Identification of Nonseparable Triangular Models with Discrete Instruments *

Xavier D'Haultfœuille[†] Philippe Février[‡]

December 2014

Abstract

We study the identification through instruments of a nonseparable function that relates a continuous outcome to a continuous endogenous variable. Using group and dynamical systems theories, we show that full identification can be achieved under strong exogeneity of the instrument and a dual monotonicity condition, even if the instrument is discrete. When identified, the model is also testable. Our results therefore highlight the identifying power of strong exogeneity when combined with monotonicity restrictions.

JEL: C14.

Keywords: nonparametric identification, discrete instrument, strong exogeneity, monotonicity.

*We thank two co-editors and six anonymous referees for their remarks. We have benefited from discussions with Ivan Canay, Clément de Chaisemartin, Sylvain Chabé-Ferret, Andrew Chesher, Pierre Février, Etienne Ghys, James Heckman, Toru Kitagawa, Manasa Patnam, Julio Rebelo, Jean-Marc Robin, Susanne Schennach, Azeem Shaikh, Elie Tamer and Edward Vytlacil. We also thank participants of the CEMMAP, Chicago, Columbia, Johns Hopkins and Northwestern seminars.

[†]CREST, xavier.dhaultfoeuille@ensae.fr

[‡]CREST, fevrier@ensae.fr

1 Introduction

We consider, in this paper, the identification of a triangular nonseparable model that takes the form:

$$\begin{aligned} Y &= g(X, \varepsilon) \\ X &= h(Z, \eta), \end{aligned} \tag{1.1}$$

where $Y \in \mathbb{R}$ denotes the outcome, $X \in \mathbb{R}$ is a continuous endogenous variable and Z is the instrument. Models of this type are also considered by, e.g., Chernozhukov & Hansen (2005), Florens et al. (2008), Imbens & Newey (2009) and Torgovitsky (2014). We focus in particular on the identification of g from the distribution of (X, Y, Z).

Our main result is that if Z is independent of (ε, η) and both $g(x, .)$ and $h(z, .)$ are strictly monotonic, then g can be identified even if the instrument is discrete.[1] Specifically, we prove, relying on group and dynamical systems theories, that unless the instrument is binary and Z has a strictly monotonic effect on X, g is fully identified under mild regularity conditions, and the model is testable. This result contrasts with those of previous papers also considering models with continuous endogenous regressors, which establish identification with continuous instruments only (see,e.g., Newey & Powell 2003, Chernozhukov & Hansen 2005, Florens et al. 2008, Imbens & Newey 2009, Hoderlein & Sasaki 2013). Chesher (2007) is an exception, but given his model, only some particular treatment effects can be identified with discrete instruments. Altonji & Matzkin (2005) also show that g can be identified with a discrete instrument (see their Theorem 4.1), but under a condition (Assumption 4.4 in their paper) that is different from usual exogeneity restrictions.

The main difference with the estimating equation approach followed by, e.g., Newey & Powell (2003) or Chernozhukov & Hansen (2005), is that Z is inde-

[1] Our analysis also applies when the instrument is continuous, but the approach is more straightforward in this case.

pendent of both error terms, not only of ε. The key difference with the control function approach considered, among others, by Chesher (2003) and Imbens & Newey (2009), is that we also impose monotonicity of $g(x, .)$. Our results, therefore, emphasize the important identifying power of combining the two approaches. Taken together, our conditions imply either that heterogeneity is univariate or that it can be aggregated in a single dimension. Such restrictions, although appealing for some applications, rule out important frameworks such as random coefficient models or simultaneous equations. We refer to Imbens (2007) and Kasy (2011) for a thorough discussion of these issues, as well as to Kasy (2014) for a positive identification result without unidimensional heterogeneity.

This note is organized as follows. The second section presents the model, while the third section displays the main identification results. The proofs are gathered in the appendix. The supplementary materials provides some additional details on the link with group theory and consider the case of a multivariate X.

2 The model

Our first assumptions are the strong exogeneity of the instrument and a dual monotonicity condition. We denote by $\mathcal{X}$ the interior of the support of X and by $\{1, ..., K\}$, $K \geq 2$, the support of Z.

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

Assumption 2 *(Dual strict monotonicity)* $\varepsilon \in \mathbb{R}$, $\eta \in \mathbb{R}$ *and for all* $(x, z) \in \mathcal{X} \times \{1, ..., K\}$, $\tau \mapsto g(x, \tau)$ *and* $v \mapsto h(z, v)$ *are strictly increasing.*

Assumption 1 is typically imposed when using the control function approach (see, e.g., Florens et al. 2008, Imbens & Newey 2009, Hoderlein & Sasaki 2013) but is a stronger than the condition $Z \perp\!\!\!\perp \varepsilon$ considered in the estimating equation approach followed by Chernozhukov & Hansen (2005) and Chesher (2010).

Assumption 2 defines a one-to-one mapping between (X, Y) and (ε, η) for a given value of Z. This condition can be reasonable in some applications, but is not satisfied in all cases. It is violated in random coefficient models, for instance. It also rules out the setting of Chesher (2003), where η directly affects the outcome equation. Finally, it cannot handle simultaneity problems, as discussed by Imbens (2007). We also impose the following regularity conditions. Hereafter, $\mathcal{H}$ denotes the interior of the support of η while, for any random variables R, S, F_R and $F_{R|S=s}$ denote respectively the cumulative distribution functions (cdfs) of R and of R conditional on $S = s$.

Assumption 3 *(Regularity conditions)*
(i) Support$(X|Z = z) = [\underline{x}, \overline{x}]$ *with* $-\infty \leq \underline{x} < \overline{x} \leq \infty$ *independent of* z.
(ii) ε *has a uniform distribution.*
(iii) F_η *is continuous and strictly increasing on Support*(η).
(iv) $(u, v) \mapsto F_{\varepsilon|\eta=v}(u)$ *is continuous on* $(0, 1) \times \mathcal{H}$ *and* $u \mapsto F_{\varepsilon|\eta=v}(u)$ *is strictly increasing on* $(0, 1)$ *for all* $v \in \mathcal{H}$.
(v) $g(., .)$ *and* $h(z, .)$ *are continuous on* $\mathcal{X} \times (0, 1)$ *and* $\mathcal{H}$ *respectively.*

Condition (i) allows the support of X conditional on $Z = z$ to be either bounded or unbounded, but restricts it to be independent of z. This is theoretically testable and, as shown in the supplementary materials, several of our results still hold if the intersections of conditional supports are large enough. Condition (ii) is the usual normalization in nonseparable models (see, e.g., Matzkin 2003). The continuity conditions in (iii) and (iv) ensure that X and Y conditional on (X, Z) are continuously distributed. Condition (iv) also implies that the support of $\varepsilon|\eta = v$ does not depend on v. Combined with Assumptions 1 and 2, it implies the testable restriction that the support of $Y|X = x, Z = z$ does not depend on z. Without such a condition, we could still point identify some treatment effects similar to those considered by Chesher (2007), but part of the proof of Theorem 1 below does not apply anymore. Whether we could still identify g in such cases remains unclear to us. Finally, Condition (v) excludes discontinuous effects of X on Y. This is necessary here as we obtain identification of $g(., \tau)$ by continuity arguments. Note that Condition (v) imposes continuity of $g(., \tau)$ on the interior $\mathcal{X}$ of the support of X, but not on the whole support. This is important, in order to include models where $g(., \tau)$ tends to infinity at the boundaries.

Our paper is closely related to independent work by Torgovitsky (2014). He imposes conditions similar to Assumptions 1-3. However, in contrast to our work, he does not require that Support$(X|Z = z)$ is independent of z. He imposes rather that this support is bounded either from above or below. He also slightly reinforces Assumptions 2 and 3-(v) by imposing strict monotonicity of $g(\underline{x}, .)$ and $g(\overline{x}, .)$, and continuity of $g(., \tau)$ on the boundaries of Support(X). Under his assumptions, the conditional cdfs $(F_{X|Z=z})_{z\in \text{Support}(Z)}$ always cross, and the crossing points can be used for identification. Using group theory, we show that identification can be achieved even if these cdfs do not cross, or without using crossing points that occur at the boundary of the support of X. Finally, we show that if identified, the model is also testable.

3 Identification results

To derive identification of the model, we want to show that for any $(x, x', \tau) \in \mathcal{X}^2 \times (0, 1)$, there exists an identified function $Q_{x'x}$ such that

$$g(x', \tau) = Q_{x'x} \circ g(x, \tau) \tag{3.1}$$

for all $(x, x', \tau) \in (0, 1)$. This relation together with the normalization that ε is uniform will ensure that g is identified.

To establish (3.1), we use exogenous changes in X obtained by moving Z while keeping η constant. We define, for all $(x, i, j) \in \mathcal{X} \times \{1, ..., K\}^2$,

$$s_{ij}(x) = F^{-1}_{X|Z=j} \circ F_{X|Z=i}(x).$$

It can be shown that $s_{ij}(x) - x$ corresponds to the shift in X when Z moves from i to j while keeping η constant and equal to $h^{-1}(i, x)$.[2] Observing the effect of such a change on Y, and using the dual monotonicity condition, we can show (see the proof of Theorem 1) that

$$g(s_{ij}(x), \tau) = \left[F^{-1}_{Y|X=s_{ij}(x), Z=j} \circ F_{Y|X=x, Z=i} \right] \circ g(x, \tau). \tag{3.2}$$

This ensures that (3.1) holds for all $x \in \mathcal{X} \times (0,1)$, but only for $x' = s_{ij}(x)$, $(i, j) \in \{1, ..., K\}^2$. By induction, however. $g(s_{kl} \circ s_{ij}(x), \tau)$ is a known function of $g(s_{ij}(x), \tau)$, for instance, and thus a known function of x. Thus, (3.1) actually holds for all $x \in \mathcal{X} \times (0,1)$ and all $x' = s_{kl} \circ s_{ij}(x)$, $(i, j, k, l) \in \{1, ..., K\}^4$. As a result, by a straightforward induction, (3.1) holds for all $x \in \mathcal{X} \times (0,1)$ and all x' that are compositions of the $(s_{ij})_{(i,j) \in \{1,...,K\}^2}$ taken at x. Let $\mathcal{S}$ denote this set of functions and $\mathcal{O}_x = \{s(x) : s \in \mathcal{S}\}$. $\mathcal{S}$ and $\mathcal{O}_x$ are called respectively the group generated by the $(s_{ij})_{(i,j) \in \{1,...,K\}^2}$ and the orbit of x.[3] It is clear that if $\mathcal{O}_x = \mathcal{X}$ for all x, the model is identified. Using continuity arguments and the strict monotonicity of $g(x, .)$, we actually prove the following stronger result.

Theorem 1 *Suppose that Assumptions 1-3 hold. Then g is identified if for any $(x, x') \in \mathcal{X}^2$, there exists $x_1, ..., x_J$ such that $\overline{\mathcal{O}}_{x_j} \cap \overline{\mathcal{O}}_{x_{j+1}} \neq \emptyset$ for $j = 0, ..., J$, where $\overline{\mathcal{O}}_x$ denotes the closure of $\mathcal{O}_x$ in $\mathcal{X}$ and $x_0 = x$, $x_{J+1} = x'$.*

Theorem 1 establishes identification under a condition on the orbits. We now show that this condition is satisfied in many cases. We introduce for that purpose the freeness and nonfreeness properties.

Definition 1 *The freeness property holds if there exists no $s \in \mathcal{S}$ different from the identity function that admits a fixed point. The nonfreeness property holds if there exists $s \in \mathcal{S}$ different from the identity function that admits a positive and finite number of fixed points.*

[2] s_{ij} corresponds to the function π in Torgovitsky (2014).

[3] Though not needed for the reading of the paper, we refer to Section 1 of the Supplementary Materials for definitions on group theory.

Whether the freeness or nonfreeness properties holds depends on the way the instrument affects X. With a binary instrument ($K = 2$), freeness is equivalent to either $h(1, .) > h(2, .)$, $h(2, .) > h(1, .)$ or $h(1, .) = h(2, .)$, the latter being uninformative. Intuitively, it can be interpreted as an homogeneity of the instrument, as all individuals either react positively or negatively to it. It can also be seen as an extension, in a continuous setting, of the monotonicity condition considered by Imbens & Angrist (1994) for dummy endogenous variables.[4] When $K \geq 3$, freeness can still be interpreted as an homogenous effect of the instrument Z on X. Any succession of exogenous shifts (corresponding to $s \in \mathcal{S}$) should either have a strictly positive, negative or null effect on all individuals. Generalized location models $h(Z, \eta) = \mu(\nu(Z) + \eta)$ (where Support$(\eta) = \mathbb{R}$) are examples of first stage equations satisfying this restriction. On the other hand, nonfreeness holds when, roughly speaking, the instrument has an heterogenous effect on X. It holds in particular when $F_{X|Z=i}$ and $F_{X|Z=j}$ cross. But when $K \geq 3$, nonfreeness may hold even if none of the functions $F_{X|Z=i}$ and $F_{X|Z=j}$ cross. We refer to the supplementary materials for a longer discussion on freeness and nonfreeness.

Our main result is that the condition on the orbits is satisfied under nonfreeness, or if freeness and the following regularity and rank condition hold.

Assumption 4 *(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* $(m, n) \in \mathbb{Z}^2$, $(m, n) \neq (0, 0)$, $s_{ij}^m \neq s_{ik}^n$.[5]

The non-periodicity condition requires that $K \geq 3$, and can be seen as 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 $Z = k$. For example, if

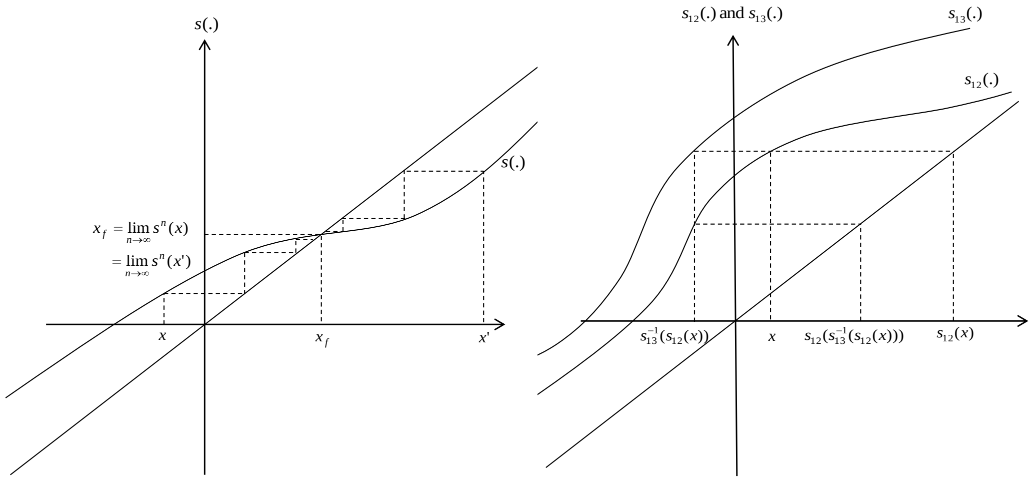
[4] The important difference with their condition, however, is that we can test it directly in the data, by checking whether $F_{X|Z=1}$ stochastically dominates (or is dominated by) $F_{X|Z=2}$ at the first-order.

[5] For any $m \in \mathbb{Z}$, we let $s_{ij}^m = s_{ij} \circ ... \circ s_{ij}$ (with $m - 1$ compositions) if $m > 0$, s_{ij}^0 =Id, the identity function on $\mathcal{X}$ and $s_{ij}^m = s_{ij}^{-1} \circ ... \circ s_{ij}^{-1}$ if $m < 0$.

$X = \nu(Z) + \eta$ with $\nu(1) = 0$ (without loss of generality), it holds if $\nu(2) \neq 0$ and $\nu(3)/\nu(2) \notin \mathbb{Q}$.

Theorem 2 *Suppose that Assumptions 1-3 hold. Then g is fully identified if either nonfreeness holds, or freeness and Assumption 4 hold. The model is also testable under these conditions.*

The arguments comprising the proofs of the free and nonfree case are quite distinct. In the nonfree case, there exists a nontrivial function $s \in \mathcal{S}$ admitting fixed points. We can then "connect" orbits and thus satisfy the condition stated in Theorem 1. The left graph of Figure 1 displays a simple example with one fixed point x_f for s. The dashed lines starting from x correspond to the sequence $(s^n(x))_{n\in\mathbb{N}}$, which belongs to $\mathcal{O}_x$. This sequence converges to x_f, which implies that $x_f \in \overline{\mathcal{O}}_x$. Similarly, $x_f \in \overline{\mathcal{O}}_{x'}$. Therefore, for any $(x, x') \in \mathcal{X}^2$, $\overline{\mathcal{O}}_x \cap \overline{\mathcal{O}}_{x'} \neq \emptyset$. Thus, the conditions of Theorem 1 are satisfied with $J = 0$. In the free case, we show that Theorem 1 can be applied with $J = 0$, by showing that any orbit is dense in $\mathcal{X}$. Consider for instance the ternary case $K = 3$. The orbit of x is the set of points that can be reached after a finite number of iterations similar to those displayed in the right graph of Figure 1. We can show that this iterative process will eventually cover the whole space under Assumption 4. Formally, this stems form Hölder's and Denjoy's theorems, two fundamental results in group and dynamical systems theories. establish that the model is testable in both cases. Note that Theorem 2 does not cover the case of a binary instrument with freeness, since Assumption 4 implies $K \geq 3$. This is a case for which Theorem 1 does not apply. The orbits are not dense and their closures do not intersect. In this case, we can show (see D'Haultfœuille & Février 2011) that the model is not identified, though some average and quantile treatment effects are still identified.



Nonfree case. Free case.

Figure 1: Intuition behind identification.

Finally, we also prove that the model is testable in both cases. The idea is the following. Consider first nonfreeness and suppose for simplicity that s_{ij} admits a fixed point x^*. Equation (3.2) then implies that $F_{Y|X=x^*,Z=i} = F_{Y|X=x^*,Z=j}$, which has no reason to hold if the model is misspecified. In the free case, by Hölder's theorem and Equation (3.2), we have, for all x and all $(i,j,k,l) \in \{1,...,K\}^4$ (see the appendix for a proof),

$$\begin{aligned}
&F^{-1}_{Y|X=s_{ij}\circ s_{kl}(x),Z=j} \circ F_{Y|X=s_{kl}(x),Z=i} \circ F^{-1}_{Y|X=s_{kl}(x),Z=l} \circ F_{Y|X=x,Z=k} \\
=&F^{-1}_{Y|X=s_{kl}\circ s_{ij}(x),Z=l} \circ F_{Y|X=s_{ij}(x),Z=k} \circ F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}.
\end{aligned}$$

Once more, this equality may not hold if Assumptions 1-3 are not satisfied, providing a basis for a test.

A Appendix: Proofs

A.1 Proof of Theorem 1

The idea is to prove that for any $(x, x', \tau) \in \mathcal{X}^2 \times (0, 1)$, there exists an identified, strictly increasing function $Q_{x'x}$ such that

$$g(x', \tau) = Q_{x'x} \circ g(x, \tau). \tag{A.1}$$

We first show that (A.1) holds when $x' \in \mathcal{O}_x$. We then prove that it actually holds when $x' \in \overline{\mathcal{O}}_x$. In a third step, we establish (A.1) for all $(x, x') \in \mathcal{X}^2$. In the last step, we prove that (A.1) ensures that g is identified.

1. (A.1) holds when $x' \in \mathcal{O}_x$.

First, following, e.g. Matzkin (2007, Section 3.3.2), we have $s_{ij}(x) = h(j, h^{-1}(i, x))$. Now,

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

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). The third equality stems from ε being independent of Z conditional on η (Assumption 1). The last equalities apply the same reasoning as the first ones, but the other way round. By Assumptions 2 and 3, $y \mapsto F_{Y|X=x,Z=i}(y)$ is strictly increasing for all (x, i). Hence, its inverse exists, and

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

In other words, there exists a strictly increasing, identified function Q_{ijx} such that $g(s_{ij}(x), \tau) = Q_{ijx} \circ g(x, \tau)$. Now suppose that $x' \in \mathcal{O}_x$. Then there exists $s_{i_1 j_1}, ..., s_{i_p j_p}$ such that $x' = s(x)$, with $s = s_{i_1 j_1} \circ ... \circ s_{i_p j_p}$. Then (A.1) holds with $Q_{x'x} = Q_{i_1 j_1 x} \circ ... \circ Q_{i_p j_p x}$.

2. (A.1) holds when $x' \in \overline{\mathcal{O}}_x$.

By definition, there exists $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{O}_x$ such that $x' = \lim_{n\to\infty} x_n$. By what precedes, we have, for any $\tau \in (0, 1)$,

$$g(x_n, \tau) = Q_{x_n x} \circ g(x, \tau). \tag{A.2}$$

Hence, for any y, $Q_{x_n x}(y) = g(x_n, g^{-1}(x, y))$, with $g^{-1}(x, .)$ the inverse of $g(x, .)$. This proves, by continuity of $g(., \tau)$, that $Q_{x_n x}$ converges to a limit $Q_{x'x}$, which is strictly increasing as the composition of two strictly increasing functions. We obtain (A.1) by making (A.2) tend to infinity.

3. (A.1) holds for all $(x, x') \in \mathcal{X}^2 \times (0, 1)$.

By assumption, there exists $x_1, ..., x_J$ such that $\overline{\mathcal{O}}_{x_j} \cap \overline{\mathcal{O}}_{x_{j+1}} \neq \emptyset$ for $j = 0, ..., J$ (with $x_0 = x$ and $x_{J+1} = x'$). Let $x_j^* \in \overline{\mathcal{O}}_{x_j} \cap \overline{\mathcal{O}}_{x_{j+1}}$. By what precedes, we have for all $j = 0, ..., J$,

$$\begin{aligned} g(x_j^*, \tau) &= Q_{x_j^* x_j} \circ g(x_j, \tau) \\ &= Q_{x_j^* x_{j+1}} \circ g(x_{j+1}, \tau). \end{aligned}$$

Because $Q_{x_j^* x_{j+1}}$ is strictly increasing, $g(x_{j+1}, \tau) = Q^{-1}_{x_j^* x_{j+1}} \circ Q_{x_j^* x_j} \circ g(x_j, \tau)$, where $Q^{-1}_{x_j^* x_{j+1}} \circ Q_{x_j^* x_j}$ is identified and strictly increasing. Thus,

$$g(x', \tau) = \left[Q^{-1}_{x_J^* x'} \circ Q_{x_J^* x_J}\right] \circ ... \circ \left[Q^{-1}_{x_0^* x_1} \circ Q_{x_0^* x}\right] \circ g(x, \tau),$$

and (A.1) holds for $(x, x', \tau) \in \mathcal{X}^2 \times (0, 1)$.

4. g is identified.

Let $G_x(u) = E\left[F_{Y|X} \circ Q_{Xx}(u)\right]$. Applying the same reasoning as in Step 1, we have, for any $(x,\tau) \in \mathcal{X} \times (0,1)$, $G_x(g(x,\tau)) = \tau$. Moreover, because $Q_{x'x}$ is identified for all $(x,x') \in \mathcal{X}^2$, so is G_x. G_x is also strictly increasing because for any x', $F_{Y|X=x'} \circ Q_{x'x}$ is strictly increasing. Hence, $g(x,\tau)$ is identified as the unique solution of $G_x(u) = \tau$ ■

A.2 Proof of Theorem 2

1. Free case, identification.

Let us first provide an informal outline of the proof. Without loss of generality, we set the indices (i,j,k) defined in Assumption 4 to $1,2,3$. We show the result by proving that for any $x_0 \in \mathcal{X}$, $\mathcal{O}_{x_0}$ is dense. Then the condition in Theorem 1 is satisfied by simply letting $J = 0$. To establish that $\mathcal{O}_{x_0}$ is dense, we prove 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. Note that $\mathcal{O}'_{x_0} \subset \mathcal{O}_{x_0}$ because we consider the subgroup generated by s_{12} and s_{13} only instead of $\mathcal{S}$. 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, 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 that $\mathcal{O}'_{x_0}$ is dense by, basically, "unrolling" the unit circle through successive applications of the translation.

a. 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 4, $s_{12} = h(2,.) \circ h^{-1}(1,.)$ 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 $s_{12} = r \circ t \circ r^{-1}$. By Lemma S1 in the supplementary materials, there exists 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 $\lim_{x\to 1} \widetilde{r}''(x) = [s_{12} \circ \widetilde{r}]''(0)$. 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 $(\underline{x}, \overline{x})$.

b. 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 $x \in \mathbb{R}$,

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

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

$$\begin{aligned} \pi(x) = \pi(y) \quad &\Leftrightarrow \quad \exists\, k \in \mathbb{Z} \;/\; x = y + k \\ &\Rightarrow \quad f(x) = f(y+k) = f(y) + k \\ &\Rightarrow \quad \pi \circ f(x) = \pi \circ f(y). \end{aligned} \tag{A.3}$$

For any $\widetilde{y}$ on the unit circle $[0,1)$, let y be such that $\pi(y) = \widetilde{y}$ and define $\widetilde{f}(\widetilde{y})$ by $\widetilde{f}(\widetilde{y}) = \pi \circ f(y)$. Note that $\widetilde{f}(\widetilde{y})$ is well defined because for any x satisfying also $\pi(x) = \widetilde{y}$, (A.3) implies that $\pi \circ f(x) = \pi \circ f(y)$.

Moreover, by construction, $\widetilde{f} \circ \pi = \pi \circ f$. 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,

$$\widetilde{f}^n \circ \pi = \pi \circ f^n, \quad \forall n \in \mathbb{Z} \tag{A.4}$$

c. 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.[6] We can thus apply Denjoy's theorem (see, e.g., Navas 2011, Theorem 3.1.1), and the orbits of the group generated by $\widetilde{f}$ are either all finite or all dense. Suppose that they are finite. Then for any $\dot{x} \in [0,1)$, 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 (A.4), 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 $n \neq 0$. By freeness, we then have $s_{13}^{-n} \circ s_{12}^m =$Id, contradicting Assumption 4. We conclude that any orbit for the group generated by $\widetilde{f}$ is dense in $[0,1)$.

d. $\mathcal{O}'_{x_0}$ 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 (A.4), $\pi \circ f^n(r^{-1}(x_0)) \in \pi\left(\mathcal{V}_y\right)$, 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, ending the proof that $\mathcal{O}_{x_0}$ is dense.

2. Free case, testability.

By Hölder's theorem, $s_{ij} \circ s_{kl} = s_{kl} \circ s_{ij}$ and thus, for all $(x, \tau) \in \mathcal{X} \times (0,1)$,

$$g(s_{ij} \circ s_{kl}(x), \tau) = g(s_{kl} \circ s_{ij}(x), \tau).$$

[6] 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 $Q(x+1) = Q(x) + 1$.

Applying Equation (3.2) twice, we obtain

$$\begin{aligned} g(s_{ij} \circ s_{ik}(x), \tau) = & F^{-1}_{Y|X=s_{ij}\circ s_{kl}(x),Z=j} \circ F_{Y|X=s_{kl}(x),Z=i} \\ & \circ F^{-1}_{Y|X=s_{kl},Z=l} \circ F_{Y|X=x,Z=k} \circ g(x, \tau) \\ g(s_{kl} \circ s_{ij}(x), \tau) = & F^{-1}_{Y|X=s_{kl}\circ s_{ij}(x),Z=l} \circ F_{Y|X=s_{ij}(x),Z=k} \\ & \circ F^{-1}_{Y|X=s_{ij},Z=j} \circ F_{Y|X=x,Z=i} \circ g(x, \tau). \end{aligned}$$

Because this holds for all (x, τ), we get

$$\begin{aligned} & F^{-1}_{Y|X=s_{ij}\circ s_{kl}(x),Z=j} \circ F_{Y|X=s_{kl}(x),Z=i} \circ F^{-1}_{Y|X=s_{kl},Z=l} \circ F_{Y|X=x,Z=k} \\ = & F^{-1}_{Y|X=s_{kl}\circ s_{ij}(x),Z=l} \circ F_{Y|X=s_{ij}(x),Z=k} \circ F^{-1}_{Y|X=s_{ij},Z=j} \circ F_{Y|X=x,Z=i}. \end{aligned}$$

This equality may not hold if Assumptions 1-3 are not satisfied, thus providing a test of these conditions together.

3. Nonfree case, identification.

Let $x^1 < ... < x^M$ denote the fixed points of s, $x^0 = \underline{x}$ and $x^{M+1} = \overline{x}$. Let i, k be such that $x \in [x^i, x^{i+1})$ and $x' \in [x^k, x^{k+1})$. Without loss of generality, suppose $i \leq k$. Then let $J = k - i + 1$ and choose any $x_1 \in (x^i, x^{i+1})$, any $x_2 \in (x^{i+1}, x^{i+2})$,..., any $x_J \in (x^k, x^{k+1})$. Let us prove that $\overline{\mathcal{O}}_{x_j} \cap \overline{\mathcal{O}}_{x_j+1} \neq \emptyset$ for $j = 0, ..., J$. The result then follows by Theorem 1. We only prove the result for $j = 0$ and, if $J \geq 2$, for $j = 1$. The reasoning for $j = J$ is similar as for $j = 0$, and the reasoning for $1 < j < J$ is similar as for $j = 1$.

First consider $j = 0$, and suppose that $s(u) < u$ for all $u \in (x^i, x^{i+1})$. The proof is identical if $s(u) > u$. A straightforward induction shows that the sequence $(s^n(x_1))_{n\in\mathbb{N}}$ is decreasing and bounded below by x^i. Thus, it converges to $l \in [x^i, x^{i+1})$ satisfying $s(l) = l$. Thus $l = x^i$, implying that $x^i \in \overline{\mathcal{O}}_{x_1}$. If $x = x^i$, then obviously $x^i \in \overline{\mathcal{O}}_x \cap \overline{\mathcal{O}}_{x_1}$ and the result is proved. Otherwise, reasoning as previously but on $(s^n(x))_{n\in\mathbb{N}}$ shows that $x^i \in \overline{\mathcal{O}}_x$. Thus once more $x^i \in \overline{\mathcal{O}}_x \cap \overline{\mathcal{O}}_{x_1}$ and the result holds for $j = 0$.

Now consider $j = 1$. Because the sequence $(s^{-n}(x_1))_{n\in\mathbb{N}}$ converges to x^{i+1}, $x^{i+1} \in \overline{\mathcal{O}}_{x_1}$. Now, if $s(u) > u$ on (x^{i+1}, x^{i+2}), the sequence $(s^n(x_2))_{n\in\mathbb{N}}$ converges to x^{i+1}. Otherwise, the sequence $(s^{-n}(x_2))_{n\in\mathbb{N}}$ also converges to x^{i+1}. Hence, in any case, $x^{i+1} \in \overline{\mathcal{O}}_{x_1} \cap \overline{\mathcal{O}}_{x_2}$, proving the result for $j = 1$.

4. Nonfree case, testability.

For any $x \in \mathcal{X}$ and $s \in \mathcal{S}$, $s = s_{i_1 j_1} \circ ... \circ s_{i_p j_p}$, let us denote, as before,

$$Q_{s(x)x} = Q_{i_1 j_1 s_{i_2 j_2} \circ ... \circ s_{i_p j_p}(x)} \circ ... \circ Q_{i_p j_p x},$$

where $Q_{ijx} = F^{-1}_{Y|X=s_{ij}(x),Z=j} \circ F_{Y|X=x,Z=i}$. As shown in the proof of Theorem 1, we have, for all $(x, \tau) \in \mathcal{X} \times (0, 1)$,

$$g(s(x), \tau) = Q_{s(x)x} \circ g(x, \tau).$$

Let $s \neq$Id admitting a positive and finite number of fixed points. For any such fixed point x^*, $Q_{s(x^*)x^*} =$Id. This condition has no reason to hold if the model is misspecified, thus providing a test of the model. ■

References

Altonji, J. & Matzkin, R. L. (2005), 'Cross section and panel data estimators for nonseparable models with endogenous regressors', *Econometrica* **73**, 1053–1102.

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

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

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

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

D'Haultfœuille, X. & Février, P. (2011), Identification of nonseparable models with endogeneity and discrete instruments. CREST working paper.

Florens, J., Heckman, J. J., Meghir, C. & 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 circle', *L'enseignement mathématique* **47**, 329–407.

Hoderlein, S. & Sasaki, Y. (2013), Outcome conditioned treatment effects. Working paper.

Imbens, G. W. (2007), Nonadditive models with endogenous regressors, *in* R. Blundell & 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. & Angrist, J. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–475.

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

Kasy, M. (2011), 'Identification in triangular systems using control functions', *Econometric Theory* **27**, 663–671.

Kasy, M. (2014), 'Instrumental variables with unrestricted heterogeneity and continuous treatment', *Review of Economic Studies* **81**, 1614–1636.

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

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

Navas, A. (2011), *Groups of Circle Diffeomorphisms*, Chicago Lectures in Mathematics, University of Chicago Press.

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

Torgovitsky, A. (2014), 'Identification of nonseparable models using instruments with small support', *Econometrica* **Forthcoming**.