**Title**

Global Identification In Nonlinear Semiparametric Models



**Author**
Komunjer, Ivana

**Publication Date**
2007-07-01

**2007-06**

# UNIVERSITY OF CALIFORNIA, SAN DIEGO

## DEPARTMENT OF ECONOMICS

Global Identification In Nonlinear Semiparametric Models

By

Ivana Komunjer
Department of Economics
University of California, San Diego

**DISCUSSION PAPER 2007-06**
**July 2007**

# GLOBAL IDENTIFICATION IN NONLINEAR SEMIPARAMETRIC MODELS

IVANA KOMUNJER

Abstract. This note derives primitive conditions for global identification in nonlinear simultaneous equations systems. Identification is semiparametric in the sense that the latent structural disturbance is only known to satisfy a number of orthogonality restrictions with respect to observed instruments. Our contribution to the literature on identification in a semiparametric context 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. Second, we provide primitive conditions under which a parameter value that solves those restrictions is unique.

**Keywords**: identification, structural systems, multiple equilibria, semiparametric models

## 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.

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

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

[1] A review of historical and recent developments on identification in economics can be found in Dufour and Hsiao (2008).

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, 2005), and Benkard and Berry (2007), 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), 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 the uniqueness of solutions to nonlinear systems of equations. While intuitive, this approach has not yet produced sufficient conditions for global identification that are not "*overly strong*".[2] Newey and McFadden (1994) remarked that, as a consequence, much of the related literature has adopted an approach in which identification is simply assumed.

This note makes two contributions to the literature on identification in a semiparametric context. First, we derive a set of unconditional moment restrictions that are the starting point for identification in nonlinear structural systems. Second, we provide primitive conditions under which a parameter value that solves those restrictions is unique. It is worth pointing out that our uniqueness results are global.

We consider nonlinear systems of simultaneous equations in which the distribution of latent disturbances and observed instruments is known to satisfy a set of orthogonality conditions. In Section 2 we show how these conditions give rise to moment restrictions on the distribution of the explanatory and dependent variables that are the starting point for identification. The derivation of the results is made non-trivial by the possible presence of multiple equilibria often found in models that are nonlinear. In particular, we relax the often used assumption that the structural system can be uniquely solved for the dependent variable.

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 non-negative (non-positive). When the system is continuously differentiable with respect to the structural parameter, this requirement is weaker than the full rank conditions

[2]We refer here to p.158 in Fisher (1966) as well as Rothenberg's (1971) discussion of Theorem 7.

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 a discussion of our results; proofs are found in Appendices.

## 2. Semiparametric Identification in Nonlinear Structural Models

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

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

The variables entering into these equations consist of: a set of observed dependent variables $Y \in \mathbb{R}^G$, a set of observed explanatory variables $X \in \mathbb{R}^K$, a structural parameter $\theta \in \Theta \subset \mathbb{R}^k$, and a set of latent variables $U \in \mathbb{R}^G$. For example, $U$ can be thought of as disturbance or unaccounted heterogeneity in the model. In what follows, we shall assume that $\theta$ is finite dimensional ($k < \infty$), and that $\rho : \mathbb{R}^{G+K} \times \Theta \to \mathbb{R}^G$ is a known mapping satisfying the following:

**Assumption A.** *For every $\theta \in \Theta$, the mapping $(y, x) \to \rho(y, x, \theta)$ is in $\mathcal{C}^1(\mathbb{R}^{G+K})$.*

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).

---

[3]Both Fisher's (1966) and Rothenberg's (1971) results exploit sufficient conditions for uniqueness of solutions to systems of nonlinear equations given in Gale and Nikaidô (1965). They require that the Jacobian be positive, and that the symmetric part of the derivative matrix of the system be positive semi-definite.

2.1. **Structure.** Say that $V \equiv (X', U')'$ is a random variable that takes values $v = (x', u')'$ in $\mathbb{R}^{K+G}$ and call $F_{XU}$ the associated measure. When for a given value of $\theta$, Equation (1) can be globally "solved" for $Y$ in terms of $X$ and $U$, then one can define (explicitly or implicitly) a single-valued map $Y = m(X, U, \theta)$ 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) $T : \mathbb{R}^{K+G} \to \mathbb{R}^{K+G}$ that is continuous. This leads to the usual definition of the image measure $F_{XY}$ of $W = T(V)$ on $\mathbb{R}^{K+G}$: $F_{XY} = F_{XU} \circ T^{-1}$. Hence, the distribution of the observables $Y$ and $X$ is generated by the structure $\mathcal{S} = (\theta, \rho, F_{XU})$.

When given $\theta$, $X$ and $U$ multiple solutions for $Y$ are possible in Equation (1), we no longer deal with a single-valued map from $V$ to $W$ but a correspondence $T : \mathbb{R}^{K+G} \rightrightarrows \mathbb{R}^{K+G}$. 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 $v = (x', u')' \in \mathbb{R}^{K+G}$ we shall let $\Gamma_v \equiv \{w \in \mathbb{R}^{K+G} : w = (x', y')' \text{ and } \rho(y, x, \theta) = u\}$. Then, the correspondence $T$ associates to every $v \in \mathbb{R}^{K+G}$ a set $\Gamma_v \subset \mathbb{R}^{K+G}$. 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 $\operatorname{Sel} T = \{t : \mathbb{R}^{K+G} \to \mathbb{R}^{K+G}$ Borel-measurable and such that $t(v) \in T(v)$ for almost every $v \in \mathbb{R}^{K+G}\}$ (see Jovanovic (1989), e.g.). That the set $\operatorname{Sel} T$ is nonempty is not always the case. We impose the following:

**Assumption B.** *For every* $(x, \theta) \in \mathbb{R}^K \times \Theta$, $\lim_{|y| \to \infty} [\rho(y, x, \theta)' y]/|y| = \infty$.

Assumption B ensures that given $(x, \theta)$ the mapping $y \mapsto \rho(y, x, \theta)$ is surjective on $\mathbb{R}^G$, i.e. that the inverse image by $\rho(\cdot, x, \theta)$ of any point in $\mathbb{R}^G$ is nonempty. The intuition behind this result is simple: say that $\rho$ is linear in $y$ so that for some positive definite $G \times G$ matrix $B_\theta$ we can write: $\rho(y, x, \theta) = B_\theta y + \tilde{\rho}(x, \theta)$. Since $\det B_\theta \neq 0$, $y \mapsto \rho(y, x, \theta)$ is surjective on $\mathbb{R}^G$. In addition, $[\rho(y, x, \theta)' y]/|y| =$

$B_\theta|y| + \tilde{\rho}(x,\theta)'[y/|y|]$, which goes to $\infty$ as $|y| \to \infty$. Assumption B states the same limit condition which proves to be sufficient for surjectivity whether or not $\rho$ is linear.

This property combined with the continuity requirement in A suffices to show that the set of measurable selections $\operatorname{Sel} T$ of $T$ 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* $\mathcal{S} = (\theta, t, F_{XU})$ *with* $t \in \operatorname{Sel} T$ *generates the distribution of the observables* $X$ *and* $Y$*.*

In particular, the image measure of the observables $X$ and $Y$ is then again obtained as: $F_{XY} = F_{XU} \circ t^{-1}$. Note that our construction of $F_{XY}$ 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 $F_{XU}$ is atomless, as shown by Jovanovic (1989).

2.2. **Identification Condition.** The structural parameter $\theta$ is said to be identifiable in $\Theta$ if every structure $\mathcal{S}^* = (\theta^*, t^*, F^*_{XU})$ whose characteristics are known to apply to $\mathcal{S} = (\theta, t, F_{XU})$ and which generates the same distribution of the observables $F_{XY}$ as $\mathcal{S}$ (i.e. is observationally equivalent to $\mathcal{S}$), satisfies $\theta^* = \theta$ (see Koopmans and Reiersøl (1950), Roehrig (1988), e.g.).

Here, we shall assume that $F_{XU}$ is known to satisfy:

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

where $A(X,\theta)$ is a $k \times G$ matrix of instruments which consists of functions of $X$ and $\theta$. The nature of the restrictions in Equation (2) is semiparametric: while the functional form of the distribution of the disturbance is left unspecified, a number of orthogonality conditions relating $X$, $U$ and $\theta$ are known to hold. Weaker than independence, such unconditional moment restrictions are typically found in models in which the structural parameter $\theta$ is to be estimated via Instrumental Variables (IV) methods.

Using the results of Proposition 1, the expectation of $A(X,\theta)U$ (computed under $F_{XU}$) can be related to that of $A(X,\theta)\rho(Y,X,\theta)$ (computed under $F_{XY}$). Since $t$ is Borel-measurable, we have $F_{XY} = F_{XU} \circ t^{-1}$. Then,

$$\begin{aligned} E[A(X,\theta)\rho(Y,X,\theta)] &= \int_{R^{K+G}} A(x,\theta)\rho(y,x,\theta)\,\mathrm{d}F_{XY}(x,y) \\ &= \int_{R^{K+G}} A(x,\theta)u\,\mathrm{d}F_{XU}(x,u) = E[A(X,\theta)U] \end{aligned}$$

by a simple change of variable $w = t(v)$ with $v = (x', u')'$ and $w = (x', y')'$. Under two structures $\mathcal{S}$ and $\mathcal{S}^*$ that are observationally equivalent we then have $E[A(X,\theta)\rho(Y,X,\theta)] = 0$ and $E[A(X,\theta^*)\rho(Y,X,\theta^*)] = 0$, where both expectations are taken with respect to $F_{XY}$. This leads to the following sufficient condition for semiparametric identification of $\theta$, that is valid in simultaneous equations systems (1) known to satisfy the unconditional moment restrictions in Equation (2):

**Theorem 1.** *Let Assumptions A and B hold. Assume that the observables $X$ and $Y$ are generated by a structure $\mathcal{S} = (\theta, t, F_{XU})$ where $\theta \in \Theta$, $t \in \operatorname{Sel} T$, and $F_{XU}$ satisfies $E[A(X,\theta)U] = 0$. Then $\theta$ is identified in $\Theta$ if $E[A(X,\theta)\rho(Y,X,\theta)] = 0$ has a unique solution $\theta_0$ in $\Theta$.*

## 3. Example and Intuition

Theorem 1 shows that in nonlinear structural models defined by unconditional moment restrictions, the conditions for parametric identification are like conditions for unique solutions of systems of nonlinear equations. 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.

Say that the parameter of interest $\theta$ is a scalar in $\Theta \subset \mathbb{R}$ and that the economic theory posits that the following moment condition holds: $E[A(X,\theta)\rho(Y,X,\theta)] = 0$ in which $A(X,\theta) \in \mathbb{R}$ is an instrument and $\rho : \mathbb{R}^2 \times \Theta \to \mathbb{R}$ is a structural map. To simplify the notation, let then

$$r(Y,X,\theta) \equiv A(X,\theta)\rho(Y,X,\theta)$$

The mapping $r : \mathbb{R}^2 \times \Theta \to \mathbb{R}$ (which is known) relates a dependent variable $Y \in \mathbb{R}$, an explanatory variable $X \in \mathbb{R}$, and the parameter $\theta$, in a possibly nonlinear fashion. We shall restrict our attention to those mappings that are continuously differentiable with respect to the parameter $\theta$ on the parameter set $\Theta$.

Using the results of Theorem 1, the parameter $\theta$ is identified if the equation $E[r(Y, X, \theta)] = 0$ has a unique solution $\theta_0 \in \Theta$. A simple way to guarantee identification in this case is to require that the mapping $\theta \mapsto E[r(Y, X, \theta)]$ be strictly monotone on $\Theta$. Then the equation $E[r(Y, X, \theta)] = 0$ either has a unique solution in $\theta$, or no solution exists. A sufficient condition for monotonicity is that $\partial E[r(Y, X, \theta)]/\partial\theta$ be positive (negative) on $\Theta$.

Example. Consider a simple nonlinear moment restriction $E[Y - \theta^2 X + \theta X^2] = 0$ taken from Example 2 in Dominguez and Lobato (2004). Here, $\partial E[r(Y, X, \theta)]/\partial\theta = E[-X(X + 2\theta)]$, provided we can exchange the orders of integration and derivation. If $E(X) = 0$ then $\theta$ is identified on $\Theta \equiv \mathbb{R}$. If on the other hand $E(X) \neq 0$, then identification holds on $\Theta_1 \equiv (-\infty, -E(X^2)/(2E(X))]$ and on $\Theta_2 \equiv [-E(X^2)/(2E(X)), +\infty)$.

While the discussion is simple in the case of a single parameter, complications arise when $\dim \Theta > 1$. In that case, the condition that the Jacobian $\det \mathrm{D}_\theta E[r(Y, X, \theta)] \neq 0$ on $\Theta$ no longer suffices to show that a solution to $E[r(Y, X, \theta)] = 0$ (when it exists) is unique on $\Theta$. A standard counterexample is the mapping $c : \mathbb{R}^2 \to \mathbb{R}^2$ which to each $(x_1, x_2) \in \mathbb{R}^2$ assigns $c(x_1, x_2) = (\exp x_1 \cos x_2, \exp x_1 \sin x_2)$. It is easy to check that the Jacobian of $c$ never vanishes, 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 the counterexample above by requiring that $\theta \mapsto E[r(Y, X, \theta)]$ be proper, i.e. that the inverse image of any compact set be compact. This condition is clearly violated by $c$ since for any $(y_1, y_2) \in \mathbb{R}^2 \backslash \{0\}$ the inverse image $c^{-1}(\{(y_1, y_2)\})$ is unbounded (hence not compact) in $\mathbb{R}^2$. Properness by itself does not guarantee that $\theta \mapsto E[r(Y, X, \theta)]$ is one-to-one on $\Theta$.

The latter is true, however, if one is willing to assume that in addition its Jacobian $\det \mathrm{D}_\theta E[r(Y,X,\theta)]$ never vanishes (see Corollary 4.3 in Palais (1959) e.g.).

In models that are nonlinear in $\theta$, everywhere non-vanishing Jacobian might be too strong of an assumption. It turns out, however, that restricting the Jacobian to be non-negative (non-positive) on $\Theta$ suffices to make the mapping $\theta \mapsto E[r(Y,X,\theta)]$ one-to-one on $\Theta$, provided its inverse images of individual points are of dimension zero, i.e. contain countably many points of $\Theta$. In particular, the latter requirement excludes the cases in which $\theta \mapsto E[r(Y,X,\theta)]$ is "flat" on subsets of $\Theta$ that have nonzero dimension.

Working with systems whose Jacobian possibly vanishes requires additional smoothness properties of the map $\theta \mapsto E[r(Y,X,\theta)]$. In particular, we shall assume the latter to be twice continuously differentiable on $\Theta$. This allows us to invoke an appropriate version of Sard's theorem when deriving our main result, to which we turn next.

## 4. Main Result

We now derive primitive conditions under which:

$$E[r(Y,X,\theta)] = 0 \text{ has a unique solution } \theta_0 \in \Theta \tag{3}$$

The mapping $r : \mathbb{R}^G \times \mathbb{R}^K \times \Theta \to \mathbb{R}^k$ is assumed known. The variables entering $r$ consist of: a set of dependent variables $Y \in \mathbb{R}^G$, a set of explanatory variables $X \in \mathbb{R}^K$, and a structural parameter of interest $\theta \in \Theta$ with $\Theta \subset \mathbb{R}^k$. In particular, if we let $r(Y,X,\theta) = A(X,\theta)\rho(Y,X,\theta)$, then according to Theorem 1 the property in Equation (3) is sufficient for semiparametric identification of $\theta$ in simultaneous equations systems (1) that satisfy (2).

To start our analysis, we need the following:

**Assumption C.** *(i) $\Theta \neq \emptyset$ is connected and open in $\mathbb{R}^k$; (ii) for any $\theta \in \Theta$, $E[r(Y,X,\theta)]$ exists and is finite.*

In particular, assumption C(ii) allows us to define a mapping $g : \Theta \to \mathbb{R}^k$ as $g(\theta) \equiv E[r(Y, X, \theta)]$. Connectedness of the parameter space in C(i) is crucial if we want to extend local results globally. Openness on the other hand allows us not to worry about the behavior of $g$ at the boundary. In applications in which $\Theta$ is closed, one needs to work with $\mathring{\Theta}$. Moreover, since $\Theta$ is a non-empty open subset of $\mathbb{R}^k$, necessarily $\dim \Theta = k$ (see Theorem IV.3 in Hurewicz and Wallman (1948) e.g.). Hereafter, we shall work with mappings $g$ that are twice continuously differentiable $g \in \mathcal{C}^2(\Theta)$. A sufficient condition is:

**Assumption D.** *(i) For every $(y, x) \in \mathbb{R}^{G+K}$, the mapping $\theta \mapsto r(y, x, \theta)$ is in $\mathcal{C}^2(\Theta)$; (ii) For every $\theta \in \Theta$ there exists $\delta > 0$ such that for every $(y, x, \theta') \in \mathbb{R}^{G+K} \times \Theta$, $|\theta' - \theta| < \delta$ implies that for every $1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant k$ and $1 \leqslant n \leqslant k$, we have: $\big|r^i(y, x, \theta') - r^i(y, x, \theta)]\big| \leqslant q_i^0(y, x, \theta) \cdot |\theta' - \theta|$, $\big|D_n r^i(y, x, \theta') - D_n r^i(y, x, \theta)]\big| \leqslant q_{ni}^1(y, x, \theta) \cdot |\theta' - \theta|$, and $\big|D_{nj}^2 r^i(y, x, \theta') - D_{nj}^2 r^i(y, x, \theta)]\big| \leqslant q_{nij}^2(y, x, \theta) \cdot |\theta' - \theta|$, with $E[q_i^0(Y, X, \theta)] < \infty$, $E[q_{ni}^1(Y, X, \theta)] < \infty$ and $E[q_{nij}^2(Y, X, \theta)] < \infty$.*

The above conditions are sufficient to show (via Lebesgue's dominated convergence theorem) that $g \in \mathcal{C}^2(\Theta)$. They are, however, not necessary. For example, when $G = 1$ and $K = k$, $r(Y, X, \theta) = \mathbb{I}(X'\theta - Y)$ fails to be continuously differentiable with respect to $\theta$ at any point in $\Theta$ satisfying $\theta' X = Y$.[4] Still letting $F_{Y|X}$ be the conditional distribution of $Y$ given $X$, and assuming the latter to be absolutely continuous (with respect to Lebesgue's measure) with continuously differentiable density on $\mathbb{R}^k$, we have $g(\theta) = E_X[F_{Y|X}(X'\theta)]$ in $\mathcal{C}^2(\Theta)$.

In cases in which it is possible to establish twice continuous differentiability of $g$ directly, assumption D may be dropped without altering the validity of our results. In what follows, we shall let $\mathrm{D}g \in L(\mathbb{R}^k, \mathbb{R}^k)$ denote the derivative of $g$ and $J_g(\theta)$ its Jacobian at $\theta$, $J_g(\theta) = \det \mathrm{D}g(\theta)$.

Next, we require that the mapping $g$ be proper, i.e. that the inverse image by $g$ of each compact subset of $g(\Theta)$ be a compact subset of $\Theta$. A sufficient condition is:

[4] $\mathbb{I}$ is the Heaviside function: $\mathbb{I}(x) = 1$ if $x \geqslant 0$ and 0 otherwise.

**Assumption E.** *Either* $\Theta$ *bounded, or* $\Theta$ *unbounded and* $\left|E[r(Y,X,\theta)]\right| \to \infty$ *whenever* $|\theta| \to \infty$.

This property of maps is crucial in ensuring that local homeomorphic properties of maps become global. It is worth pointing out that assuming a map is locally homeomorphic by itself does not ensure that it is a homeomorphism: previously defined mapping $c : (x_1, x_2) \mapsto (\exp x_1 \cos x_2, \exp x_1 \sin x_2)$ is a counterexample. We are now ready to state our main result:

**Theorem 2.** *Let assumptions C, D and E hold. If for every* $p \in \mathbb{R}^k$ *the equation* $E[r(Y,X,\theta)] = p$ *has countably many (possibly zero) solutions in* $\Theta$*, and if* $\det E[\mathrm{D}_\theta r(Y,X,\theta)]$ *is non-negative (non-positive) on* $\Theta$*, then* $E[r(Y,X,\theta)] = 0$ *either has no solution in* $\Theta$ *or has a unique solution* $\theta_0 \in \Theta$.

The requirement that $E[r(Y,X,\theta)] = p$ 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 $\theta' \in \Theta$ such that $\det E[\mathrm{D}_\theta r(Y,X,\theta')] \neq 0$) then the set of solutions to $E[r(Y,X,\theta)] = p$ is finite (possibly empty).[5]

We note that the condition on the non-negativity (non-positivity) of the Jacobian of $E[r(Y,X,\cdot)]$ is a weakening of the condition that $\det E[\mathrm{D}_\theta r(Y,X,\theta)] \neq 0$ on $\Theta$; the latter, combined with the continuity and properness assumptions D and E, is known to guarantee that $E[r(Y,X,\cdot)]$ is a homeomorphism from $\Theta$ onto $E[r(Y,X,\Theta)]$ (see Corollary 4.3 in Palais (1959) e.g.).

## 5. Discussion and Conclusion

From Theorem 1 it follows that the semiparametric identification of the structural parameter $\theta$ is a combined property of the structural mapping $\rho(Y,X,\theta)$ in Equation

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

(1), the instruments $A(X,\theta)$ in Equation (2), and the parameter set $\Theta$. While the functional form of $\rho$ is often dictated by an economic theory, Theorem 2 can aid the econometrician in her choice of valid instruments and point identified parameter sets.

For example, consider a linear structural system $\rho(Y,X,\theta) = Y - X'\theta$, in which $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$ and $\theta \in \Theta$ open and connected in $\mathbb{R}^k$, with instrument $A(X,\theta) = X$. Using Theorem 2, the parameter set $\Theta$ is point identified provided $\det E[XX'] \geqslant 0$ and $\dim(\operatorname{Ker} E[r(Y,X,\cdot)] - p) = 0$, where $r(Y,X,\theta) = XY - XX'\theta$ and $p \in \mathbb{R}$. Given linearity of $r(Y,X,\cdot)$, the latter is equivalent to $\dim(\operatorname{Im} E[r(Y,X,\cdot)] - p) = k$, i.e. $\det E[XX'] \neq 0$. Hence, Theorem 2 reduces to the usual full rank condition $\det E[XX'] > 0$ which is sufficient for the point identification of $\Theta$.

In structural systems that are nonlinear in $\theta$ yet with no "flats" on $\Theta$, instruments are valid provided the Jacobian condition $\det E[\mathrm{D}_\theta r(Y,X,\theta)] \geqslant 0$ (or $\leqslant 0$) holds. This condition is weaker than the requirement $\det E[\mathrm{D}_\theta r(Y,X,\theta)] > 0$ imposed in standard asymptotic normality proofs for nonlinear IV estimators of $\theta$ (see Amemiya (1977), e.g.). In particular, assuming $\det E[\mathrm{D}_\theta r(Y,X,\theta)] > 0$ is not necessary for identification.[6] While convenient, the primitive conditions of Theorem 2 are still likely to lead to more than one parameter set that is point identified; this was the case in our example in Section 3. In such cases, additional economically meaningful restrictions on $\theta$ have to be used to discriminate between different solutions.

## References

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

Amemiya, T. (1977): "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955–968.

[6]A simple counterexample is $E[r(Y,X,\theta)] = \theta^3$.

Benkard, C. L., and S. Berry (2007): "On the Nonparametric Identification of Non-Linear Simultaneous Equations Models: Comment on Brown (1983) and Rhoerig (1988)," *Econometrica*, forthcoming.

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.

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.

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

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.

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.

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.

——— (2005): "Identification in Nonparametric Simultaneous Equations," mimeo, Northwestern University.

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.

RUDIN, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill.

SARD, A. (1942): "The Measure of the Critical Values of Differentiable Maps," *Bulletin of the American Mathematical Society*, 48, 883–890.

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

SMALE, S. (1965): "An Infinite Dimensional Version of Sard's Theorem," *American Journal of Mathematics*, 87, 861–866.

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

## APPENDIX A. PROOF OF PROPOSITION 1

The proof is done in four steps.

STEP 1: We first show that, given $\theta \in \Theta$, the correspondence $T$ is closed-valued, i.e. for any $v \in \mathbb{R}^{K+G}$, $T(v) = \Gamma_v$ is a closed subset of $\mathbb{R}^{K+G}$.

Fix $\theta \in \Theta$ and let $h_\theta : \mathbb{R}^{K+G} \to \mathbb{R}^{K+G}$ be such that for any $w \equiv (x', y')' \in \mathbb{R}^{K+G}$ we have $h_\theta(w) = v$ where $v = (x', u')'$ and $u = \rho(y, x, \theta)$. Since by assumption A, $h_\theta \in \mathcal{C}(\mathbb{R}^{K+G})$ and

$$\frac{h_\theta(w)'w}{|\, w|} = \frac{|x|^2 + \rho(y,x,\theta)'y}{\sqrt{|x|^2 + |y|^2}} \geqslant \frac{|x|^2 + \rho(y,x,\theta)'y}{2\max\{|x|, |y|\}}$$

we have by assumption B, $\lim_{|w|\to\infty}[h_\theta(w)'w]/|\, w| = \infty$. By Theorem 3.3 in Deimling (1985), we then have that $h_\theta$ is surjective, i.e. $h(\mathbb{R}^{K+G}) = \mathbb{R}^{K+G}$. We now show that $h$ is also proper, i.e. that the inverse image by $h$ of each compact subset of $h(\mathbb{R}^{K+G})$ is compact in $\mathbb{R}^{K+G}$. For this, note that

$$|h_\theta(w)'w|/|\, w| \leqslant |h_\theta(w)|$$

so assumption B also implies $\lim_{|w|\to\infty} |h_\theta(w)| = \infty$. Let then $K \subset h(\mathbb{R}^{K+G})$ be compact. Since $h(\mathbb{R}^{K+G}) = \mathbb{R}^{K+G}$, $K$ is compact if and only if it is closed and bounded. By A we know that $h$ is continuous, hence $h^{-1}(K)$ is closed in $\mathbb{R}^{K+G}$. To show it is bounded, consider a sequence $\{h(w_n)\}$ in $K$. Since $K$ is compact, $h(w_n) \to h(w_0) \in K$, which by $\lim_{|w|\to\infty} |h_\theta(w)| = \infty$ implies that the sequence $\{w_n\}$ is bounded. Hence, $h^{-1}(K)$ is bounded, therefore compact in $\mathbb{R}^{K+G}$.

We can now show that the correspondence $T$ is closed-valued. Take any $v \in \mathbb{R}^{K+G}$. Since $\mathbb{R}^{K+G} = h(\mathbb{R}^{K+G})$, we have $\Gamma_v = h_\theta^{-1}(v)$ which by properness of $h_\theta$ is compact, hence closed in $\mathbb{R}^{K+G}$.

STEP 2: We next show that, given $\theta \in \Theta$, 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 $\mathbb{R}^{K+G}$ 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 Proposition 2 in Berliocchi and Lasry (1973)). Here, $\mu$ is a $K + G$ dimensional Lebesgue's measure. Consider the graph of the correspondence $T|_K$, $\mathrm{Gr}\,(T|_K) = \{(V, W) \in \mathbb{R}^{2(K+G)} : V \in K, W \in \Gamma_V\}$. We need to show that $\mathrm{Gr}\,(T|_K)$ is a closed subset of $\mathbb{R}^{2(K+G)}$. We know that $\mathrm{Gr}\,(T|_K)$ is closed if and only if for any sequence $\{(v_n', w_n')'\}$ in $\mathrm{Gr}\,(T|_K)$, $v_n \to a$, $w_n \to b$ imply that $(a', b')' \in \mathrm{Gr}\,(T|_K)$. Take then $\{(v_n', w_n')'\}$ in $\mathrm{Gr}\,(T|_K)$. By continuity of $h$, $w_n \to b$ implies $v_n = h(w_n) \to h(b) = a$. So, $b \in \Gamma_a$. Since $K$ is compact, $a \in K$, therefore $(a, b) \in \mathrm{Gr}\,(T|_K)$.

STEP 4: Finally, we can show that $\mathrm{Sel}\, T \neq \emptyset$. This is an immediate consequence of a corollary to Kuratowski's theorem: *if a correspondence $T : \mathbb{R}^{K+G} \rightrightarrows \mathbb{R}^{K+G}, V \to \Gamma_V$ is Borel-measurable, closed-valued and such that $\Gamma_V \neq \emptyset$ for almost every $V$, then* $\mathrm{Sel}\, T \neq \emptyset$ (see Corollary 1 in Berliocchi and Lasry (1973), or Theorem 18.13 in Aliprantis and Border (2007), e.g.). Note that $\Gamma_v \neq \emptyset$ for every $v \in \mathbb{R}^{K+G}$ by surjectivity of the map $h_\theta$. Hence, $W = (X', Y')'$ exists.

## Appendix B. Proof of Theorem 2

We proceed in six steps. The first step may be omitted if twice continuous differentiability of $g$ has been established directly.

STEP 1: We start by showing that under assumptions C-D the mapping $g$ is in $\mathcal{C}^2(\Theta)$.

Recall that $g$ maps the open set $\Theta \subset \mathbb{R}^k$ into $\mathbb{R}^k$. Then $g \in \mathcal{C}^1(\Theta)$ if and only if the partial derivatives $D_n g^i$ exist and are continuous on $\Theta$ for $1 \leqslant i \leqslant k$ and $1 \leqslant n \leqslant k$ (see Theorem 9.21 in Rudin (1976) e.g.). For $\theta \in \Theta$, $1 \leqslant i \leqslant k$ and $1 \leqslant n \leqslant k$ we have:

$$D_n g^i(\theta) = \lim_{t \to 0} \frac{g^i(\theta + t\, e_n) - g^i(\theta)}{t}$$

provided the limit exists, where $g^i$ is the $i$th component of $g$ and $\{e_1, \ldots, e_k\}$ the standard basis of $\mathbb{R}^k$. For $(y, x) \in \mathbb{R}^{G+K}$, let $r^i(y, x, \cdot)$ denote the $i$th component of $r(y, x, \cdot)$. From definition of $g$, we have

$$\begin{aligned} \left| \frac{g^i(\theta + t\, e_n) - g^i(\theta)}{t} \right| &\leqslant \int_{\mathbb{R}^{K+G}} \left| \frac{r^i(y, x, \theta + t\, e_n) - r^i(y, x, \theta)}{t} \right| dF_{XY}(x, y) \\ &\leqslant \int_{\mathbb{R}^{K+G}} q_i^0(y, x, \theta) dF_{XY}(x, y) \end{aligned}$$

provided $|t| < \delta$, where the second inequality uses the first condition in assumption D(ii). Now, using the facts that $r(y, x, \cdot) \in \mathcal{C}^1(\Theta)$ from D(i), and $\int_{\mathbb{R}^{K+G}} q(y, x, \theta) dF_{XY}(x, y) < \infty$, yields by Lebesgue's dominated convergence theorem (see Theorem 11.32 in Rudin (1976) e.g.)

$$\begin{aligned} \lim_{t \to 0} \frac{g^i(\theta + t\, e_n) - g^i(\theta)}{t} &= \int_{\mathbb{R}^{K+G}} \lim_{t \to 0} \left[ \frac{r^i(y, x, \theta + t\, e_n) - r^i(y, x, \theta)}{t} \right] dF_{XY}(x, y) \\ &= \int_{\mathbb{R}^{K+G}} D_n r^i(y, x, \theta) dF_{XY}(x, y) \end{aligned}$$

where $D_n r^i$ denotes the partial derivative of $r^i(y, x, \cdot)$ with respect to the $n$th component of $\theta$. Thus $D_n g^i$ exists and equals

$$D_n g^i(\theta) = \int_{\mathbb{R}^{K+G}} D_n r^i(y, x, \theta) dF_{XY}(x, y) \tag{4}$$

for every $\theta \in \Theta$. $D_n g^i$ is continuous if to every $\theta \in \Theta$ and to every $\varepsilon > 0$ corresponds a $\delta' > 0$ such that $|D_n g^i(\theta) - D_n g^i(\theta')| < \varepsilon$ if $\theta' \in \Theta$ and $|\theta - \theta'| < \delta'$. Using Equation

(4) we have:

$$\begin{aligned}|D_n g^i(\theta) - D_n g^i(\theta')| &\leqslant \int_{\mathbb{R}^{K+G}} |D_n r^i(y,x,\theta) - D_n r^i(y,x,\theta')| dF_{XY}(x,y) \\ &\leqslant |\theta - \theta'| \int_{\mathbb{R}^{K+G}} q^1_{ni}(y,x,\theta) dF_{XY}(x,y)\end{aligned}$$

provided $|\theta - \theta'| < \delta$, where the second inequality uses the second condition in assumption D(ii). Now let $M_\theta \equiv \int_{\mathbb{R}^{K+G}} q^1_{ni}(y,x,\theta) dF_{XY}(x,y) < \infty$. Then for any $\varepsilon > 0$ letting $\delta' \equiv \min\{\delta, \varepsilon/(2M_\theta)\}$ yields $g \in \mathcal{C}^1(\Theta)$.

Applying the same reasoning as above with the functions $D_n g^i$ instead of the functions $r^i$ shows that all the second order partial derivatives $D^2_{nj} g^i$ ($1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant k$, $1 \leqslant n \leqslant k$) exist and are continuous on $\Theta$; hence (by Definition 9.39 in Rudin (1976)) $g \in \mathcal{C}^2(\Theta)$.

STEP 2: Next, we show that under assumptions C, D and E, the mapping $g$ is proper, i.e. that the inverse image by $g$ of each compact subset of $g(\Theta)$ be a compact subset of $\Theta$. The proof is straightforward: let $K \subset g(\Theta)$ be compact. Since $g(\Theta) \subset \mathbb{R}^k$, $K$ is compact if and only if it is closed and bounded. Given that $g$ is continuous, $g^{-1}(K)$ is closed in $\Theta$. It remains to be shown that it is bounded. When $\Theta$ is bounded, $g^{-1}(K) \subset \Theta$ is bounded. When $\Theta$ is unbounded, E implies $|g(\theta)| \to \infty$ whenever $|\theta| \to \infty$. Let $(g(\theta_n))$ be a sequence in $K$. Since $K$ is compact, $g(\theta_n) \to g(\theta_0) \in K$, which by assumption E implies that $(\theta_n)$ is bounded. Hence, $g^{-1}(K)$ is compact in $\Theta$.

Before continuing, let us note that a continuous proper map $g$ is also closed, i.e. $g(B)$ closed whenever $B \subset \Theta$ closed (see Corollary in Palais (1970) e.g.). That $g$ is closed is, perhaps surprisingly, not equivalent to $g$ being open, i.e. $g(B)$ open whenever $B \subset \Theta$ open. Having $g$ closed or open are two possible ways of defining the continuity of its inverse $g^{-1}$; the two are generally different and coincide only if the mapping $g$ is one-to-one.

STEP 3: If $J_g(\theta) = 0$ then $\theta$ is a critical point of $g$; the set of all such points is called the critical set of $g$, $B_g = \{\theta \in \Theta : J_g(\theta) = 0\}$. The image by $g$ of the critical set is the set of singular values of $g$, $S_g = g(B_g)$. We are now ready to prove the following intermediate result:

**Lemma 1.** *Let assumptions C, D and E hold. If* $\Theta \backslash g^{-1}(S_g)$ *connected, then* $g$ *is a homeomorphism from* $\Theta \backslash g^{-1}(S_g)$ *onto* $g(\Theta)\backslash S_g$.

First, we shall establish that $\tilde{g} = g|_A$ where $A = \Theta \backslash g^{-1}(S_g)$ is a local homeomorphism. We have $\tilde{g} : A \to g(\Theta)\backslash S_g$ and since $g^{-1}(S_g) \supset B_g$ the critical set of $\tilde{g}$ is $B_{\tilde{g}} = \emptyset$. Since $g \in \mathcal{C}^1(\Theta)$, we also have $\tilde{g} \in \mathcal{C}^1(A)$. Then by the inverse function theorem (see Theorem 9.24 in Rudin (1976) e.g.), $\tilde{g}$ is a local homeomorphism on $\mathring{A}$.

Next, we show that $\tilde{g}$ is proper: let $C$ be a compact in $g(\Theta)\backslash S_g$ and note that $\tilde{g}^{-1}(C) = g^{-1}(C)$ since $\tilde{g}^{-1} = g^{-1}|_{g(\Theta)\backslash S_g}$. Then $C$ is compact in $g(\Theta)$ so by properness of $g$ we have that $g^{-1}(C)$ is compact in $\Theta$. Since $C \cap S_g = \emptyset$ it follows that $g^{-1}(C) \cap g^{-1}(S_g) = \emptyset$ and so $g^{-1}(C)$ is compact in $A$. Same reasoning as above shows that $\bar{g}|_{g^{-1}(S_g)}$ is proper.

Finally, we show that $A$ is open. Consider $\theta \in \Theta \backslash B_g$. Then $J_g(\theta) \neq 0$ and the inverse function theorem applied to $g$ implies that there exists an open neighborhood $U$ of $\theta$ on which $g$ is a local homeomorphism; so $U \cap B_g = \emptyset$ and $U \subset \Theta \backslash B_g$, which shows that $\Theta \backslash B_g$ is open. Since $\Theta$ is open $B_g$ is closed. Using our previous observation that a continuous proper map is closed, we know that $\bar{g}$ is closed. Since $B_g \subset g^{-1}(S_g)$ we have that $S_g = \bar{g}(B_g)$ is closed. Continuity of $g$ then guarantees that $g^{-1}(S_g)$ is closed as well, thus $A$ is open.

The result of Lemma 1 follows by the global homeomorphism theorem: *if* $A \subset \mathbb{R}^k$ *is open connected,* $\tilde{g} : A \to \mathbb{R}^k$ *a local homeomorphism and* $\tilde{g}$ *proper, then* $\tilde{g}$ *is a homeomorphism onto* $\tilde{g}(A)$ (see Exercise 4.3 in Deimling (1985) e.g.).

The requirement that $\Theta \backslash g^{-1}(S_g)$ be connected is by no means trivial to satisfy. For example, consider the case $\Theta = \mathbb{R}$. It is well known that the only connected sets in $\mathbb{R}$ are the intervals; then necessarily $g^{-1}(S_g) = (-\infty, \lambda\rangle \cup \langle \mu, +\infty)$ where $(\lambda, \mu) \in \bar{\mathbb{R}}$ and $\langle$ denotes either ( or [ with analogous definition for $\rangle$. So $S_g = g((-\infty, \lambda\rangle) \cup g(\langle \mu, +\infty))$, which by continuity of $g$ and assumption E equals $S_g =$

$(-\infty, \sup_{(-\infty,\lambda\rangle} g] \cup [\inf_{\langle\mu,+\infty)} g, +\infty)$. Now, from Sard's lemma (Sard, 1942) we know that when $g \in \mathcal{C}^1(\mathbb{R})$ its singular set $S_g$ has Lebesgue measure zero, so that necessarily $\lambda = -\infty$ and $\mu = +\infty$; then $g^{-1}(S_g) = \emptyset$ which implies $B_g = \emptyset$, so $g$ is necessarily monotone on $\mathbb{R}$.

In a sense, when the dimension of the parameter set $\Theta$ equals 1 then removing single points from $\Theta$ suffices to make the resulting set disconnected. In dimensions $k > 1$, there is still hope that the set $g^{-1}(S_g)$ be "small enough" for $\Theta \backslash g^{-1}(S_g)$ to remain connected.

Before proceeding, we note that Lemma 1 can also be shown to hold by using the theory of space covering maps (see Palais (1959) e.g.); this approach has been taken in Plastock (1978). An alternative proof is as follows. First, note that under assumption D $g$ is a $\mathcal{C}^1$ Fredholm map with index zero: its derivative $\mathrm{D}g$ is a linear operator from $\mathbb{R}^k$ to $\mathbb{R}^k$, so $\dim(\mathrm{Ker}\,\mathrm{D}g) + \dim(\mathrm{Im}\,\mathrm{D}g) = k$. Letting the index of $g$ be defined as: $\mathrm{Index}\, g \equiv \dim(\mathrm{Ker}\,\mathrm{D}g) - \dim(\mathrm{Coker}\,\mathrm{D}g)$, where $\mathrm{Coker}\,\mathrm{D}g$ is the quotient space $\mathbb{R}^k/\mathrm{Im}\,\mathrm{D}g$, we then have the simple equality: $\mathrm{Index}\, g = k - k = 0$. So $g$ is a $\mathcal{C}^1$ Fredholm map with index 0 (Smale, 1965). Next, we rely on the results of Theorem 3 in Plastock (1978): *if $g : \Theta \to \mathbb{R}^k$ is a $\mathcal{C}^1$ proper Fredholm map of index 0, and if $g(\Theta) \neq S_g$ and $g(\Theta)\backslash S_g$ connected, then $\tilde{g} = g|_{\Theta\backslash g^{-1}(S_g)}$ is a covering space map.* Assumption E implies properness of $g$; continuity of $g$ implies $g(\Theta)\backslash S_g$ connected whenever $\Theta\backslash g^{-1}(S_g)$ connected; trivially, $0 \in g(\Theta)\backslash S_g$ implies $g(\Theta) \neq S_g$. Using Theorem 3 in Plastock (1978) together with connectedness of the domain $\Theta\backslash g^{-1}(S_g)$ of $\tilde{g}$ then yields the result of Lemma 1.

STEP 4: We show the following lemma:

**Lemma 2.** *Let assumptions C, D and E hold. If for every $\theta \in B_g$ , $\mathrm{rank}\,\mathrm{D}g(\theta) < k - 1$, and for every $p \in S_g$ , $\dim g^{-1}(p) = 0$, then $\Theta\backslash g^{-1}(S_g)$ connected.*

Our proof relies on the following dimension result: *any connected $k$-dimensional set $\Theta$ in $\mathbb{R}^k$ cannot be disconnected by a subset of dimension $< k - 1$* (see Theorem IV.4 in Hurewicz and Wallman (1948)).

To ensure that the inverse image $g^{-1}(S_g)$ of the set of critical values is of dimension $< k-1$, we combine the following two results: Sard's lemma and a dimension lowering theorem for closed maps. By Theorem 2 in Sard (1965) we know that when $g \in \mathcal{C}^2$, the image by $g$ of all the critical points whose rank is $< k-1$ is of dimension $< k-1$.

In order derive a similar result for the image by $g$ of its singular set, we apply Theorem VI.7 in Hurewicz and Wallman (1948) to the map $\bar{g} : g^{-1}(S_g) \to S_g$ defined previously. Recall that we established $\bar{g}$ closed. Then, if for every point $p$ in $S_g$, the inverse image $g^{-1}(p)$ is of dimension zero, i.e. countable, then $\dim g^{-1}(S_g) \leqslant \dim S_g$ and so $\dim g^{-1}(S_g) < k-1$, as desired.

STEP 5: Note that from the definition of $B_g$, we have $\operatorname{rank} \mathrm{D}g(\theta) \leqslant k-1$. Lemma 2 requires, however, that the latter inequality be strict. We now try to replace this requirement with a more intuitive condition on the Jacobian of $g$.

Recall that under assumptions C-D, $J_g$ is a continuous function on $\Theta$. If $\Theta \backslash g^{-1}(S_g)$ is connected, then $J_g(\theta)$ is necessarily of the same sign ($> 0$ or $< 0$) throughout $\Theta \backslash g^{-1}(S_g)$; otherwise, there would exist a point $\theta' \in \Theta \backslash g^{-1}(S_g)$ such that $J_g(\theta') = 0$, which is impossible given our definition of $B_g$.

The natural question then is whether having $J_g$ non-negative (non-positive) on $\Theta$ in turn implies $\Theta \backslash g^{-1}(S_g)$ connected. We establish the following:

**Lemma 3.** *Let assumptions C, D and E hold. If $J_g$ is (non-negative) non-positive on $\Theta$, and for every $p \in g(\Theta)$ , $\dim g^{-1}(p) = 0$, then: (i) $\dim g^{-1}(S_g) \leqslant \dim S_g < k-1$, and (ii) $g$ is a homeomorphism from $\Theta \backslash g^{-1}(S_g)$ onto $g(\Theta) \backslash S_g$.*

In the proof of Lemma 3, we shall again combine two results. First is: *every $g : \Theta \to \mathbb{R}^k$ (with $\Theta$ open in $\mathbb{R}^k$) of class $\mathcal{C}^1$, such that $\dim g^{-1}(p) = 0$ for every $p \in g(\Theta)$, and whose Jacobian $J_g$ is non-negative (non-positive) on $\Theta$, is open* (see Theorem 2 in Titus and Young (1952) e.g.). Second is an extension of the inverse function theorem for open maps: *if $g : \Theta \to \mathbb{R}^k$ (with $\Theta$ open in $\mathbb{R}^k$) of class $\mathcal{C}^1$ is open, then for every $\theta \in B_g$ we have* $\operatorname{rank} \mathrm{D}g(\theta) < k-1$ (see Theorem 1.4 in Church

(1963) e.g.). Together with Lemmas 1 and 2, the above results show that Lemma 3 holds.

STEP 6: We are now ready to prove Theorem 2.

When $g$ is open, then not only is its restriction to $\Theta\backslash g^{-1}(S_g)$ a homeomorphism onto $g(\Theta)\backslash S_g$, but the same holds for its restriction to $g^{-1}(S_g)$. By construction, $\bar{g} : g^{-1}(S_g) \to S_g$ is onto. We now show that it is also one-to-one: let $p \in S_g$ and assume that $g^{-1}(p) \supset \{\theta', \theta''\}$ with $\theta' \neq \theta''$. Since $\mathbb{R}^k$ is separated, there exist two disjoint open sets $U'$ and $U''$ containing $\theta'$ and $\theta''$, respectively. Given that $g$ is open, $V' = g(U')$ and $V'' = g(U'')$ are open, and so $V' \cap V'' \supset \{p\} \neq \emptyset$ is open in $g(\Theta)$ which is itself a non-empty open subset of $\mathbb{R}^k$; by Theorem IV.3 in Hurewicz and Wallman (1948) then $\dim V' \cap V'' = \dim g(\Theta) = k$. In particular, $V' \cap V''$ contains a point $q \in g(\Theta)\backslash S_g$; otherwise, $V' \cap V'' \subset S_g$ which would imply $\dim S_g = k$ and is contradictory with $\dim S_g < k - 1$. Now, $\tilde{g}$ being homeomorphic from $\Theta\backslash g^{-1}(S_g)$ onto $g(\Theta)\backslash S_g$ is in contradiction with $U' \cap U'' = \emptyset$.

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 $g^{-1}(S_g)$ onto $S_g$. This completes the proof of Theorem 2.