GLOBAL IDENTIFICATION IN NONLINEAR SEMIPARAMETRIC MODELS

IVANA KOMUNJER

Abstract. This paper derives primitive conditions for global identification in nonlinear simultaneous equations systems. Identification is semiparametric in the sense that it is based on a set of unconditional moment restrictions. Our contribution to the literature is twofold. First, we derive a set of unconditional moment restrictions on the observables that are the starting point for identification in nonlinear structural systems even when multiple equilibria are present. Second, we provide primitive conditions under which a parameter value that solves those restrictions is unique. We apply our results to a nonlinear transformed regression model with multiple equilibria and give sufficient conditions for identifiability of its parameters.

Keywords: identification, structural systems, multiple equilibria, correspondences, semiparametric models, proper mappings, global homeomorphisms

1. Introduction

The problem of identification of economic relations has a long standing history, with systematic discussions given in a collective work of the Cowles foundation edited by Koopmans (1950).[1] In a nutshell, the identification problem is concerned with the unambiguous definition of the parameters to be estimated. Thus, it precedes the problem of statistical estimation. When identification fails, the properties of

Affiliation and Contact information: Department of Economics, University of California at San Diego, komunjer@ucsd.edu.

Acknowledgments: I would like to thank the participants at Brown, Concordia, CREATES, Stanford, Toulouse, UC Berkeley, UCL, UCLA, and Yale econometric seminars for their valuable comments and suggestions.

[1]See, e.g., Dufour and Hsiao (2008) for a review of historical and recent developments on identification in economics.

conventional statistical procedures are likely to change. The objective of this paper is to provide primitive conditions under which identification is guaranteed to hold.

Based on the work of Koopmans and Reiersøl (1950), a complete treatment of identification in a parametric context was given in Rothenberg (1971) and Bowden (1973). Using an approach based on information criteria, they provided conditions under which parametric models are locally and globally identified. Unfortunately, such results may only be applied in models in which it is possible to specify the likelihood function of the dependent variables.

Situations in which the distribution of the dependent variables is left unspecified require conditions for identification in a nonparametric context. Those have been derived in the work of Brown (1983), Roehrig (1988), Matzkin (1994, 2008), and Benkard and Berry (2006), among others. Common to all the studies is an assumption of independence between the (observed) explanatory variables and latent disturbances to the structural system.

Semiparametric models, which are the focus of this paper, fall in between the fully parametric and nonparametric models. They arise when the distribution of the disturbances is only known to satisfy certain moment restrictions. These are typically expressed as conditions for orthogonality between the disturbances and instruments—functions of explanatory variables—and are hence weaker than an assumption of independence.

The present paper examines identification in semiparametric models defined by unconditional moment restrictions. Thus, its contributions are complementary to the existing literature that considers models with conditional moment restrictions, such as Chesher (2003), Newey and Powell (2003), Chernozhukov and Hansen (2005), Severini and Tripathi (2006), Chernozhukov, Imbens, and Newey (2007), for example. It is worthwhile distinguishing these two cases, as identification in some unconditional moment models implied by the conditional ones may fail even when the conditional model is identified. Examples of such failures can be found in Dominguez and Lobato (2004).

The basic semiparametric results for linear simultaneous equation systems under linear parameter constraints were given in Koopmans (1950). These criteria are the well-known rank conditions that were extended by Fisher (1961, 1965) to nonlinear systems that are still linear in parameters. An important step towards a full treatment of identification in general nonlinear models was made by Fisher (1966) and Rothenberg (1971). Their insight was to treat the identification problem simply as a question of uniqueness of solutions to nonlinear systems of equations.

Both Fisher's (1966) and Rothenberg's (1971) results exploit the uniqueness conditions given in Gale and Nikaidô (1965).[2] In particular, they require that the derivative matrix of the system of (nonlinear) equations be weakly positive quasi-definite, i.e. that its symmetric part be positive semi-definite and that its Jacobian be positive everywhere. In many instances, however, this approach produces sufficient conditions for global identification that—in the words of Rothenberg (1971)— are"*overly strong*".

This paper makes two contributions to the literature on identification in a semi-parametric context. First, we derive a set of unconditional moment restrictions that are the starting point for identification in nonlinear structural systems even in the presence of multiple equilibria. Second, we provide primitive conditions under which a parameter value that solves those restrictions is globally unique. Our uniqueness results exploit the sufficient conditions for proper mappings to be homeomorphic, pioneered by Palais (1959). Hence, we are able to relax the weak positive quasi-definite condition of Gale-Nikaidô-Fisher-Rothenberg.

The paper is organized as follows. Throughout, we consider nonlinear systems of simultaneous equations in which the distribution of the disturbances and instruments is known to satisfy a set of unconditional moment conditions. In Section 2 we show how these conditions give rise to the moment restrictions on the distribution of the explanatory and dependent variables that are the starting point for identification.

[2]Specifically, they rely on the results of Theorem 6w (p.89) in Gale and Nikaidô (1965).

This result is non-trivial to derive as we allow the structural system to possess multiple equilibria. Because of equilibrium multiplicity, we model the relation between the disturbances and dependent variables as a correspondence rather than a single-valued mapping. The use of correspondences is relatively novel in the econometric literature (see, e.g., Galichon and Henry, 2006; Beresteanu and Molinari, 2008).

In Section 3, we consider a simple example which gives the key idea behind the main result of the paper. By the same token, we illustrate and discuss the difficulties of finding primitive conditions for identification in general nonlinear models.

Our main result is in Section 4. It derives a set of conditions which guarantee that a solution to a nonlinear system of equations is unique. Two of those conditions are key to the identification: one concerns the Jacobian of the system, while the other excludes "flats". In particular, we assume that the Jacobian of the system is either everywhere non-negative or everywhere non-positive. When the system is continuously differentiable with respect to the structural parameter, this requirement is weaker than the full rank conditions given in Theorem 5.10.2 in Fisher (1966) and Theorem 7 in Rothenberg (1971).[3] In other words, we allow the rank of the derivative matrix to be less than full, provided this only happens over sufficiently small regions in the parameter space. The latter is our second main requirement: that the system does not have any "flats", i.e. does not remain constant over regions in the parameter space that have nonzero dimension. Our results exploit well established results of nonlinear functional analysis.

We conclude in Section 5 with an application of our results to a nonlinear IV model with multiple equilibria. The model is nonlinear in both variables and parameters, and no transformation reduces it to linearity. In addition the model fails to satisfy Gale-Nikaidô-Fisher-Rothenberg assumptions. We show however that it satisfies the conditions of our Corollary 2, thereby establishing the identifiability of the model parameters. All of our proofs are found in an Appendix.

[3]It is worth pointing out that we place conditions only on the sign of the Jacobian. Unlike Gale-Nikaidô-Fisher-Rothenberg, we do not make any positive definiteness assumptions on the derivative matrix of the system.

2. Semiparametric Identification in Nonlinear Structural Models

Let an economic theory specify a system of nonlinear simultaneous equations:

$$\rho(Y, X, \theta) = U \tag{1}$$

in which $\rho : \mathbb{D}_Y \times \mathbb{D}_X \times \mathbb{R}^k \to \mathbb{D}_U$ is a known mapping, and $\mathbb{D}_Y \subseteq \mathbb{R}^G$, $\mathbb{D}_X \subseteq \mathbb{R}^K$, $\mathbb{D}_U \subseteq \mathbb{R}^G$, with $G < \infty$ and $K < \infty$. The variables entering into these equations consist of: a set of observed dependent variables $Y \in \mathbb{D}_Y$, a set of observed explanatory variables $X \in \mathbb{D}_X$, a finite dimensional parameter $\theta \in \mathbb{R}^k$ ($k < \infty$), and a set of latent variables $U \in \mathbb{D}_U$. For example, U can be thought of as disturbance or unaccounted heterogeneity in the model. The object of interest is the true value θ_0 of the structural parameter θ in Equation (1).

We begin our discussion of semiparametric identification with a description of a structure relevant in the context of nonlinear simultaneous equations systems such as the one in Equation (1).

2.1. Structure.

Say that the random variables X and U take values $x \in \mathbb{D}_X$ and $u \in \mathbb{D}_U$, respectively. Fix $x \in \mathbb{D}_X$, $\theta_0 \in \mathbb{R}^k$ and $u \in \mathbb{D}_U$, and consider the structural equations $\rho(y, x, \theta_0) = u$. Under what conditions can those equations be solved for $y \in \mathbb{D}_Y$ in terms of x, u and θ_0? In other words, when is it the case that the structural system in Equation (1) admits at least one equilibrium for the dependent variable?

An equilibrium for Y is guaranteed to exist whenever for a given $(x', \theta_0')' \in \mathbb{D}_X \times \mathbb{R}^k$ the mapping $y \mapsto \rho(y, x, \theta_0)$ is surjective from $\mathbb{D}_Y$ onto $\mathbb{D}_U$. Hereafter, we shall consider the case in which $\mathbb{D}_U = \mathbb{R}^G$ and $\mathbb{D}_Y = \mathbb{R}^G$. We start by assuming the following:

Assumption A. *The mapping $(y, x) \to \rho(y, x, \theta_0)$ is in $\mathcal{C}^1(\mathbb{R}^G \times \mathbb{D}_X)$.*

In order to guarantee surjectivity, we impose the following additional requirement:

Assumption B. *For every $x \in \mathbb{D}_X$, $\lim_{|y|\to\infty}[\rho(y, x, \theta_0)'y]/|y| = \infty$.*

Assumptions A and B ensure that given $(x', \theta_0')' \in \mathbb{D}_X \times \mathbb{R}^k$ the mapping $y \mapsto \rho(y, x, \theta_0)$ is surjective on $\mathbb{R}^G$, i.e. that the inverse image by $\rho(\cdot, x, \theta_0)$ of any point in $\mathbb{R}^G$ is nonempty.[4] Surjectivity guarantees that there is always at least one equilibrium for Y. It does not guarantee, however, that this equilibrium is unique. Yet, equilibrium uniqueness plays an important role in computing the distribution of the dependent variable induced by that of the explanatory variable and the disturbance.

Let $V \equiv (X', U')'$ be the random variable that takes values $v \equiv (x', u')'$ in $\mathbb{D}_X \times \mathbb{R}^G$ and call μ_{XU} the associated measure. When for a given $(x', u', \theta_0')' \in \mathbb{D}_X \times \mathbb{R}^G \times \mathbb{R}^k$, the structural equations $\rho(y, x, \theta_0) = u$ can be globally uniquely solved for $y \in \mathbb{R}^G$ in terms of x, u and θ_0, then one can define (explicitly or implicitly) a single-valued map $y = m(x, u, \theta_0)$ that is continuous in x and u. The transformation T which to each v associates $w \equiv (x', y')'$ is then a single-valued mapping (or function) $T : \mathbb{D}_X \times \mathbb{R}^G \to \mathbb{D}_X \times \mathbb{R}^G$ that is continuous. This leads to the usual definition of the image measure μ_{XY} on $T(\mathbb{D}_X \times \mathbb{R}^G) \subseteq \mathbb{D}_X \times \mathbb{R}^G$ of the random variable $W \equiv (X', Y')'$; we have $W = T(V)$ so $\mu_{XY} = \mu_{XU} \circ T^{-1}$, and $T^{-1} : T(\mathbb{D}_X \times \mathbb{R}^G) \to \mathbb{D}_X \times \mathbb{R}^G$ is the mapping which to each $w = (x', y')'$ in $T(\mathbb{D}_X \times \mathbb{R}^G)$ associates $v = (x', u')' = (x', \rho(y, x, \theta_0)')'$ in $\mathbb{D}_X \times \mathbb{R}^G$. Hence, the distribution of the observables Y and X is generated by the structure $\mathcal{S} = (\theta_0, T, \mu_{XU})$.

When given $(x', u', \theta_0')' \in \mathbb{D}_X \times \mathbb{R}^G \times \mathbb{R}^k$, multiple solutions for $y \in \mathbb{R}^G$ are possible that satisfy $\rho(y, x, \theta_0) = u$, we no longer deal with a single-valued map from V to W but a correspondence $T : \mathbb{D}_X \times \mathbb{R}^G \rightrightarrows \mathbb{D}_X \times \mathbb{R}^G$. Multiple equilibria for the dependent variable are likely to arise in structural systems that are nonlinear in variables. A complete determination of the distribution of the observables X and Y must then include a rule according to which a particular y is chosen from the set of solution points.

More formally, for any $v = (x', u')' \in \mathbb{D}_X \times \mathbb{R}^G$ we shall let $\Gamma_v \equiv \{w \in \mathbb{D}_X \times \mathbb{R}^G : w = (x', y')' \text{ and } \rho(y, x, \theta_0) = u\}$. Then, the correspondence T associates to every $v \in$

[4]See Step 1 in the proof of Proposition 1.

$\mathbb{D}_X \times \mathbb{R}^G$ a set $\Gamma_v \subseteq \mathbb{D}_X \times \mathbb{R}^G$. The random variable W is obtained by transforming V with a single-valued map t that belongs to the class of measurable selections $\operatorname{Sel} T$ of T, whereby $\operatorname{Sel} T = \{t : \mathbb{D}_X \times \mathbb{R}^G \to \mathbb{D}_X \times \mathbb{R}^G$ Borel-measurable and such that $t(v) \in T(v)$ for almost every $v \in \mathbb{D}_X \times \mathbb{R}^G\}$ (see, e.g., Aliprantis and Border, 2007). Together, Assumptions A and B suffice to show that the set $\operatorname{Sel} T$ is nonempty; hence $W = (X', Y')'$ is well defined. We then have the following result:

Proposition 1. *Let Assumptions A and B hold. Then* $\operatorname{Sel} T \neq \emptyset$*, and the structure* $\mathcal{S} = (\theta_0, t, \mu_{XU})$ *with* $\theta_0 \in \mathbb{R}^k$ *and* $t \in \operatorname{Sel} T$ *generates the distribution* μ_{XY} *of the observables* X *and* Y*.*

In particular, since t is Borel-measurable the image measure μ_{XY} on $t(\mathbb{D}_X \times \mathbb{R}^G) \subseteq \mathbb{D}_X \times \mathbb{R}^G$ of the observables X and Y is then again obtained as $\mu_{XY} = \mu_{XU} \circ t^{-1}$, and $t^{-1} : t(\mathbb{D}_X \times \mathbb{R}^G) \to \mathbb{D}_X \times \mathbb{R}^G$ is the mapping which to each $w = (x', y')'$ in $t(\mathbb{D}_X \times \mathbb{R}^G)$ associates $v = (x', u')' = (x', \rho(y, x, \theta_0)')'$ in $\mathbb{D}_X \times \mathbb{R}^G$. Note that our construction of μ_{XY} does not allow for any extrinsic randomness in the choice of equilibria for Y. This, however, is not a serious restriction on the attainable distributions of Y when μ_{XU} is atomless, as shown by Jovanovic (1989).

2.2. **Identification Condition.** The true value θ_0 of the structural parameter θ is said to be identifiable if every structure $\mathcal{S}^* = (\theta_0^*, t^*, \mu_{XU}^*)$ whose characteristics are known to apply to $\mathcal{S} = (\theta_0, t, \mu_{XU})$ and which generates the same distribution of the observables μ_{XY} as $\mathcal{S}$ (i.e. is observationally equivalent to $\mathcal{S}$), satisfies $\theta_0^* = \theta_0$ (see, e.g., Koopmans and Reiersøl, 1950; Roehrig, 1988). Here, we shall assume that $\mathcal{S}$ is known to satisfy:

$$E[G(U, X, \theta_0)] = 0 \tag{2}$$

where $G : \mathbb{R}^G \times \mathbb{D}_X \times \mathbb{R}^k$ is a known moment function of U, X and θ.

The nature of the restrictions in Equation (2) is semiparametric: while the distribution μ_{XU} is left unknown, a number of unconditional moment conditions relating X, U and θ_0 are known to hold. When the moment function is of the form

$G(U, X, \theta) = A(X, \theta)U$ then the moment restrictions in (2) reduce to the familiar orthogonality conditions between the disturbance U and a $k \times G$ matrix of instruments $A(X, \theta)$. Weaker than independence, such orthogonality conditions are typically found in models in which θ_0 is to be estimated via Instrumental Variables (IV) methods.

Using the results of Proposition 1, the expectation of $G(U, X, \theta_0)$ (computed under μ_{XU}) can be related to that of $G\big(\rho(Y, X, \theta_0), X, \theta_0\big)$ (computed under μ_{XY}). To simplify the notation, let $\bar{G}(v, \theta_0) \equiv G(u, x, \theta_0)$ where $v = (x', u')' \in \mathbb{D}_X \times \mathbb{R}^G$, so:

$$E[G(U, X, \theta_0)] = \int_{\mathbb{D}_X \times \mathbb{R}^G} G(u, x, \theta_0)\, \mu_{XU}(\mathrm{d}x\mathrm{d}u) = \int_{\mathbb{D}_X \times \mathbb{R}^G} \bar{G}(v, \theta_0)\, \mu_{XU}(\mathrm{d}v)$$

Recall that we have $\mu_{XU} = \mu_{XY} \circ t$. Then,

$$\begin{aligned}
\int_{\mathbb{D}_X \times \mathbb{R}^G} \bar{G}(v, \theta_0)\, \mu_{XU}(\mathrm{d}v) &= \int_{\mathbb{D}_X \times \mathbb{R}^G} \bar{G}(v, \theta_0)\, \mu_{XY} \circ t(\mathrm{d}v) \\
&= \int_{t(\mathbb{D}_X \times \mathbb{R}^G)} \bar{G}\big(t^{-1}(w), \theta_0\big)\, \mu_{XY}(\mathrm{d}w) \\
&= \int_{t(\mathbb{D}_X \times \mathbb{R}^G)} G\big(\rho(y, x, \theta_0), x, \theta_0\big)\, \mu_{XY}(\mathrm{d}x\mathrm{d}y) \\
&= E\big[G\big(\rho(Y, X, \theta_0), X, \theta_0\big)\big]
\end{aligned}$$

where the second equality follows by a change of variable $w = t(v)$ with $v = (x', u')'$ and $w = (x', y')'$ (see, e.g., Theorem 16.13 in Billingsley, 1995), and the third equality uses the fact t^{-1} is the mapping which to each $w = (x', y')' \in t(\mathbb{D}_X \times \mathbb{R}^G)$ associates $v = (x', u')' = (x', \rho(y, x, \theta_0)')'$. Under two observationally equivalent structures $\mathcal{S}$ and $\mathcal{S}^*$ we then have $E\big[G\big(\rho(Y, X, \theta_0), X, \theta_0\big)\big] = 0$ and $E\big[G\big(\rho(Y, X, \theta_0^*), X, \theta_0^*\big)\big] = 0$, where both expectations are taken with respect to μ_{XY}. This leads to the following necessary and sufficient condition for identification of θ_0, that is valid in simultaneous equations systems in (1) known to satisfy the unconditional moment restrictions in (2):

Definition 1. *Let Assumptions A and B hold. Assume that the observables X and Y are generated by a structure $\mathcal{S} = (\theta_0, t, \mu_{XU})$ in which $\theta_0 \in \mathbb{R}^k$, $t \in \operatorname{Sel} T$, and*

$E[G(U,X,\theta_0)] = 0$. *Then* θ_0 *is identifiable if and only if* $E\big[G\big(\rho(Y,X,\theta),X,\theta\big)\big] = 0$ *has a unique solution* $\theta = \theta_0$ *on* $\mathbb{R}^k$.

To simplify the notation, we shall hereafter let:

$$r(Y,X,\theta) \equiv G\big(\rho(Y,X,\theta),X,\theta\big)$$

for every $\theta \in \mathbb{R}^k$. The expectation $E[r(Y,X,\theta)]$ in Definition 1 is taken with respect to μ_{XY} obtained under the true parameter value θ_0. In order to guarantee that the expectation is well defined, we impose the following:

Assumption C. *For every* $\theta \in \mathbb{R}^k$, $E[r(Y,X,\theta)]$ *exists and is finite.*

We now derive a useful property that is equivalent to that of identifiability of θ_0 given in Definition 1. For this, we introduce two new objects. Let R be any non-stochastic $k \times k$ matrix with full rank. We allow R to depend on the parameter θ and its true value θ_0, as well as on the distribution of the observables μ_{XY}, so long as this dependence remains deterministic and such that $\det R$ never vanishes. In addition, let $h : \mathbb{R}^k \to \mathbb{R}^k$ be any mapping that is a homeomorphism.[5] Similar to previously, we allow the functional form of h to depend on θ, θ_0 and μ_{XY} so long as its homeomorphic property is preserved.

Define a mapping $g : \mathbb{R}^k \to \mathbb{R}^k$ which to each $\theta \in \mathbb{R}^k$ assigns $g(\theta) \equiv h\big(R\,E[r(Y,X,\theta)]\big)$. As previously, the expectation is taken with respect to μ_{XY}. We then have the following property:

$$E[r(Y,X,\theta)] = 0 \text{ if and only if } g(\theta) = h(0). \tag{3}$$

The property in Equation (3) states that $E[r(Y,X,\theta)] = 0$ has a unique solution $\theta = \theta_0$ if and only if $\theta = \theta_0$ is the unique solution to $g(\theta) = h(0)$. Note that the matrix R effectively rotates the moment conditions without affecting the solution; indeed, $E[r(Y,X,\theta)] = 0$ is equivalent to $R\,E[r(Y,X,\theta)] = 0$ so long as R is of full

[5]A mapping is said to be a homeomorphism if it is continuous, one-to-one, onto, and has a continuous inverse.

rank. The mapping h on the other hand changes the image by $R\,E[r(Y,X,\theta)]$ of the true value θ_0 without affecting uniqueness; indeed $h\big(R\,E[r(Y,X,\theta)]\big) = h(0)$ if and only if $R\,E[r(Y,X,\theta)] = 0$ so long as h is a homeomorphism.

The identifiability condition in Definition 1 is thus equivalent to the condition that $g(\theta) = h(0)$ be uniquely solved at $\theta = \theta_0$, which is the well-known GMM identification condition. As pointed out by Newey and McFadden (1994) (Section 2.2.3, p.2127) "here conditions for identification are like conditions for unique solutions of nonlinear equations [...], which are known to be difficult." Before proceeding, we consider a simple example which illustrates the difficulties associated with a general treatment of the identification problem, and gives the insights of our approach.

3. Example and Intuition

Say that the structural parameter θ in Equation (1) is a scalar ($k = 1$) whose true value $\theta_0 \in \mathbb{R}$ is known to satisfy $E[r(Y,X,\theta_0)] = 0$. As previously, the expectation is taken with respect to μ_{XY} obtained under θ_0, and the map $r : \mathbb{R}^3 \to \mathbb{R}$ is known. Using the property in (3), the true parameter value θ_0 is identifiable if and only if the equation $g(\theta) \equiv h\big(R\,E[r(Y,X,\theta)]\big) = h(0)$ has a unique solution $\theta = \theta_0$ on $\mathbb{R}$. Here, $R \in \mathbb{R}$ is positive (negative) and h is a homeomorphism from $\mathbb{R}$ onto $\mathbb{R}$.

A simple way to guarantee uniqueness is to require that the mapping g be a homeomorphism from $\mathbb{R}$ onto $\mathbb{R}$. That g is a homeomorphism ensures that for every $p \in \mathbb{R}$ the equation $g(\theta) = p$ has a unique solution in θ. Identifiability of θ_0 then follows by considering $p = h(0)$. When θ is a scalar, requiring that g be continuous and strictly monotone on $\mathbb{R}$ is sufficient to guarantee that it is a homeomorphism from $\mathbb{R}$ onto $\mathbb{R}$. If we restrict our attention to those mappings that are continuously differentiable with respect to θ on the parameter space $\mathbb{R}$, then a sufficient condition for identification is simply that g' be positive (negative) on $\mathbb{R}$.

> Example. Consider a simple nonlinear moment restriction $E[Y - \theta_0^2 X + \theta_0 X^2] = 0$ (with $\theta_0 \in \mathbb{R}$) taken from Example 2 in Dominguez and Lobato (2004). Let $R = 1$ and h be the identity map on

$\mathbb{R}$, so $g(\theta) = E[r(Y, X, \theta)]$. Here, $g'(\theta) \equiv \partial E[r(Y, X, \theta)]/\partial\theta = E[-X(X + 2\theta)]$, provided we can exchange the orders of integration and derivation. Assume that $E(X^2) > 0$. If $E(X) = 0$, then any θ_0 in $\mathbb{R}$ is identifiable. If on the other hand $E(X) \neq 0$, then both θ_0 and $E(X^2)/E(X) - \theta_0$ solve the moment restriction. So unless $\theta_0 = E(X^2)/[2E(X)]$, there are two distinct solutions on $\mathbb{R}$ to the moment restriction, and identification of θ_0 fails.

While the discussion is simple in the case of a single parameter, complications arise when $k > 1$. In that case, θ and $r(Y, X, \theta)$ are both vectors in $\mathbb{R}^k$, the map g is from $\mathbb{R}^k$ to $\mathbb{R}^k$ and we are brought to consider its Jacobian instead of the above derivative. Unfortunately, requiring that the Jacobian of g be positive (negative) on $\mathbb{R}^k$ no longer suffices to show that g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$. A standard counterexample is the mapping $c : \mathbb{R}^2 \to \mathbb{R}^2$ which to each $(\theta_1, \theta_2)' \in \mathbb{R}^2$ assigns $c(\theta_1, \theta_2) = (\exp\theta_1 \cos\theta_2, \exp\theta_1 \sin\theta_2)$. It is easy to check that its Jacobian is everywhere positive, yet the inverse image by c of any point in $\mathbb{R}^2\backslash\{0\}$ has an infinite number of distinct elements. Our solution is to first eliminate the mappings such as c by requiring that g be proper, i.e. that the inverse image of any compact set be compact. This condition is clearly violated by c since for any $(p_1, p_2)' \in \mathbb{R}^2\backslash\{0\}$ the inverse image $c^{-1}(\{(p_1, p_2)'\})$ is unbounded (hence not compact) in $\mathbb{R}^2$.

Properness by itself does not guarantee that g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$. The latter is true if one is willing to assume that in addition its Jacobian $\det \mathrm{D}g$ never vanishes (see, e.g., Corollary 4.3 in Palais, 1959). Still, in structural models that are nonlinear in θ, everywhere non-vanishing Jacobian might be too strong of an assumption. Many simple mappings on $\mathbb{R}^k$ fail to satisfy this requirement. Even in the scalar case ($k = 1$), requiring that the derivative g' be positive (negative) everywhere on $\mathbb{R}$ would rule out such simple nonlinear mappings as $g(\theta) = \theta^3$. It turns out, however, that when $k \neq 2$, restricting the Jacobian to be either non-negative on $\mathbb{R}^k$ or non-positive on $\mathbb{R}^k$ suffices to make a proper mapping g homeomorphic, provided its inverse images of individual points are of dimension zero, i.e. contain countably many points of $\mathbb{R}^k$. In particular, the latter requirement excludes the cases in which g remains "flat" on subsets of $\mathbb{R}^k$ that have nonzero dimension.[6]

Working with systems whose Jacobian possibly vanishes requires additional restrictions on the dimension of the branch set, i.e. the set of points where g fails to be a local homeomorphism. We separately consider two cases: one where the branch set is known to be bounded, and a second one where the branch set is possibly unbounded but its dimension does not exceed $k-3$. The two cases are covered, respectively, in our main Theorem and Corollary, to which we now turn.

4. Main Result

Consider again the simultaneous equations system in (1) for which the true parameter value θ_0 satisfies the moment restrictions in (2). As previously, $r(Y,X,\theta) = G\big(\rho(Y,X,\theta),X,\theta\big)$ and $g(\theta) = h\big(R\,E[r(Y,X,\theta)]\big)$ for any $\theta \in \mathbb{R}^k$. The $k \times k$ matrix R is of full rank, and h is homeomorphic from $\mathbb{R}^k$ onto $\mathbb{R}^k$. We now derive primitive conditions under which:

$$g \text{ is a homeomorphism from } \mathbb{R}^k \text{ onto } \mathbb{R}^k \tag{4}$$

According to Equation (3), the property in (4) is sufficient for θ_0 to be identifiable. Notice, however, that the homeomorphic property on $\mathbb{R}^k$ is not strictly necessary since identification only restricts the behavior of g around $g(\theta) = h(0)$.

Hereafter, we shall work with mappings g that are twice continuously differentiable on $\mathbb{R}^k$.

Assumption D. *The map g is in $\mathcal{C}^2(\mathbb{R}^k)$.*

In what follows, we shall let $\mathrm{D}g \in L(\mathbb{R}^k,\mathbb{R}^k)$ denote the derivative of g. The following assumption restricts the behavior of the Jacobian $J \equiv \det \mathrm{D}g$ of g on $\mathbb{R}^k$.

---

[6]Unfortunately, the result does not hold in dimension $k=2$. A simple counterexample is the mapping $b : \mathbb{R}^2 \to \mathbb{R}^2$ which to each $(\theta_1,\theta_2)' \in \mathbb{R}^2$ assigns $b(\theta_1,\theta_2) = (\theta_1^2 - \theta_2^2, \theta_1\theta_2)$ (for a discussion see, e.g., Church and Hemmingsen, 1960; Chua and Lam, 1972).

Unlike all of our other assumptions on the mapping g, this condition on the Jacobian of g effectively restricts the equilibrium selection procedure t.

Assumption E. *For every $\theta \in \mathbb{R}^k$, $J(\theta)$ is non-negative (non-positive).*

The condition on the non-negativity (non-positivity) of the Jacobian J is a weakening of the Gale-Nikaidô-Fisher-Rothenberg condition that the latter be positive. Note that unlike Gale-Nikaidô-Fisher-Rothenberg, Assumption E does not require the matrix of derivatives $\mathrm{D}g$ to be quasi-positive definite.

It is worth pointing out that the sign condition in Assumption E is also a weakening of the condition that the Jacobian be non-vanishing on $\mathbb{R}^k$. Indeed, if g is twice continuously differentiable then its Jacobian J is continuous, so requiring that for every $\theta \in \mathbb{R}^k$, $J(\theta) \neq 0$ is equivalent to requiring that J be either positive or negative on $\mathbb{R}^k$.

Next, we require that the mapping g be proper, i.e. that the inverse image by g of each compact subset of $\mathbb{R}^k$ be a compact in $\mathbb{R}^k$. A sufficient condition is:

Assumption F. $|g(\theta)| \to \infty$ *whenever* $|\theta| \to \infty$.

Finally, we impose the following:

Assumption G. *For every $p \in \mathbb{R}^k$ the equation $g(\theta) = p$ has countably many (possibly zero) solutions in $\mathbb{R}^k$.*

The requirement that $g(\theta) = p$ have at most countably many solutions is only binding for values of p that are not regular (such values are called critical values). Indeed, if p is a regular value (meaning that the inverse image of $\{p\}$ contains only the parameter values $\theta_r \in \mathbb{R}^k$ for which the Jacobian $J(\theta_r)$ is different from zero) then the set of solutions to $g(\theta) = p$ is finite.[7] This requirement excludes the situations

[7]By properness, the inverse image of $\{p\}$ is a compact set in $\mathbb{R}^k$; the inverse function theorem guarantees that this set is discrete, hence it is finite (see, e.g., step 5 in the proof of Theorem by Debreu, 1970).

in which the map g remains "flat" over regions in the parameter space that are of dimension greater or equal than 1.

We are now ready to state our main result:

Theorem 1. *Assume $k \neq 2$ and let Assumptions A through G hold. If the set of points $\theta_s \in \mathbb{R}^k$ for which* $\operatorname{rank} \mathrm{D}g(\theta_s) < k-1$ *is bounded, then θ_0 is identified on $\mathbb{R}^k$.*

We first comment on the strength of the conditions imposed in Theorem 1. Our identification result does not hold if the dimension of the parameter set is $k = 2$. The reason behind is that our Assumptions C-G, when combined with the boundedness condition in Theorem 1, do not guarantee that the mapping g is a homeomorphism from $\mathbb{R}^2$ onto $\mathbb{R}^2$.[8] Now consider the second requirement of Theorem 1 and let R_q $(0 \leqslant q \leqslant k)$ denote the set of parameter values θ_s at which $\operatorname{rank} \mathrm{D}g(\theta_s) \leqslant q$. A simple sufficient condition for R_{k-2} to be bounded, is that the Jacobian does not vanish at infinity. Indeed, if for large enough values of $|\theta|$ the Jacobian remains positive (negative), then the set R_{k-1} remains bounded. A fortiori, its subset R_{k-2} is then bounded as well.

The intuition behind Theorem 1 is simple: Assumptions A and B are used to translate the problem of identification into that of uniqueness of solutions to the system of equations $g(\theta) = h(0)$. Assumptions C through G on the other hand ensure that the the map g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$, provided $k \neq 2$ and the boundedness condition of Theorem 1 holds. Of course, there exist alternative sets of assumptions guaranteeing that g is a homeomorphism. One example is Chua and Lam (1972), who replace Assumption E and our boundedness condition with the requirement that the set R_{k-1} be of dimension less or equal than 0 (see Theorem 2.2 in Chua and Lam, 1972).[9] The following result relaxes Chua and Lam's (1972)

[8] Church and Hemmingsen (1960) and Chua and Lam (1972) contain simple examples of mappings that are not homeomorphisms yet satisfy all the requirements of Theorem 1 except $k \neq 2$. See also our footnote 6.

[9] The empty set is the only set that has dimension -1.

dimension requirement on R_{k-1} by replacing it with a weaker requirement on $R_{k-2} \subseteq R_{k-1}$:

Corollary 2. *Assume $k > 2$ and let Assumptions A through G hold. If the set of points $\theta_s \in \mathbb{R}^k$ for which* $\operatorname{rank} \mathrm{D}g(\theta_s) < k-1$ *is of dimension less or equal than $k-3$, then θ_0 is identified on $\mathbb{R}^k$.*

Theorem 1 and its Corollary 2 give sufficient conditions for global identification of θ_0 under alternative assumptions on the set R_{k-2}. If the latter is bounded, then the result of Theorem 1 applies. If boundedness cannot be established, then Corollary 2 still holds provided the dimension of R_{k-2} remains sufficiently small relative to the dimension k of the parameter space.

5. Application and Conclusion

We conclude by first giving an application of our main result to a nonlinear IV model with multiple equilibria. Consider a nonlinear transformed regression model relating a scalar dependent variable $Y \in \mathbb{R}$ to a scalar explanatory variable $X \in \mathbb{D}_X \subseteq \mathbb{R}$ and a disturbance $U \in \mathbb{R}$:

$$Y^3 - (\beta + \tfrac{1}{3}\beta^3 + \gamma)Y = \alpha + (\beta + \gamma + \tfrac{1}{3}\gamma^3)X + U, \quad (\alpha, \beta, \gamma)' \in \mathbb{R}^3. \tag{5}$$

Note that when $\beta\gamma \neq 0$ the above model is nonlinear in the parameter $\theta \equiv (\alpha, \beta, \gamma)' \in \mathbb{R}^3$, and no transformation reduces it to a linear-in-parameters model. In addition, the model is nonlinear in the variables and exhibits multiple equilibria for Y. We denote by $\theta_0 \equiv (\alpha_0, \beta_0, \gamma_0)'$ the true value of interest of the structural parameter θ. Hereafter, we assume that the structure $\mathcal{S} = (\theta_0, t, \mu_{XU})$ in (5) satisfies the following:

Condition H. *(i)* $E(|X|^4) < \infty$ *and* $E(|Y|^6) < \infty$*; (ii)* $\mathrm{Var}(X) \neq 0$*, and* $\mathrm{Cov}(Y, X) \neq 0$ *or* $\mathrm{Cov}(Y, X^2) \neq 0$*; (iii)* $\mathrm{Cov}(Y, X)\mathrm{Cov}(X, X^2) \neq \mathrm{Cov}(Y, X^2)\mathrm{Var}(X)$.

Condition H(i) is used to establish several moment existence and finiteness results.[10] Condition H(ii) excludes the situations in which X is not random, and in which Y is uncorrelated with both X and X^2. This property is used to establish properness of the mapping g. Condition H(iii) on the other hand ensures that the Jacobian of g is not identically equal to zero on $\mathbb{R}^3$.

We now show that under Condition H any θ_0 on $\mathbb{R}^3$ is identifiable by the orthogonality conditions $E(A(X,\theta)U) = 0$ in which the instruments are defined as:

$$A(X,\theta) \equiv (1, X, X^2)' \tag{6}$$

For this, we let $\rho(Y,X,\theta) \equiv Y^3 - (\beta + \frac{1}{3}\beta^3 + \gamma)Y - \alpha - (\beta + \gamma + \frac{1}{3}\gamma^3)X$, let $G(U,X,\theta) \equiv A(X,\theta)U$, and define the rotation matrix R as:

$$R \equiv \begin{pmatrix} 1 & 0 & 0 \\ -E(X) & 1 & 0 \\ -E(X^2) & 0 & 1 \end{pmatrix}$$

Under Condition H(i) the expected value of $r(Y,X,\theta) = G\big(\rho(Y,X,\theta), X, \theta\big)$ exists and is finite for every $\theta \in \mathbb{R}^3$, which satisfies our Assumption C. Then, letting h be the identity map from $\mathbb{R}^3$ to $\mathbb{R}^3$, we can calculate the mapping g which to every $\theta \in \mathbb{R}^3$ maps $g(\theta) = h\big(R\ E[r(Y,X,\theta)]\big) \in \mathbb{R}^3$. Letting $g(\theta) \equiv (g_1(\theta), g_2(\theta), g_3(\theta))'$ we have:

$$\begin{aligned}
g_1(\theta) &= E(Y^3) - \alpha - (\beta + \tfrac{1}{3}\beta^3 + \gamma)E(Y) - (\beta + \gamma + \tfrac{1}{3}\gamma^3)E(X) \\
g_2(\theta) &= \mathrm{Cov}(Y^3, X) - (\beta + \tfrac{1}{3}\beta^3 + \gamma)\mathrm{Cov}(Y, X) - (\beta + \gamma + \tfrac{1}{3}\gamma^3)\mathrm{Var}(X) \\
g_3(\theta) &= \mathrm{Cov}(Y^3, X^2) - (\beta + \tfrac{1}{3}\beta^3 + \gamma)\mathrm{Cov}(Y, X^2) - (\beta + \gamma + \tfrac{1}{3}\gamma^3)\mathrm{Cov}(X, X^2)
\end{aligned}$$

[10]Similar to Assumption E, the moment restriction on Y restricts the equilibrium selection t. Echenique and Komunjer (2007) provide examples that show how—in models with multiple equilibria—the equilibrium selection affects the moments of the dependent variable.

The map g is twice continuously differentiable on $\mathbb{R}^3$, which satisfies our Assumption D, and its Jacobian equals:

$$J(\theta) = -[\text{Cov}(Y, X)\text{Cov}(X, X^2) - \text{Cov}(Y, X^2)\text{Var}(X)](\beta^2 + \gamma^2 + \beta^2\gamma^2)$$

By Assumption H(iii) we have $\text{Cov}(Y, X)\text{Cov}(X, X^2) \neq \text{Cov}(Y, X^2)\text{Var}(X)$. So J is either non-positive or non-negative everywhere on $\mathbb{R}^3$, which satisfies the requirement in Assumption E. It is worth pointing out $J(\theta) = 0$ whenever $\beta = \gamma = 0$, so Gale-Nikaidô-Fisher-Rothenberg identification result does not apply here.

We now briefly check that all the remaining conditions of our Corollary 2 hold. The model in Equation (5) is continuous in both Y and X which satisfies our Assumption A. Moreover, as $y \to \infty$ the dominant term in $\rho(y, x, \theta_0)'y$ is y^4 so $[\rho(y, x, \theta_0)'y]/|y| \to \infty$, which satisfies our Assumption B.

To check the properness condition in Assumption F, we consider all possible cases under which $|\theta| \to \infty$. Say that $|\gamma| \to \infty$; then, if β is such that $|\beta+\gamma+\frac{1}{3}\gamma^3| \to \infty$ we have from H(ii) that $|g_2(\theta)| \to \infty$; if on the other hand β is such that $|\beta+\gamma+\frac{1}{3}\gamma^3| \to C$ with C a real constant, then $|\beta + \frac{1}{3}\beta^3 + \gamma| \to \infty$ so by H(ii) either $|g_2(\theta)| \to \infty$ or $|g_3(\theta)| \to \infty$. Now assume that γ remains constant; then, if $|\beta| \to \infty$ we have again $|g_2(\theta)| \to \infty$ or $|g_3(\theta)| \to \infty$, depending on the nonzero term in H(ii); if on the other hand β remains constant and $|\alpha| \to \infty$ we have $|g_1(\theta)| \to \infty$. So $|\theta| \to \infty$ implies $|g(\theta)| \to \infty$.

We now show that Assumption G holds as well. For this, note that $J(\theta) = 0$ only if $\beta = \gamma = 0$. The critical values of g are then of the form: $p = (E(Y^3) - -\alpha, \text{Cov}(Y^3, X), \text{Cov}(Y^3, X^2))'$. Hence, the points in the inverse image $g^{-1}(\{p\})$ necessarily solve the following equations:

$$\begin{aligned}
(\beta + \tfrac{1}{3}\beta^3 + \gamma)\text{Cov}(Y, X) + (\beta + \gamma + \tfrac{1}{3}\gamma^3)\text{Var}(X) &= 0 \\
(\beta + \tfrac{1}{3}\beta^3 + \gamma)\text{Cov}(Y, X^2) + (\beta + \gamma + \tfrac{1}{3}\gamma^3)\text{Cov}(X, X^2) &= 0
\end{aligned}$$

We now show that the unique solution to the above system is $(\beta, \gamma) = (0, 0)$. For this, let P be a matrix defined as:

$$P \equiv \begin{pmatrix} \mathrm{Cov}(Y, X) & \mathrm{Var}(X) \\ \mathrm{Cov}(Y, X^2) & \mathrm{Cov}(X, X^2) \end{pmatrix}$$

Note that under Condition H(iii) we have $\det P \neq 0$ so P is of full rank. Premultiplying the above system of equations by the matrix P^{-1} we get an equivalent system of equations:

$$\begin{aligned} \beta + \tfrac{1}{3}\beta^3 + \gamma &= 0 \\ \beta + \gamma + \tfrac{1}{3}\gamma^3 &= 0 \end{aligned}$$

whose unique solution is $(\beta, \gamma) = (0, 0)$. Hence, if $p \equiv (p_1, p_2, p_3)'$ is a critical value of g then its inverse image equals $g^{-1}(\{p\}) = \{(E(Y^3) - p_1, 0, 0)'\}$, which is of dimension 0.

Finally, we check that the dimension requirement of Corollary 2 holds. For this, note that the points θ_s for which $\mathrm{rank}\, \mathrm{D}g(\theta_s) < 3$ are necessarily of the form $\theta_s = (\alpha, 0, 0)'$. We now show that under Condition H for any such θ_s we have $\mathrm{rank}\, \mathrm{D}g(\theta_s) = 2$. We reason by contradiction: assume that $\mathrm{rank}\, \mathrm{D}g(\theta_s) = 1$; then necessarily $\mathrm{Cov}(Y, X) + \mathrm{Var}(X) = 0$ and $\mathrm{Cov}(Y, X^2) + \mathrm{Cov}(X, X^2) = 0$, which in turn implies $\mathrm{Cov}(Y, X)\mathrm{Cov}(X, X^2) = \mathrm{Cov}(Y, X^2)\mathrm{Var}(X)$; the last property is in contradiction with Condition H(iii). It follows that the set of points θ_s for which $\mathrm{rank}\, \mathrm{D}g(\theta_s) < 2$ is empty, hence of dimension -1. Thus, under Condition H, the result of Corollary 2 applies, which shows that any value θ_0 of the structural parameter θ in the nonlinear model (5) is identifiable via the instruments given in (6).

The structural model in (5) is an example of nonlinear models with multiple equilibria that satisfy the primitive conditions for global identification derived in Corollary 2. The main goal of this paper was to investigate the issue of global identification of finite dimensional parameters that appear in nonlinear systems of structural equations. We focused on semiparametric identification based on a set of unconditional moment restrictions known to be satisfied by the true value of interest of the structural parameter. Our main results—given in Theorem 1 and its Corollary 2—provide sufficient primitive conditions for identification to hold globally.

Appendix A. Proofs

Proof of Proposition 1. The proof is done in three steps.

STEP 1: We first show that, given $\theta_0 \in \mathbb{R}^k$, the correspondence T is closed-valued, i.e. for any $v \equiv (x', u')' \in \mathbb{D}_X \times \mathbb{R}^G$, $T(v) = \Gamma_v$ is a closed subset of $\mathbb{D}_X \times \mathbb{R}^G$. We have $\Gamma_v = \{x\} \times \Lambda_v$ where $\Lambda_v \equiv \{y \in \mathbb{R}^G : \rho(y, x, \theta_0) = u\}$, so it suffices to show that Λ_v is a closed subset of $\mathbb{R}^G$.

Fix $(x, \theta_0) \in \mathbb{D}_X \times \mathbb{R}^k$ and let $\tilde{h} : \mathbb{R}^G \to \mathbb{R}^G$ be a mapping which to each $y \in \mathbb{R}^G$ assigns $\tilde{h}(y) = \rho(y, x, \theta_0)$. Since by Assumption A, $\tilde{h} \in \mathcal{C}(\mathbb{R}^G)$ and by Assumption B, $\lim_{|y|\to\infty}[\tilde{h}(y)'y]/|\, y| = \infty$, we then have by Theorem 3.3 in Deimling (1985) that $\tilde{h}$ is surjective, i.e. $\tilde{h}(\mathbb{R}^G) = \mathbb{R}^G$.

We now show that $\tilde{h}$ is also proper, i.e. that the inverse image by $\tilde{h}$ of each compact subset of $\mathbb{R}^G$ is compact in $\mathbb{R}^G$. For this, note that

$$|\tilde{h}(y)'y|/|\, y| \leqslant |\tilde{h}(y)|$$

so Assumption B also implies $\lim_{|y|\to\infty} |\tilde{h}(y)| = \infty$. Let then $K \subset \mathbb{R}^G$ be compact, i.e. closed and bounded. By Assumption A we know that $\tilde{h}$ is continuous, hence $\tilde{h}^{-1}(K)$ is closed in $\mathbb{R}^G$. To show that $\tilde{h}^{-1}(K)$ is bounded, consider a sequence $\{\tilde{h}(y_n)\}$ $(n \in \mathbb{N})$ in K. Since K is compact, $\tilde{h}(y_n) \underset{n\to\infty}{\longrightarrow} \tilde{h}(y_0) \in K$, which by $\lim_{|y|\to\infty} |\tilde{h}(y)| = \infty$ implies that the sequence $\{y_n\}$ $(n \in \mathbb{N})$ is bounded. Hence, $\tilde{h}^{-1}(K)$ is bounded, therefore compact in $\mathbb{R}^G$.

We can now show that the correspondence T is closed-valued. Take any $v = (x', u')' \in \mathbb{D}_X \times \mathbb{R}^G$. Since $\mathbb{R}^G = \tilde{h}(\mathbb{R}^G)$, we have $\Lambda_v = \tilde{h}^{-1}(u)$ which by properness of $\tilde{h}$ is compact, hence closed in $\mathbb{R}^G$.

STEP 2: We next show that, given $\theta_0 \in \mathbb{R}^k$, the correspondence T is Borel-measurable.

A necessary and sufficient condition for T to be Borel-measurable and closed-valued is: *for any K compact in $\mathbb{D}_X \times \mathbb{R}^G$ and any $\varepsilon > 0$, there exists H compact $H \subset K$ such that $\mu(K \backslash H) < \varepsilon$ and $T|_H$ is closed-graph* (see, e.g., Proposition 2 in Berliocchi and Lasry, 1973). Consider the graph of the correspondence $T|_K$, $\operatorname{Gr}(T|_K) = \{(v', w')' \in (\mathbb{D}_X \times \mathbb{R}^G)^2 : v \in K, w \in \Gamma_v\}$. We need to show that $\operatorname{Gr}(T|_K)$ is a closed subset of $\mathbb{R}^{2(K+G)}$. We know that $\operatorname{Gr}(T|_K)$ is closed if and only if for any sequence $\{(v_n', w_n')'\}$ $(n \in \mathbb{N})$ in $\operatorname{Gr}(T|_K)$, $v_n \underset{n\to\infty}{\longrightarrow} a$, $w_n \underset{n\to\infty}{\longrightarrow} b$ imply that $(a', b')' \in \operatorname{Gr}(T|_K)$. Take then $\{(v_n', w_n')'\}$ in $\operatorname{Gr}(T|_K)$. Let $\bar{h} : \mathbb{D}_X \times \mathbb{R}^G \to \mathbb{D}_X \times \mathbb{R}^G$ be a mapping which to each $w \equiv (x', y')' \in \mathbb{D}_X \times \mathbb{R}^G$ assigns $\bar{h}(w) = (x', \rho(y, x, \theta_0)')'$ so $\bar{h}(w) = v$. From Assumption A, we know that $\bar{h} \in \mathcal{C}(\mathbb{D}_X \times \mathbb{R}^G)$. By continuity of $\bar{h}$, $w_n \underset{n\to\infty}{\longrightarrow} b$ implies $v_n = \bar{h}(w_n) \underset{n\to\infty}{\longrightarrow} \bar{h}(b) = a$. So, $b \in \Gamma_a$. Since K is compact, $a \in K$, therefore $(a', b')' \in \operatorname{Gr}(T|_K)$.

STEP 3: Finally, we can show that $\operatorname{Sel} T \neq \emptyset$. This is an immediate consequence of a corollary to Kuratowski's theorem: *if a correspondence $T : \mathbb{D}_X \times \mathbb{R}^G \rightrightarrows \mathbb{D}_X \times \mathbb{R}^G, v \to \Gamma_v$ is Borel-measurable, closed-valued and such that $\Gamma_v \neq \emptyset$ for almost every v, then* $\operatorname{Sel} T \neq \emptyset$; see, e.g., Corollary 1 in Berliocchi and Lasry (1973) or Theorem 18.13 in Aliprantis and Border (2007). Note that $\Gamma_v \neq \emptyset$ for every $v \in \mathbb{D}_X \times \mathbb{R}^G$ by the surjectivity of the map h. Hence, $\operatorname{Sel} T \neq \emptyset$ and $W = (X', Y')'$ exists. □

Proof of Theorem 1. We start by fixing the notation. We denote by $g(D)$ the image by g of any subset $D \subseteq \mathbb{R}^k$ of its domain, and by $g^{-1}(R)$ the inverse image by g of any subset $R \subseteq g(\mathbb{R}^k)$ of its range. We let B denote the set of all points $\theta \in \mathbb{R}^k$ at which g fails to be local homeomorphism; the set of all such points is called the branch set of g. We denote by $\mathrm{D}g \in L(\mathbb{R}^k, \mathbb{R}^k)$ the derivative of g. The set of points θ_s at which $\operatorname{rank} \mathrm{D}g(\theta_s) \leqslant q$ with $0 \leqslant q \leqslant k$ is denoted R_q; it follows that $B \subseteq R_{k-1}$. Only the zero matrix has rank 0. Finally, we let J denote the Jacobian of g, i.e. $J = \det \mathrm{D}g$. The proof is done in five steps.

STEP 1: We start by showing that the mapping $g : \mathbb{R}^k \to \mathbb{R}^k$ is open, i.e. that the image of every open subset of $\mathbb{R}^k$ is open in $\mathbb{R}^k$, and that we have: $B = \emptyset$ when $k = 1$ and $B \subseteq R_{k-2}$ when $k > 1$.

Openness of g is an immediate consequence of Theorem 2 in Titus and Young (1952): *every* $g : \mathbb{R}^k \to \mathbb{R}^k$ *of class* $\mathcal{C}^1$ *such that* $\dim g^{-1}(p) = 0$ *for every* $p \in g(\mathbb{R}^k)$ *and whose Jacobian* J *is non-negative (non-positive) on* $\mathbb{R}^k$ *is open.* Now we can use an extension of the inverse function theorem for open maps given in Theorem 1.4 by Church (1963): *if* $g : \mathbb{R}^k \to \mathbb{R}^k$ *of class* $\mathcal{C}^1$ *is open, then* g *is locally a homeomorphism at* $\theta \in \mathbb{R}^k$ *whenever* $\operatorname{rank} J(\theta) \geqslant k-1$. Hence, for $k = 1$, openness of g implies that $B = \emptyset$. For $k > 1$, we have $B \subseteq R_{k-2}$.

STEP 2: We now show that when $k > 1$, $\dim B = \dim g(B) = \dim g^{-1}(g(B)) \leqslant k-2$.

For this, we use two results. First, we use Theorem 2 in Sard (1965): $g(R_q)$ *is of dimension* $\leqslant q$*, if* $g \in \mathcal{C}^n(\mathbb{R}^k)$ *with* $n \geqslant k-q$. When $q = k-1$, this result reduces to the well-known Sard's lemma: $\dim g(R_{k-1}) < k$ as long as $g \in \mathcal{C}^1(\mathbb{R}^k)$. When $q = k-2$, then we get the following implication of Sard's (1965) result: $\dim g(R_{k-2}) < k-1$ as long as $g \in \mathcal{C}^2(\mathbb{R}^k)$. We are now ready to combine this property with a second result, which is Corollary 2.3 in Church and Hemmingsen (1960): *If* g *is open and such that* $\dim g^{-1}(p) = 0$ *for every* $p \in g(\mathbb{R}^k)$*, and if* $\dim g(B) < k$*, then* $\dim B = \dim g(B) = \dim g^{-1}(g(B))$. Now, from Step 1 we have $g(B) \subseteq g(R_{k-2})$ so $\dim g(B) \leqslant \dim g(R_{k-2})$. Together with Sard's (1965) theorem, the latter shows that $\dim B = \dim g(B) = \dim g^{-1}(g(B)) \leqslant k-2$.

STEP 3: We show that $g(\mathbb{R}^k) = \mathbb{R}^k$, that g is proper, i.e. that the inverse image by g of any compact subset $K \subset \mathbb{R}^k$ is compact in $\mathbb{R}^k$, and that $\mathbb{R}^k \backslash g^{-1}(g(B))$ is connected when $k > 1$.

Recall from Step 1 that g is open on $\mathbb{R}^k$; this implies that $g(\mathbb{R}^k)$ is open in $\mathbb{R}^k$. Now using Assumption F we show that $g(\mathbb{R}^k)$ is closed in $\mathbb{R}^k$: take a sequence $\{g(\theta_n)\}$ ($n \in \mathbb{N}$) such that $g(\theta_n) \underset{n\to\infty}{\longrightarrow} p$. By Assumption F we then have $\{\theta_n\}$ ($n \in \mathbb{N}$) bounded, so $\theta_n \underset{n\to\infty}{\longrightarrow} \bar{\theta}$ and by continuity of g, $p = g(\bar{\theta})$ and $g(\mathbb{R}^k)$ is closed. Since $g(\mathbb{R}^k)$ is both open and closed in $\mathbb{R}^k$, we have $g(\mathbb{R}^k) = \mathbb{R}^k$.

We now show that the mapping g is proper, i.e. that the inverse image by g of each compact subset of $\mathbb{R}^k$ is a compact subset of $\mathbb{R}^k$. The proof is straightforward:

let $K \subset \mathbb{R}^k$ be compact, i.e. closed and bounded. Given that g is continuous, $g^{-1}(K)$ is closed in $\mathbb{R}^k$. It remains to be shown that $g^{-1}(K)$ is bounded. The proof of boundedness is similar to before. Assumption F implies $|g(\theta)| \to \infty$ whenever $|\theta| \to \infty$. Let $\{g(\theta_n)\}$ $(n \in \mathbb{N})$ be a sequence in K. Since K is compact, $g(\theta_n) \underset{n\to\infty}{\longrightarrow} g(\theta_0) \in K$, which by Assumption F implies that $\{\theta_n\}$ $(n \in \mathbb{N})$ is bounded. Hence $g^{-1}(K)$ is bounded in $\mathbb{R}^k$. Before continuing, let us note that a continuous proper map g is also closed, i.e. $g(B)$ closed whenever $B \subset \mathbb{R}^k$ closed (see, e.g., Corollary in Palais, 1970).

Finally, to show that $\mathbb{R}^k \backslash g^{-1}(g(B))$ is connected for any $k > 1$, we use Theorem IV.4 in Hurewicz and Wallman (1948): *any connected k-dimensional set in $\mathbb{R}^k$ cannot be disconnected by a subset of dimension* $< k-1$. The desired result follows by using the connectedness of $\mathbb{R}^k$ together with $\dim g^{-1}(g(B)) < k-1$ obtained in Step 2.

STEP 4: We now show that the restriction of g to $\mathbb{R}^k \backslash g^{-1}(g(B))$ is a covering map.

For this we use Covering Space Theorem 1 in Plastock (1978): *Let A be a connected open set in $\mathbb{R}^k$. Then $\tilde{g} : A \to \tilde{g}(A)$ is a covering space map if (i) $\tilde{g}$ is a local homeomorphism, and (ii) $\tilde{g}$ is proper.* When $k = 1$, Plastock's (1978) result applies to $A \equiv \mathbb{R}$ and $\tilde{g} = g$, since from Step 1 we known that g is a local homeomorphism on $\mathbb{R}$, and from Step 3 we know that g is proper. Hence, when $k = 1$, g is a covering map.

Now, consider the case $k > 1$. We need to check that all the conditions of Plastock's (1978) theorem are satisfied when $A \equiv \mathbb{R}^k \backslash g^{-1}(g(B))$ and $\tilde{g} \equiv g|_A$. First, we shall establish that $\tilde{g}$ is a local homeomorphism. We have $g^{-1}(g(B)) \supseteq B$ so $A \cap B = \emptyset$ and $\tilde{g} : A \to \mathbb{R}^k \backslash g(B)$ is a local homeomorphism. Next, we show that $\tilde{g}$ is proper: let C be a compact in $\mathbb{R}^k \backslash g(B)$ and note that $\tilde{g}^{-1}(C) = g^{-1}(C)$ since $\tilde{g}^{-1} = g^{-1}|_{\mathbb{R}^k \backslash g(B)}$. Then by properness of g we have that $g^{-1}(C)$ is compact in $\mathbb{R}^k$. Since $C \cap g(B) = \emptyset$ it follows that $g^{-1}(C) \cap g^{-1}(g(B)) = \emptyset$ and so $g^{-1}(C)$ is compact in A.

Finally, we show that A is open. Consider $\theta \in \mathbb{R}^k \backslash B$. Then g is a local homeomorphism at θ, i.e. there exists an open neighborhood U of θ such that $g(U)$ is open in $\mathbb{R}^k$ and $g|_U : U \to g(U)$ is a homeomorphism. So $U \cap B = \emptyset$ and $U \subset \mathbb{R}^k \backslash B$, which shows that $\mathbb{R}^k \backslash B$ is open; hence B is closed. Using our previous observation (made in Step 3) that a continuous proper map is closed, we know that g is closed, so $g(B)$ is closed in $\mathbb{R}^k$. Continuity of g then guarantees that $g^{-1}(g(B))$ is closed in $\mathbb{R}^k$, thus A is open.

From Step 3 we know that when $k > 1$, the set $\mathbb{R}^k \backslash g^{-1}(g(B))$ is connected. We can then apply Plastock's (1978) Covering Space Theorem to show that the restriction of g to $\mathbb{R}^k \backslash g^{-1}(g(B))$ is a covering map.

STEP 5: Finally, we show that when $k \neq 2$ g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$.

For this, we use Theorem 1.3 in Church and Hemmingsen (1960): *Let g be an open map of $\mathbb{R}^k$ onto $\mathbb{R}^k$, $k \neq 2$, such that $\dim g(B) \leqslant k - 2$. If the restriction of g to $\mathbb{R}^k \backslash g^{-1}(g(B))$ is a covering map, and if B is compact, then g is a homeomorphism.*

We now check that all the conditions of Church and Hemmingsen's (1960) result hold. That g is open follows from Step 1; that g is onto $\mathbb{R}^k$ follows from Step 3. In Step 1 we show that when $k = 1$ the set B is empty, so $g(B) = \emptyset$ and $\dim g(B) = -1$. When $k > 2$, Step 2 shows that $\dim g(B) \leqslant k - 2$. That $g|_{\mathbb{R}^k \backslash g^{-1}(g(B))}$ is a covering map follows from Step 4. It remains to show that B is compact. When $k = 1$, the result is trivial. When $k > 2$, we know from Step 4 that B is closed; so if B is in addition bounded then it is compact. From Step 1 we know that $B \subseteq R_{k-2}$; so the condition that R_{k-2} is bounded from Theorem 1 implies that B is bounded.

Under Assumptions C-G, the conditions given in Theorem 1 is sufficient to guarantee that g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$. Together with Assumptions A-B this guarantees, by Definition 1, that θ_0 is identified on $\mathbb{R}^k$. This completes the proof of Theorem 1. $\square$

Proof of Corollary 2. The proof is identical to that of Theorem 1 except for the proof of the result in Step 5 which should be modified as follows:

STEP 5: To show that when $k > 2$ g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$, we first show that the restriction of g to $A = \mathbb{R}^k \backslash g^{-1}(g(B))$ is a homeomorphism.

For this, we use Lemma 1 in Plastock (1978): *If $\tilde{g} : A \to \tilde{g}(A)$ is a covering space map, A and $\tilde{g}(A)$ pathwise connected and $\tilde{g}(A)$ simply connected, then $\tilde{g}$ is a global homeomorphism.* From Step 2 we know that $\dim g(B) = \dim g^{-1}(g(B)) \leqslant k-2$. By using the same reasoning as in Step 3, we then have that A and $\tilde{g}(A) = \mathbb{R}^k \backslash g(B)$ are connected. Recall in addition from Step 4 that A is open and that $g(B)$ is closed, so that $\tilde{g}(A) = \mathbb{R}^k \backslash g(B)$ is open in $\mathbb{R}^k$. Hence, A and $\tilde{g}(A)$ are two open subsets of $\mathbb{R}^k$ that are connected; this implies that they are also pathwise connected. To show that $\tilde{g}(A)$ is simply connected, we use Theorem 25 in Basye (1935): *If K is a closed subset of $\mathbb{R}^k$ of dimension $k-3$ or less, then $\mathbb{R}^k \backslash K$ is simply connected.* Letting $K \equiv g(B)$, we know that K is closed in $\mathbb{R}^k$. Moreover, from Step 2 we know that $\dim g(B) = \dim B \leqslant \dim R_{k-2}$, which from the condition of Corollary 2 is less or equal than $k-3$; this implies that $\tilde{g}(A)$ is simply connected. Hence, $g|_{\mathbb{R}^k \backslash g^{-1}(g(B))}$ is a homeomorphism from $\mathbb{R}^k \backslash g^{-1}(g(B))$ onto $\mathbb{R}^k \backslash g(B)$.

It remains to show that $g|_{g^{-1}(g(B))}$ is a homeomorphism from $g^{-1}(g(B))$ onto $g(B)$. Let then $\bar{g} \equiv g|_{g^{-1}(g(B))}$. By construction, $\bar{g} : g^{-1}(g(B)) \to g(B)$ is onto. We now show that it is also one-to-one: let $p \in g(B)$ and assume that $g^{-1}(p) \supset \{\theta_1, \theta_2\}$ with $\theta_1 \neq \theta_2$. Since $\mathbb{R}^k$ is separated, there exist two disjoint open sets U_1 and U_2 containing θ_1 and θ_2, respectively. Given that g is open, $V_1 = g(U_1)$ and $V_2 = g(U_2)$ are open, and so $V_1 \cap V_2 \supset \{p\} \neq \emptyset$ is open in $\mathbb{R}^k$; by Theorem IV.3 in Hurewicz and Wallman (1948) then $\dim V_1 \cap V_2 = k$. In particular, $V_1 \cap V_2$ contains a point $q \in \mathbb{R}^k \backslash g(B)$; otherwise, $V_1 \cap V_2 \subseteq g(B)$ which would imply $\dim g(B) = k$ and is contradictory with $\dim g(B) < k-1$ shown in Step 2. Now, $g|_{\mathbb{R}^k \backslash g^{-1}(g(B))}$ being a homeomorphism from $\mathbb{R}^k \backslash g^{-1}(g(B))$ onto $\mathbb{R}^k \backslash g(B)$ is in contradiction with $U_1 \cap U_2 = \emptyset$. Hence, $\bar{g}$ is one-to-one, onto, continuous, and both open and closed; hence its inverse is also continuous, and $\bar{g}$ is a homeomorphism from $g^{-1}(g(B))$ onto $g(B)$.

Combining all of the above shows that g is a homeomorphism from $\mathbb{R}^k$ onto $\mathbb{R}^k$. The remainder of the proof is identical to that of Theorem 1 $\square$

References

Aliprantis, C. D., and K. C. Border (2007): *Infinite Dimensional Analysis*. Springer-Verlag, Berlin, 3rd edn.

Basye, R. E. (1935): "Simply Connected Sets," *Transactions of the American Mathematical Society*, 38, 341–356.

Benkard, C. L., and S. Berry (2006): "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on Brown (1983) and Rhoerig (1988)," *Econometrica*, 74, 1429–1440.

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

Berliocchi, H., and J.-M. Lasry (1973): "Intégrandes Normales et Mesures Paramétrées En Calcul Des Variations," *Bulletin de la S.M.F.*, 101, 129–184.

Billingsley, P. (1995): *Probability and Measure*. John Wiley & Sons.

Bowden, R. (1973): "The Theory of Parametric Identification," *Econometrica*, 41, 1069–1074.

Brown, B. W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175–196.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Chua, L. O., and Y.-F. Lam (1972): "Global Homeomorphism of Vector-Valued Functions," *Journal of Mathematical Analysis and Applications*, 39, 600–624.

Church, P. T. (1963): "Differentiable Open Maps on Manifolds," *Transactions of the American Mathematical Society*, 109, 87–100.

Church, P. T., and E. Hemmingsen (1960): "Light Open Maps on n-Manifolds," *Duke Mathematical Journal*, 27, 527–536.

Debreu, G. (1970): "Economies with a Finite Set of Equilibria," *Econometrica*, 38, 387–392.

Deimling, K. (1985): *Nonlinear Functional Analysis*. Springer-Verlag, Berlin.

Dominguez, M. A., and I. N. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601–1615.

Dufour, J.-M., and C. Hsiao (2008): "Identification," in *The New Palgrave Dictionary of Economics: Second Edition*, ed. by L. Blume, and S. Durlauf. Palgrave Macmillan, England.

Echenique, F., and I. Komunjer (2007): "Testing Models with Multiple Equilibria by Quantile Methods," Mimeo, Caltech and UCSD.

Fisher, F. M. (1961): "Identifiability Criteria in Nonlinear Systems," *Econometrica*, 29, 574–590.

——— (1965): "Identifiability Criteria in Nonlinear Systems: A Further Note," *Econometrica*, 33, 197–205.

——— (1966): *The Identification Problem in Economics*. McGraw-Hill, New York.

Gale, D., and H. Nikaidô (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

Galichon, A., and M. Henry (2006): "Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models," Manuscript, Harvard University and Columbia University.

Hurewicz, W., and H. Wallman (1948): *Dimension Theory*. Princeton University Press, Princeton.

Jovanovic, B. (1989): "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 57, 1431–1437.

Koopmans, T. C. (ed.) (1950): *Statistical Inference in Dynamic Economic Models*, vol. 10 of *Cowles Commission Monograph*. John Wiley & Sons, Inc.

Koopmans, T. C., and O. Reiersøl (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21, 165–181.

Matzkin, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics Volume*, ed. by R. F. Engle, and D. L. McFadden, pp. 2524–2559. North-Holland.

——— (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, forthcoming.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, pp. 2113–2247. Elsevier Science.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Palais, R. S. (1959): "Natural Operations on Differential Forms," *Transactions of the American Mathematical Society*, 92, 125–141.

——— (1970): "When Proper Maps are Closed," *Proceedings of the American Mathematical Society*, 24, 835–836.

Plastock, R. A. (1978): "Nonlinear Fredholm Maps of Index Zero and Their Singularities," *Proceedings of the American Mathematical Society*, 68, 317–322.

Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Sard, A. (1965): "Hausdorff Measure of Critical Images on Banach Manifolds," *American Journal of Mathematics*, 87, 158–174.

Severini, T. A., and G. Tripathi (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory*, 22, 258–278.

Titus, C. J., and G. S. Young (1952): "A Jacobian Condition for Interiority," *Michigan Mathematical Journal*, 1, 69–72.