\right),$$

for some $(\alpha_1, ..., \alpha_d) \in \mathbb{R}^d$ and $(e_2, ..., e_K) \in \mathbb{Z}^{K-1}$. Moreover, any $e \in \mathbb{Z}^{K-1}$ corresponds to a function $s \in \mathcal{S}$.

Noting $\beta_m = \prod_{k=2}^{K}\sigma_m(k)^{e_k}$, the function s admits a unique attractive fixed point if, for all m, $0 < \beta_m < 1$. Indeed $\mu_m(\alpha_m + \beta_m\mu_m^{-1}(x_{m,f})) = x_{m,f}$ if and only if $\mu_m^{-1}(x_{m,f}) = \frac{\alpha_m}{1-\beta_m}$. Moreover, $\mu_m(\alpha + \beta\mu_m^{-1}(x_m)) > x_m$ for $x_m < x_{m,f}$. Thus, by Theorem 5.2, it suffices to show that there exists $(e_2, ..., e_K) \in \mathbb{Z}^{K-1}$ such that

$$\left(\prod_{k=2}^{K}\sigma_m(k)^{e_k}\right) < 1 \text{ for all } m \in \{1, ..., d\}. \tag{6.11}$$

Let M denote the $d \times K - 1$ matrix of typical (i, j) element $\ln\sigma_i(j + 1)$. Because M is full rank by assumption, there exists $u \in \mathbb{R}^{K-1}$ such that $Mu = (1, ..., 1)'$. Thus, by density of $\mathbb{Q}^{K-1}$, there exists $\widetilde{u} \in \mathbb{Q}^{K-1}$ such that $M\widetilde{u} < 0$ (where the inequality should be understood componentwise). Moreover, $\widetilde{u}$ can be written $(e_2/D, ..., e_K/D)'$, where $(e_2, ...e_K, D) \in \mathbb{Z}^K$. This implies that $M(e_2, ...e_K)' < 0$ which is equivalent to (6.11). ■

References

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

Chandhuri, A. (1991), 'Global nonparametric estimation of conditional quantile functions and their derivatives', *Journal of Multivariate Analysis* **39**, 246–269.

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

Chernozhukov, V., Imbens, G. and Newey, W. (2007), 'Instrumental variable estimation of nonseparable models', *Journal of Econometrics* **139**, 4–14.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–1441.

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

Chesher, A., Rosen, A. and Smolinski, K. (2011), An instrumental variable model of multiple discrete choice. cemmap working paper CWP06/11.

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 Sasaki, Y. (2011*a*), Necessary and sufficient conditions for identification of causal effects in endogenous nonseparable models. Working paper.

Hoderlein, S. and Sasaki, Y. (2011*b*), On the role of time in nonseparable panel data models. Working paper.

Imbens, G. W. (2007), Nonadditive models with endogenous regressors, *in* R. Blundell and W. N. abd T. Persson, eds, 'Advances in Economics and Econometrics, Ninth World Congress of the Econometric Socieity', Vol. III, Econometric Society Monographs.

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.

Matzkin, R. L. (2003), 'Nonparametric estimation of nonadditive random functions', *Econometrica* **71**, 1339–1375.

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. Working paper.

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.

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