Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models[1]

Richard Blundell
UCL and IFS

and

Rosa L. Matzkin
UCLA

First version: March 2008
This version: October 2010

Abstract

The control function approach (Heckman and Robb (1985)) in a system of linear simultaneous equations provides a convenient procedure to estimate one of the functions in the system using reduced form residuals from the other functions as additional regressors. The conditions on the structural system under which this procedure can be used in nonlinear and nonparametric simultaneous equations has thus far been unknown. In this note, we define a new property of functions called *control function separability* and show it provides a complete characterization of the structural systems of simultaneous equations in which the control function procedure is valid.
Key Words: Nonseparable models, Simultaneous equations, control functions. **JEL Classification:** C3.

[1] **Acknowledgements:** We are grateful to Whitney Newey, Arthur Lewbel, and participants at the Cemmap-Cowles Conference on "Advancing Applied Microeconometrics", in Beijing, August 2010, for helpful comments. The research is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Funding from the ESRC is gratefully acknowledged. Rosa L. Matzkin gratefully acknowledges as well support from NSF. Yong Hyeon Yang provided excellent research assisstance. The usual disclaimer applies.

1 Introduction

A standard situation in applied econometrics is where one is interested in estimating a nonseparable model of the form

$$y_1 = m^1 (y_2, \varepsilon_1)$$

when it is suspected or known that y_2 is itself a function of y_1. Additionally there is an observable variable x, which might be used as an instrument for the estimation of m^1. Specifically, one believes that for some function m^2 and unobservable ε_2,

$$y_2 = m^2 (y_1, x, \varepsilon_2).$$

The nonparametric identification and estimation of m^1 under different assumptions on this model has been studied in Roehrig (1988), Newey and Powell (1989, 2003), Brown and Matzkin (1998), Darrolles, Florens, and Renault (2002), Ai and Chen (2003), Hall and Horowitz (2003), Benkard and Berry (2004, 2006), Chernozhukov and Hansen (2005), and Matzkin (2005, 2008, 2010a) among others (see Blundell and Powell (2003), Matzkin (2007), and many others, for partial surveys).

If the model were linear and with additive unobservables, one could estimate m^1 by first estimating a reduced form function for y_2, which would also turn out to be linear,

$$y_2 = h^2 (x, \eta) = \gamma x + \eta,$$

and then using η as an additional conditioning variable in the estimation of m^1, an idea dating back to Telser (1964).[2]

[2] Heckman (1978) references this paper in his comprehensive discussion of estimating simultaneous models with discrete endogenous variables. Blundell and Powell (2003) note that it is difficult to locate a definitive early reference to the control function version of

If the structural model were triangular, in the sense that y_1 is not an argument in m^2, a generalized version of this procedure could be applied to nonparametric, nonadditive versions of the model, as developed in Chesher (2003) and Imbens and Newey (2009). Their control function methods can be used in the triangular structural model

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= s(x, \eta) \end{aligned}$$

when x independent of (ε_1, η), m^1 strictly increasing in ε_1, and s strictly increasing in the unobservable η.

When the simultaneous model cannot be expressed in a triangular form, one can consider alternative restrictions in the joint distribution of $(\varepsilon_1, \varepsilon_2)$ and use the estimation approach in Matzkin (2010a), or one can assume that ε_1 is independent of x and use the instrumental variable estimator, see Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007).[3]

The question we aim to answer is the following: Suppose that we were interested in estimating the function m^1 when the structural model is of the form

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= m^2(y_1, x, \varepsilon_2) \end{aligned}$$

2SLS. Dhrymes (1970, equation 4.3.57) shows that the 2SLS coefficients can be obtained by a least squares regression of y_1 on $\widehat{y}_2$ and $\widehat{\eta}$, while Telser (1964) shows how the seemingly unrelated regressions model can be estimated by using residuals from other equations as regressors in a particular equation of interest.

[3]Unlike the control function approach and the Matzkin approach, the instrumental variable estimator requires dealing with the ill-posed inverse problem. (See also Hahn and Ridder (2010) regarding estimation of nonseparable models using conditional moment restrictions.)

and x is independent of $(\varepsilon_1, \varepsilon_2)$. Under what conditions on m^2 can we do this by first estimating a function for y_2 of the type

$$y_2 = s(x, \eta)$$

and then using η as an additional conditioning variable in the estimation of m^1?

More specifically, we seek an answer to the question: Under what conditions on m^2 is it the case that the simultaneous equations *Model (S)*

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= m^2(y_1, x, \varepsilon_2) \end{aligned}$$

with x independent of $(\varepsilon_1, \varepsilon_2)$, is observationally equivalent to the triangular *Model (T)*

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= s(x, \eta) \end{aligned}$$

with x independent of (ε_1, η)?

In what follows we first define a new property of functions, *control function separability*. We then show, in Section 3, that this property completely characterizes systems of simultaneous equations where a function of interest can be estimated using a control function. An example of a utility function whose system of demand functions satisfies control function separability is presented in Section 4 and illustrates the restrictiveness of the CF assumptions.

Section 5 describes how to extend our results to Limited Dependent Variable models with simultaneity in latent or observable continuous variables.

The Appendix provides conditions in terms of the derivatives of the structural functions in the system and conditions in terms of restrictions on the reduced form system. Section 6 concludes.

2 Assumptions and Definitions

2.1 The structural model and control function separability

We will consider the structural model

$$\text{Model (S)} \qquad \begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= m^2(y_1, x, \varepsilon_2) \end{aligned}$$

satisfying the following assumptions.

Assumption S.1 (differentiability): *For all values* $y_1, y_2, x, \varepsilon_1, \varepsilon_2$ *of* $Y_1, Y_2, X, \varepsilon_1, \varepsilon_2$, *the functions* m^1 *and* m^2 *are continuously differentiable.*

Assumption S.2 (independence): $(\varepsilon_1, \varepsilon_2)$ *is distributed independently of* X.

Assumption S.3 (support): *Conditional on any value* x of X, *the densities of* $(\varepsilon_1, \varepsilon_2)$ *and of* (Y_1, Y_2) *are continuous and have convex support.*

Assumption S.4 (monotonicity): *For all values* y_2 *of* Y_2, *the function* m^1 *is strictly increasing in* ε_1; *and for all values* (y_1, x) *of* (Y_1, X), *the function* m^2 *is strictly increasing in* ε_2.

Assumption S.5 (crossing): *For all values* $(y_1, y_2, x, \varepsilon_1, \varepsilon_2)$ *of* $(Y_1, Y_2, X, \varepsilon_1, \varepsilon_2)$,

$\left(\partial m^1(y_2, \varepsilon_1)/\partial y_2\right)\left(\partial m^2(y_1, x, \varepsilon_2)/\partial y_1\right) < 1.$

The technical assumptions S.1-S.3 could be partially relaxed at the cost of making the presentation more complex. Assumption S.4 guarantees that the function m^1 can be inverted in ε_1 and that the function m^2 can be inverted in ε_2. Hence, this assumption allows us to express the direct system of structural equations (S), defined by (m^1, m^2), in terms of a structural inverse system (I) of functions (r^1, r^2), which map any vector of observable variables (y_1, y_2, x) into the vector of unobservable variables $(\varepsilon_1, \varepsilon_2)$,

$$\text{Model (I)} \qquad \begin{aligned} \varepsilon_1 &= r^1(y_1, y_2) \\ \varepsilon_2 &= r^2(y_1, y_2, x). \end{aligned}$$

Assumption S.5 is a weakening of the common situation where the value of the endogenous variables is determined by the intersection of a downwards and an upwards slopping function. Together with Assumption S.4, this assumption guarantees the existence of a unique reduced form system (R) of equations, defined by functions (h^1, h^2), which map the vector of exogenous variables $(\varepsilon_1, \varepsilon_2, x)$ into the vector of endogenous variables (y_1, y_2),

$$\text{Model (R)} \qquad \begin{aligned} y_1 &= h^1(x, \varepsilon_1, \varepsilon_2) \\ y_2 &= h^2(x, \varepsilon_1, \varepsilon_2). \end{aligned}$$

These assumptions also guarantee that the reduced form function h^1 is monotone increasing in ε_1 and the reduced form function h^2 is monotone increasing in ε_2. These results are established in the following Lemma.

Lemma 1: *Suppose that Model (S) satisfies Assumptions S.1–S.5. Then, there exist unique functions h^1 and h^2 representing Model (S). Moreover, for all $x, \varepsilon_1, \varepsilon_2$, h^1 and h^2 are continuously differentiable, $\partial h^1 (x, \varepsilon_1, \varepsilon_2) / \partial \varepsilon_1 > 0$ and $\partial h^2 (x, \varepsilon_1, \varepsilon_2) / \partial \varepsilon_2 > 0$.*

Proof of Lemma 1: Assumption S.4 guarantees the existence of the structural inverse system (I) of differentiable functions (r^1, r^2) satisfying

$$\begin{aligned} y_1 &= m^1 \left(y_2, r^1 (y_1, y_2) \right) \\ y_2 &= m^2 \left(y_1, x, r^2 (y_1, y_2, x) \right) \end{aligned}$$

By Assumption S.1, we can differentiate these equations with respect to y_1 and y_2, to get

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\partial m^1}{\partial \varepsilon_1} \frac{\partial r^1}{\partial y_1} & \frac{\partial m^1}{\partial y_2} + \frac{\partial m^1}{\partial \varepsilon_1} \frac{\partial r^1}{\partial y_2} \\ \frac{\partial m^2}{\partial y_1} + \frac{\partial m^2}{\partial \varepsilon_2} \frac{\partial r^2}{\partial y_1} & \frac{\partial m^2}{\partial \varepsilon_2} \frac{\partial r^2}{\partial y_2} \end{pmatrix}$$

Hence, $\partial r^1 / \partial y_1 = \left(\partial m^1 / \partial \varepsilon_1 \right)^{-1}$, $\partial r^2 / \partial y_2 = \left(\partial m^2 / \partial \varepsilon_2 \right)^{-1}$, $\partial r^1 / \partial y_2 =$ $= \; - \left(\partial m^1 / \partial \varepsilon_1 \right)^{-1} \left(\partial m^1 / \partial y_2 \right)$, and $\partial r^2 / \partial y_1 \; = \; - \left(\partial m^2 / \partial \varepsilon_2 \right)^{-1} \left(\partial m^2 / \partial y_1 \right)$. These expressions together with Assumptions S.4 and S.5 imply that $\partial r^1 / \partial y_1 > 0$, $\partial r^2 / \partial y_2 > 0$, and $\left(\partial r^1 / \partial y_2 \right) \left(\partial r^2 / \partial y_1 \right) < 0$. Hence the determinants of all principal submatrices of the Jacobian matrix

$$\begin{pmatrix} \frac{\partial r^1 (y_1, y_2)}{\partial y_1} & \frac{\partial r^1 (y_1, y_2)}{\partial y_2} \\ \frac{\partial r^2 (y_1, y_2, x)}{\partial y_1} & \frac{\partial r^2 (y_1, y_2, x)}{\partial y_2} \end{pmatrix}$$

of (r^1, r^2) with respect to (y_1, y_2) are positive. It follows by Gale and Nikaido

(1965) that there exist unique functions (h^1, h^2) such that for all $(x, \varepsilon_1, \varepsilon_2)$

$$\begin{aligned} \varepsilon_1 &= r^1\left(h^1(x,\varepsilon_1,\varepsilon_2), h^2(x,\varepsilon_1,\varepsilon_2)\right) \\ \varepsilon_2 &= r^2\left(h^1(x,\varepsilon_1,\varepsilon_2), h^2(x,\varepsilon_1,\varepsilon_2), x\right) \end{aligned}$$

We have then established the existence of the reduced form system (R). The Implicit Function Theorem implies by Assumption S.1 that h^1 and h^2 are continuously differentiable. Moreover, the Jacobian matrix of (h^1, h^2) with respect to $(\varepsilon_1, \varepsilon_2)$ is the inverse of the Jacobian matrix of (r^1, r^2) with respect to (y_1, y_2). Assumptions S.4 and S.5 then imply that for all $x, \varepsilon_1, \varepsilon_2$, $\partial h^2(x,\varepsilon_1,\varepsilon_2)/\partial\varepsilon_2 > 0$ and $\partial h^2(x,\varepsilon_1,\varepsilon_2)/\partial\varepsilon_2 > 0$. This completes the proof of Lemma 1.//

We next define a new property, which we call *control function separability.*

Definition: *A structural inverse system of equations* $(r^1(y_1,y_2), r^2(y_1,y_2,x))$ *satisfies control function separability if there exist functions* $v : R^2 \to R$ *and* $q : R^2 \to R$ *such that*

(a) for all (y_1, y_2, x),

$$r^2(y_1,y_2,x) = v\left(q(y_2,x), r^1(y_1,y_2)\right)$$

(b) for any value of its second argument, v *is strictly increasing in its first argument, and*

(c) for any value of its first argument, q *is strictly increasing in its second argument.*

2.2 The triangular model and observational equivalence

We will consider triangular models of the form

$$\text{Model (T)} \qquad \begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= s(x, \eta) \end{aligned}$$

satisfying the following assumptions.

Assumption T.1 (differentiability): *For all values of $y_1, y_2, x, \varepsilon_1, \eta$ of $Y_1, Y_2, X, \varepsilon_1, \eta$ the functions m^1 and s are continuously differentiable.*

Assumption T.2 (independence): *(ε_1, η) is distributed independently of X.*

Assumption T.3 (support): *Conditional on any value x* of *X, the densities of (ε_1, η) and of (Y_1, Y_2) are continuous and have convex support.*

Assumption T.4 (monotonicity): *For all values of y_2, the function m^1 is strictly increasing in ε_1; and for all values of x, the function s is strictly increasing in η.*

Using the standard definition of observational equivalence, we will say that Model (S) is observationally equivalent to Model (T) if the distributions of the observable variables generated by each of these models is the same:

Definition: *Model (S) is observationally equivalent to model (T) iff for all* y_1, y_2, x *such that* $f_X(x) > 0$

$$f_{Y_1,Y_2|X=x}(y_1, y_2; S) = f_{Y_1,Y_2|X=x}(y_1, y_2; T).$$

In the next section, we establish that control function separability completely characterizes observational equivalence between Model (S) and Model (T).

3 Characterization of Observational Equivalence and Control Function Separability

Our characterization theorem is the following:

Theorem 1: *Suppose that Model (S) satisfies Assumptions S.1-S.5 and Model (T) satisfies Assumptions T.1-T.4. Then, Model (S) is observationally equivalent to Model (T) if and only if the inverse system of equations* $(r^1(y_1, y_2), r^2(y_1, y_2, x))$ *derived from (S) satisfies control function separability.*

Proof of Theorem 1: Suppose that Model (S) is observationally equivalent to Model (T). Then, for all y_1, y_2, x such that $f_X(x) > 0$

$$f_{Y_1,Y_2|X=x}(y_1, y_2; S) = f_{Y_1,Y_2|X=x}(y_1, y_2; T).$$

Consider the transformation

$$\begin{aligned} \varepsilon_1 &= r^1(y_1, y_2) \\ y_2 &= y_2 \\ x &= x \end{aligned}$$

The inverse of this transformation is

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= y_2 \\ x &= x \end{aligned}$$

Hence, the conditional density of (ε_1, y_2) given $X = x$, under Model T and under Model S are, respectively

$$f_{\varepsilon_1, Y_2|X=x}(\varepsilon_1, y_2; T) = f_{Y_1,Y_2|X=x}\left(m^1(y_2, \varepsilon_1), y_2; T\right) \left|\frac{\partial m^1(y_2, \varepsilon_1)}{\partial \varepsilon_1}\right|$$

and

$$f_{\varepsilon_1, Y_2|X=x}(\varepsilon_1, y_2; S) = f_{Y_1,Y_2|X=x}\left(m^1(y_2, \varepsilon_1), y_2; S\right) \left|\frac{\partial m^1(y_2, \varepsilon_1)}{\partial \varepsilon_1}\right|.$$

In particular, for all y_2, all x such that $f_X(x) > 0$, and for $\varepsilon_1 = r^1(y_1, y_2)$

$$(T1.1) \quad f_{Y_2|\varepsilon_1 = r^1(y_1,y_2), X=x}(y_2; T) = f_{Y_2|\varepsilon_1 = r^1(y_1,y_2), X=x}(y_2; S).$$

That is, the distribution of Y_2 conditional on $\varepsilon_1 = r^1(y_1, y_2)$ and $X = x$, generated by either Model (S) or Model (T) must be the same. By Model (T), the conditional distribution of Y_2 conditional on $(\varepsilon_1, X) = (r^1(y_1, y_2), x)$ can be expressed as

$$\begin{aligned} &\Pr\left(Y_2 \leq y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\ = &\Pr\left(s(x, \eta) \leq y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\ = &\Pr\left(\eta \leq \widetilde{s}(x, y_2) | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\ = &F_{\eta|\varepsilon_1 = r^1(y_1,y_2)}(\widetilde{s}(x, y_2)). \end{aligned}$$

where $\widetilde{s}$ denotes the inverse of s with respect to η. The existence of $\widetilde{s}$ and its strict monotonicity with respect to y_2 is guaranteed by Assumption T.4. The last equality follows because Assumption T.2 implies that conditional on ε_1, η is independent of X. On the other side, by Model (S), we have that

$$\begin{aligned}
& \Pr\left(Y_2 \leq y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\
= & \Pr\left(h^2(x, \varepsilon_1, \varepsilon_2) \leq y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\
= & \Pr\left(\varepsilon_2 \leq \widetilde{h}^2(x, \varepsilon_1, y_2) | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\
= & \Pr\left(\varepsilon_2 \leq r^2\left(m^1(y_2, \varepsilon_1), y_2, x\right) | \varepsilon_1 = r^1(y_1, y_2), X = x\right) \\
= & F_{\varepsilon_2 | \varepsilon_1 = r^1(y_1, y_2)}\left(r^2\left(m^1(y_2, \varepsilon_1), y_2, x\right)\right).
\end{aligned}$$

where $\widetilde{h}^2$ denotes the inverse of h^2 with respect to ε_2. The existence of $\widetilde{h}^2$ and its strict monotonicity with respect to y_2 follows by Lemma 1. The third equality follows because when $\varepsilon_1 = r^1(y_1, y_2)$, the value of ε_2 such that

$$y_2 = h^2(x, \varepsilon_1, \varepsilon_2)$$

is

$$\varepsilon_2 = r^2(y_1, y_2, x)$$

The last equality follows because Assumption S.2 implies that conditional on ε_1, ε_2 is independent of X.

Equating the expressions that we got for $\Pr(Y_2 \leq y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x)$ from Model (T) and from Model (S), we can conclude that for all y_2, x, ε_1

$$(T1.2) \qquad F_{\varepsilon_2 | \varepsilon_1 = r^1(y_1, y_2)}\left(r^2\left(m^1(y_2, \varepsilon_1), y_2, x\right)\right) = F_{\eta | \varepsilon_1 = r^1(y_1, y_2)}\left(\widetilde{s}(x, y_2)\right)$$

Substituting $m^1(y_2, \varepsilon_1)$ by $y_1,$ we get that for all y_1, y_2, x

$$F_{\varepsilon_2 | \varepsilon_1 = r^1(y_1, y_2)}\left(r^2(y_1, y_2, x)\right) = F_{\eta | \varepsilon_1 = r^1(y_1, y_2)}\left(\widetilde{s}(x, y_2)\right)$$

Note that the distribution of ε_2 conditional on ε_1 can be expressed as an unknown function $G(\varepsilon_2, \varepsilon_1)$, of two arguments. Analogously, the distribution of η conditional on ε_1 can be expressed as an unknown function $H(\eta, \varepsilon_1)$. Denote the (possibly infinite) support of ε_2 conditional on $\varepsilon_1 = r^1(y_1, y_2)$ by $[\varepsilon_L^2, \varepsilon_U^2]$, and the (possibly infinite) support of η conditional on $\varepsilon_1 = r^1(y_1, y_2)$ by $[\eta_L, \eta_U]$. Our assumptions S.2 and S.3 imply that the distribution $F_{\varepsilon_2|\varepsilon_1=r^1(y_1,y_2)}(\cdot)$ is strictly increasing on $[\varepsilon_L^2, \varepsilon_U^2]$ and maps $[\varepsilon_L^2, \varepsilon_U^2]$ onto $[0,1]$. Our Assumptions T.2 and T.3 imply that the distribution $F_{\eta|\varepsilon_1=r^1(y_1,y_2)}(\cdot)$ is strictly increasing on $[\eta_L, \eta_U]$ and maps $[\eta_L, \eta_U]$ onto $[0,1]$. Hence, $(T1.2)$ and our assumptions imply that there exists a function $\widetilde{s}$, strictly increasing in its second argument, and functions $G(\varepsilon_2, \varepsilon_1)$ and $H(\eta, \varepsilon_1)$, such that for all y_1, y_2, x with $f_X(x) > 0$ and $f_{Y_1,Y_2|X=x}(y_1, y_2) > 0$

$$G\left(r^2(y_1, y_2, x), r^1(y_1, y_2)\right) = H\left(\widetilde{s}(x, y_2), r^1(y_1, y_2)\right)$$

and G and H are both strictly increasing in their first arguments at, respectively, $\varepsilon^2 = r^2(y_1, y_2, x)$ and $\eta = \widetilde{s}(x, y_2)$. Let $\widetilde{G}$ denote the inverse of G, with respect to its first argument. Then, $\widetilde{G}(\cdot, r^1(y_1, y_2)) : [0,1] \to [r_L^2 . r_U^2]$ is strictly increasing on $(0,1)$ and

$$r^2(y_1, y_2, x) = \widetilde{G}\left(H\left(\widetilde{s}(x, y_2), r^1(y_1, y_2)\right), r^1(y_1, y_2)\right)$$

This implies that r^2 is weakly separable into $r^1(y_1, y_2)$ and a function of (x, y_2), strictly increasing in y_2. Moreover, since H and $\widetilde{G}$ are both strictly increasing with respect to their first argument on their respective relevant domains, r^2 must be strictly increasing in the value of $\widetilde{s}$. Extending the function $\widetilde{s}$ to be strictly increasing at all $y_2 \in R$ and extending the function $\widetilde{G} \circ H$ to be strictly increasing on all values $\widetilde{s} \in R$, we can conclude that $(T1.1)$, and hence also the observational equivalence between Model (T) and Model (S), implies that $(r^1(y_1, y_2), r^2(y_1, y_2, x))$ satisfies control function separability.

To show that control function separability implies the observational equivalence between Model (S) and Model (T), suppose that Model (S), satisfying Assumptions S.1-S.5, is such that there exist continuously differentiable functions $v : R^2 \to R$ and $q : R^2 \to R$ such that for all (y_1, y_2, x),

$$r^2(y_1, y_2, x) = v\left(q(y_2, x), r^1(y_1, y_2)\right),$$

where for any value of $r^1(y_1, y_2)$, q is strictly increasing in y_2 and v is strictly increasing in its first argument. Let $\varepsilon_1 = r^1(y_1, y_2)$ and $\overline{\eta} = q(x, y_2)$. Then

$$\varepsilon_2 = r^2(y_1, y_2, x) = v(\overline{\eta}, \varepsilon_1)$$

where v is strictly increasing in $\overline{\eta}$. Letting $\widetilde{v}$ denote the inverse of v with respect to $\overline{\eta}$, it follows that,

$$q(y_2, x) = \overline{\eta} = \widetilde{v}(\varepsilon_2, \varepsilon_1).$$

Since $\overline{\eta}$ is a function of $(\varepsilon_1, \varepsilon_2)$, Assumption S.2 implies Assumption T.2. Since $\widetilde{v}$ is strictly increasing in ε_2, Assumption S.3 implies that $(\varepsilon_1, \overline{\eta})$ has a continuous density on a convex support. Let $\widetilde{q}$ denote the inverse of q with respect to y_2. The function $\widetilde{q}$ exists because q is strictly increasing in y_2. Then,

$$y_2 = \widetilde{q}(\overline{\eta}, x) = \widetilde{q}(\widetilde{v}(\varepsilon_2, \varepsilon_1), x).$$

Since

$$y_2 = h^2(x, \varepsilon_1, \varepsilon_2)$$

it follows that

$$y_2 = h^2(x, \varepsilon_1, \varepsilon_2) = \widetilde{q}(\widetilde{v}(\varepsilon_2, \varepsilon_1), x)$$

where $\widetilde{q}$ is strictly increasing with respect to its first argument. Hence,

$$y_2 = h^2(x, \varepsilon_1, \varepsilon_2) = \widetilde{q}(\overline{\eta}, x)$$

where $\widetilde{q}$ is strictly increasing in $\overline{\eta}$. This implies that control function separability implies that the system composed of the structural form function for y_1 and the reduced form function for y_2 is of the form

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= h^2(x, \varepsilon_1, \varepsilon_2) = \widetilde{q}(\widetilde{v}(\varepsilon_2, \varepsilon_1), x) = \widetilde{q}(\overline{\eta}, x) \end{aligned}$$

where $\widetilde{q}$ is strictly increasing in $\overline{\eta}$ and $(\varepsilon_1, \overline{\eta})$ is independent of X. To show that the model generated by (m^1, h^2) is observationally equivalent to the model generated by $(m^1, \widetilde{q})$, we note that the model generated by (m^1, h^2) implies that for all x such that $f_X(x) > 0$,

$$\begin{aligned} & f_{Y_1,Y_2|X=x}(y_1, y_2; S) \\ = \; & f_{\varepsilon_1,\varepsilon_2}\left(r^1(y_1, y_2), r^2(y_1, y_2, x)\right) \left|r^1_{y_1} r^2_{y_2} - r^1_{y_2} r^2_{y_1}\right| \end{aligned}$$

where $r^1_{y_1} = r^1_{y_1}(y_1, y_2)$, $r^2_{y_2} = r^2_{y_2}(y_1, y_2, x)$, $r^1_{y_2} = r^1_{y_2}(y_1, y_2)$, and $r^2_{y_1} = r^2_{y_1}(y_1, y_2, x)$. On the other side, for the model generated by $(m^1, \widetilde{q})$, we have that,

$$\begin{aligned} & f_{Y_1,Y_2|X=x}(y_1, y_2; T) \\ = \; & f_{\varepsilon_1,\overline{\eta}}\left(r^1(y_1, y_2), \widetilde{v}\left(r^2(y_1, y_2, x), r^1(y_1, y_2)\right)\right) \left|r^1_{y_1}\left(\widetilde{v}_1 r^2_{y_2} + \widetilde{v}_2 r^1_{y_2}\right) - r^1_{y_2}\left(\widetilde{v}_1 r^2_{y_1} + \widetilde{v}_2 r^1_{y_1}\right)\right| \end{aligned}$$

where $\widetilde{v}_1$ denotes the derivative of $\widetilde{v}$ with respect to its first coordinate and $\widetilde{v}_2$ denotes the derivative of $\widetilde{v}$ with respect to its second coordinate. Since

$$\left|r^1_{y_1}\left(\widetilde{v}_1 r^2_{y_2} + \widetilde{v}_2 r^1_{y_2}\right) - r^1_{y_2}\left(\widetilde{v}_1 r^2_{y_1} + \widetilde{v}_2 r^1_{y_1}\right)\right| = \widetilde{v}_1 \left|r^1_{y_1} r^2_{y_2} - r^1_{y_2} r^2_{y_1}\right|$$

and

$$f_{\varepsilon_2|\varepsilon_1=r^1(y_1,y_2)}\left(r^2\left(y_1,y_2,x\right)\right)=f_{\overline{\eta}|\varepsilon_1=r^1(y_1,y_2)}\left(\widetilde{v}\left(r^2\left(y_1,y_2,x\right),r^1\left(y_1,y_2\right)\right)\right)\ \widetilde{v}_1$$

it follows that for all x such that $f_X(x)>0$,

$$f_{Y_1,Y_2|X=x}\left(y_1,y_2;S\right)=f_{Y_1,Y_2|X=x}\left(y_1,y_2;T\right)$$

Hence, control function separability implies that Model (S) is observationally equivalent to Model (T). This completes the proof of Theorem 1.//

Theorem 1 provides a characterization of two-equation systems with simultaneity where one of the functions can be estimated using the other to derive a control function. One of the main conclusions of the theorem is that to verify whether one of the equations can be used to derive a control function, it must be that the inverse function of that equation, which maps the observable endogenous and observable exogenous variables into the value of the unobservable, must be separable into the inverse function of the first equation and a function not involving the dependent variable of the first equation. That is, the function

$$y_2=m^2\left(y_1,x,\varepsilon_2\right)$$

can be used to derive a control function to identify the function m^1, where

$$y_1=m^1\left(y_2,\varepsilon_1\right)$$

if and only if the inverse function of m^2 with respect to ε_2 is separable into r^1 and a function of y_2 and x.

In the Appendix we provide equivalent characterizations of these conditions in terms of the derivatives of the structural functions and of the reduced form system (R).

4 An example

We next provide an example of an optimization problem, for which the first order conditions satisfy control function separability. Our results then imply that one can estimate the structural equation using a control function approach. The objective function in our example is specified as

$$\begin{aligned} & V\left(y_1, y_2, x_1, x_2, x_3\right) \\ = \; & \left(\varepsilon_1+\varepsilon_2\right)\; u\left(y_2\right)+\varepsilon_1\; \log \left(y_1-u\left(y_2\right)\right)-y_1 x_1-y_2 x_2+x_3 \end{aligned}$$

This can be the objective function of a consumer choosing demand for three products, (y_1, y_2, y_3) subject to a linear budget constraint, $x_1 y_1+x_2 y_2+y_3 \leq x_3$, with x_1 and x_2 denoting the prices of, respectively, y_1 and y_2 and x_3 denoting income.

The first order conditions with respect to y_1 and y_2 are

$$(5.1)\ \tfrac{\partial}{\partial y_1}: \qquad \tfrac{\varepsilon_1}{\left(y_1-u\left(y_2\right)\right)}-x_1=0$$

$$(5.2)\ \tfrac{\partial}{\partial y_2}: \qquad \left(\varepsilon_1+\varepsilon_2\right) u^{\prime}\left(y_2\right)-u^{\prime}\left(y_2\right) \tfrac{\varepsilon_1}{\left(y_1-u\left(y_2\right)\right)}-x_2=0$$

The Hessian of the objective function is

$$\begin{bmatrix} \frac{-\ \varepsilon_1}{(y_1-u(y_2))^2} & \frac{\varepsilon_1\ u'(y_2)}{(y_1-u(y_2))^2} \\ \frac{\varepsilon_1\ u'(y_2)}{(y_1-u(y_2))^2} & \left(\varepsilon_1+\varepsilon_2-\frac{\varepsilon_1}{(y_1-u(y_2))}\right)u''(y_2)-\left(u'(y_2)\right)^2\frac{\varepsilon_1}{(y_1-u(y_2))^2} \end{bmatrix}.$$

This Hessian is negative definite when $\varepsilon_1 > 0$, $u'(y_2) > 0$, $u''(y_2) < 0$ and

$$\left(\varepsilon_1+\varepsilon_2-\frac{\varepsilon_1}{(y_1-u(y_2))}\right) > 0$$

Since at the values of (y_1, y_2) that satisfy the First Order conditions, $\varepsilon_1/(y_1-u(y_2)) = x_1$ and $(\varepsilon_1+\varepsilon_2-(\varepsilon_1/(y_1-u(y_2))))\,u'(y_2) = x_2$,as long as $x_1 > 0$ and $x_2 > 0$, the objective function is strictly concave at (y_1, y_2) that solve the optimization problem.

To obtain the system of structural equations, note that from (5.1), we get

$$(5.3) \qquad \varepsilon_1 = [y_1-u(y_2)]\ x_1$$

And using (5.3) in (5.2), we get

$$(5.4) \qquad [(\varepsilon_1+\varepsilon_2)-x_1]\,u'(y_2) = x_2$$

Hence,

$$\begin{aligned} \varepsilon_2 &= \frac{x_2}{u'(y_2)} - y_1x_1 + u(y_2)\,x_1 + x_1 \\ &= \left(\frac{x_2}{u'(y_2)} + x_1\right) - (y_1-u(y_2))\,x_1 \end{aligned}$$

We can then easily see that the resulting *system of structural equations,* which is

$$\varepsilon_1 = [y_1 - u(y_2)]\ x_1$$

$$\varepsilon_2 = \left(\frac{x_2}{u'(y_2)} + x_1 \right) - (y_1 - u(y_2))\, x_1$$

satisfy control function separability. The *triangular system of equations*, which can then be estimated using a control function for nonseparable models, is

$$y_1 = u(y_2) + \frac{\varepsilon_1}{x_1}$$

$$y_2 = (u')^{-1} \left(\frac{x_2}{\varepsilon_1 + \varepsilon_2 - x_1} \right)$$

The unobservable $\eta = \varepsilon_1 + \varepsilon_2$ is the control function for y_2 in the equation for y_1. Conditional on $\eta = \varepsilon_1 + \varepsilon_2$, y_2 is a function of only (x_1, x_2), which is independent of ε_1. Hence, conditional on $\eta = \varepsilon_1 + \varepsilon_2$, y_2 is independent of ε_1, exactly the conditions one needs to use η as the control function in the estimation of the equation for y_1.

5 Simultaneity in Latent Variables

Our results can be applied to a wide range of Limited Dependent Variable models with simultaneity in the latent variables, when additional exogenous variables are observed and some separability conditions are satisfied. In particular, suppose that we were interested in estimating m^1 in the model

$$y_1^* = m^1(y_2^*, w_1, w_2, \varepsilon_1)$$
$$y_2^* = m^2(y_1^*, w_1, w_2, x, \varepsilon_2)$$

where instead of observing (y_1^*, y_2^*), we observed a transformation, (y_1, y_2), of (y_1^*, y_2^*) defined by a known vector function (T_1, T_2),

$$\begin{aligned} y_1 &= T_1(y_1^*, y_2^*) \\ y_2 &= T_2(y_1^*, y_2^*) \end{aligned}$$

Assume that (w_1, w_2, x) is independent of $(\varepsilon_1, \varepsilon_2)$ and that for known functions b_1 and b_2 and unknown functions $\overline{m}^1$ and $\overline{m}^2$ the system of simultaneous equations can be written as

$$\begin{aligned} b_1(y_1^*, w_1) &= \overline{m}^1(b_2(y_2^*, w_2), \varepsilon_1) \\ b_2(y_2^*, w_2) &= \overline{m}^2(b_1(y_1^*, w_1), x, \varepsilon_2) \end{aligned}$$

Then, under support conditions on (w_1, w_2, x) and on the range of (T_1, T_2), and under invertibility conditions on (b_1, b_2), one can express this system as

$$\begin{aligned} \overline{b}_1 &= \overline{m}^1(\overline{b}_2, \varepsilon_1) \\ \overline{b}_2 &= \overline{m}^2(\overline{b}_1, x, \varepsilon_2) \end{aligned}$$

where the distribution of $(\overline{b}_1, \overline{b}_2, x)$ is known. (See Matzkin (2010b) for formal assumptions and arguments and more general models.) The identification and estimation of $\widetilde{m}^1$ can then proceed using a control function approach, as developed in the previous sections, when this system satisfies control function separability.

To provide a simple specific example of the arguments that are involved in the above statements, we consider a special case of a binary threshold crossing model analyzed in Briesch, Chintagunta and Matzkin (1997, 2009),

$$y_1^* = m^1(y_2, \varepsilon_1) + w_1$$

$$\begin{aligned} y_1 &= 0 \quad \text{if} \quad y_1^* \leq 0 \\ &= 1 \quad \text{otherwise} \end{aligned}$$

Suppose that instead of assuming as they did, that (y_2, w_1) is independent of ε_1, we assume that

$$y_2 = m^2 \left(y_1^* - w_1, x, \varepsilon_2\right)$$

and that (x, w_1) is independent of $(\varepsilon_1, \varepsilon_2)$. An example of such a model is where y_2 is discretionary expenditure by an individual in a store for which expenditures are observable, w_1 is an exogenous observable expenditure, and $y_1^* - w_1$ is unobserved discretionary expenditure over the fixed amount w_1. Assuming that m^1 is invertible in ε_1 and m^2 is invertible in ε_2, we can rewrite the two equation system as

$$\begin{aligned} \varepsilon_1 &= r^1 \left(y_1^* - w_1, y_2\right) \\ \varepsilon_2 &= r^2 \left(y_1^* - w_1, y_2, x\right) \end{aligned}$$

If this system can be expressed as

$$\begin{aligned} \varepsilon_1 &= r^1 \left(y_1^* - w_1, y_2\right) \\ \varepsilon_2 &= v \left(r^1 \left(y_1^* - w_1, y_2\right), s\left(y_2, x\right)\right) \end{aligned}$$

for some unknown functions r^1, v and s, satisfying our regularity conditions, then one can identify and estimate m^1 using a control function approach. To shed more light on this result, let $\overline{b}_1 = y_1^* - w_1$. Then, the model becomes

$$\begin{aligned} \overline{b}_1 &= m^1 \left(y_2, \varepsilon_1\right) \\ y_2 &= m^2 \left(\overline{b}_1, x, \varepsilon_2\right) \end{aligned}$$

with a system of reduced form functions

$$\begin{aligned}\bar{b}_1 &= h^1(x, \varepsilon_1, \varepsilon_2) \\ y_2 &= h^2(x, \varepsilon_1, \varepsilon_2)\end{aligned}$$

Following Matzkin (2010b), we extend arguments for identification of semiparametric binary threshold crossing models using conditional independence (Lewbel (2000)), and arguments for identification of nonparametric and nonadditive binary threshold crossing models using independence (Matzkin (1992), Briesch, Chintagunta, and Matzkin (1997, 2009)) to models with simultaneity. For this, we assume that (X, W) has an everywhere positive density. Our independence assumption implies that W is independent of $(\varepsilon_1, \varepsilon_2)$ conditional on X. Then, since conditional on X, $\left(\bar{b}_1, y_2\right)$ is only a function of $(\varepsilon_1, \varepsilon_2)$, we have that for all w_1, t_1

$$\begin{aligned}\Pr\left(\left(\overline{B}_1, Y_2\right) \leq (t_1, y_2) | X = x\right) &= \Pr\left(\left(\overline{B}_1, Y_2\right) \leq (t_1, y_2) | W_1 = w_1, X = x\right) \\ &= \Pr\left((Y_1^* - W_1, Y_2) \leq (t_1, y_2) | W_1 = w_1, X = x\right) \\ &= \Pr\left((Y_1^*, Y_2) \leq (t_1 + w_1, y_2) | W_1 = w_1, X = x\right)\end{aligned}$$

Letting $w_1 = -t_1$, we get that

$$\Pr\left(\left(\overline{B}_1, Y_2\right) \leq (t_1, y_2) | X = x\right) = \Pr\left((Y_1, Y_2) \leq (0, y_2) | W_1 = -t_1, X = x\right)$$

Hence, the distribution of $\left(\bar{b}_1, y_2\right)$ conditional on X is identified. The analysis of the system

$$\begin{aligned}\bar{b}_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= m^2\left(\bar{b}_1, x, \varepsilon_2\right)\end{aligned}$$

when this identified distribution is given is analogous to the analysis of the system

$$y_1 = m^1(y_2, \varepsilon_1)$$
$$y_2 = m^2(y_1, x, \varepsilon_2)$$

with the distribution of (y_1, y_2) given X, considered in our previous sections. In particular, if the system satisfies control function separability, we can first estimate the model $y_2 = \widetilde{s}(x, \eta)$ where $\widetilde{s}$ is an unknown function increasing in η, and then use the estimated η as a control in the estimation of m^1.

6 Conclusions

In this note we have provided a conclusive answer to the question of when it is possible to use a control function approach to identify and estimate a function in a simultaneous equations model. We define a new property of functions, called control function separability, which characterizes systems of simultaneous equations where a function of interest can be estimated using a control function derived from the second equation. We show that this condition is equivalent to requiring that the reduced form function for the endogenous regressor in the function of interest is separable into a function of all the unobservable variables. We also provide conditions in terms of the derivatives of the two functions in the system.

An example a system of structural equations, which is generated by the first order conditions of an optimization problem, and which satisfies control function separability, was presented. We have also shown how our results can be used to identify and estimate Limited Dependent Variable models with simultaneity in the latent or observable continuous variables.

7 Appendix A

7.1 Characterization in terms of Derivatives

Taking advantage of the assumed differentiability, we can characterize systems where one of the functions can be estimated using a control function approach using a condition in terms of the derivatives of the functions of Models (T) and (S). The following result provides such a condition. Let $r_x^2 = \partial r^2 (y_1, y_2, x) / \partial x$, $r_{y_1}^2 = \partial r^2 (y_1, y_2, x) / \partial y_1$, and $r_{y_2}^2 = \partial r^2 (y_1, y_2, x) / \partial y_2$ denote the derivatives of r^2, $\widetilde{s}_x = \partial \widetilde{s} (y_2, x) / \partial x$ and $\widetilde{s}_{y_2} = \partial \widetilde{s} (y_2, x) / \partial y_2$ denote the derivatives of $\widetilde{s}$, and let $m_{y_2}^1 = \partial m^1 (y_2, \varepsilon_1) / \partial y_2$ denote the derivative of the function of interest m^1 with respect to the endogenous variable y_2.

Theorem 2: *Suppose that Model (S) satisfies Assumptions S.1-S.5 and that Model (T) satisfies Assumptions T.1-T.4. Then, Model (S) is observationally equivalent to Model (T) if and only if for all* x, y_1, y_2,

$$\frac{r_x^2}{r_{y_1}^2 \; m_{y_2}^1 \; + \; r_{y_2}^2} = \frac{\widetilde{s}_x}{\widetilde{s}_{y_2}}$$

Proof of Theorem 2: As in the proof of Theorem 1, observational equivalence between Model (T) and Model (S) implies that for all y_2, x, and $\varepsilon_1 = r^1 (y_1, y_2)$

$$(T1.2) \qquad F_{\varepsilon_2|\varepsilon_1} \left(r^2 \left(m^1 (y_2, \varepsilon_1), y_2, x \right) \right) = F_{\eta|\varepsilon_1} \left(\widetilde{s} (y_2, x) \right)$$

Differentiating both sides of $(T1.2)$ with respect to y_2 and x, we get that

$$f_{\varepsilon_2|\varepsilon_1}\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)\right)\ \left(r^2_{y_1}\ m^1_{y_2}\ +\ r^2_{y_2}\right)\ =\ f_{\eta|\varepsilon_1}\left(\widetilde{s}\left(y_2,x\right)\right)\ \widetilde{s}_{y_2}$$
$$f_{\varepsilon_2|\varepsilon_1}\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)\right)\ r^2_x\ =\ f_{\eta|\varepsilon_1}\left(\widetilde{s}\left(y_2,x\right)\right)\ \widetilde{s}_x$$

where, as defined above, $r^2_{y_1}=\partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial y_1$, $r^2_{y_2}=\partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial y_2$, $r^2_x=\partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial x$, $m^1_{y_2}=\partial m^1\left(y_2,\varepsilon_1\right)/\partial y_2$, $\widetilde{s}_{y_2}=\partial\widetilde{s}\left(y_2,x\right)/\partial y_2$, and $\widetilde{s}_x=\partial\widetilde{s}\left(y_2,x\right)/\partial x$. Taking ratios, we get that

$$\frac{r^2_x}{r^2_{y_1}\ m^1_{y_2}\ +\ r^2_{y_2}}\ =\frac{\widetilde{s}_x}{\widetilde{s}_{y_2}}$$

Conversely, suppose that for all y_2,x,ε_1,

$$(T2.1)\qquad \frac{r^2_x}{r^2_{y_1}\ m^1_{y_2}\ +\ r^2_{y_2}}\ =\frac{\widetilde{s}_x}{\widetilde{s}_{y_2}}\ .$$

Define

$$b\left(y_2,x,\varepsilon_1\right)=r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right).$$

$(T2.1)$ implies that, for any fixed value of ε_1, the function $b\left(y_2,x,\varepsilon_1\right)$ is a transformation of $\widetilde{s}\left(y_2,x\right)$. Let $t\left(\cdot,\cdot,\varepsilon_1\right):R\rightarrow R$ denote such a transformation. Then, for all y_2,x,

$$b\left(y_2,x,\varepsilon_1\right)=r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)=t\left(\widetilde{s}\left(y_2,x\right),\varepsilon_1\right).$$

Substituting $m^1\left(y_2,\varepsilon_1\right)$ with y_1 and ε_1 with $r^1\left(y_1,y_2\right)$, it follows that

$$r^2\left(y_1,y_2,x\right)=t\left(\widetilde{s}\left(y_2,x\right),r^1\left(y_1,y_2\right)\right)$$

Hence, $(T2.1)$ implies control function separability. This implies, by Theorem 1, that Model (T) and Model (S) are observationally equivalent, and it completes the proof of Theorem 2.//

Instead of characterizing observationally equivalence in terms of the derivatives of the functions m^1 and r^2, we can express observational equivalence in terms of the derivatives of the inverse structural form functions. Differentiating with respect to y_1 and y_2 the identity

$$y_1 = m^1\left(y_2, r^1\left(y_1, y_2\right)\right)$$

and solving for $m^1_{y_2}$, we get that

$$m^1_{y_2} = \frac{-\, r^1_{y_2}}{r^1_{y_1}}$$

Hence, the condition that for all y_1, y_2, x

$$\frac{r^2_x}{r^2_{y_1}\ m^1_{y_2}\ +\ r^2_{y_2}} = \frac{\widetilde{s}_x}{\widetilde{s}_{y_2}}$$

is equivalent to the condition that for all y_1, y_2, x

$$\frac{r^1_{y_1}\left(y_1, y_2\right)\ r^2_x\left(y_1, y_2, x\right)}{r^1_{y_1}\left(y_1, y_2\right)\ r^2_{y_2}\left(y_1, y_2, x\right) - r^1_{y_2}\left(y_1, y_2\right)\ r^2_{y_1}\left(y_1, y_2, x\right)} = \frac{\widetilde{s}_x(y_2, x)}{\widetilde{s}_{y_2}(y_2, x)}$$

or

$$\frac{r^1_{y_1}\left(y_1, y_2\right)\ r^2_x\left(y_1, y_2, x\right)}{\left|r_y\left(y_1, y_2, x\right)\right|} = \frac{\widetilde{s}_x(y_2, x)}{\widetilde{s}_{y_2}(y_2, x)}$$

where $\left|r_y\left(y_1, y_2, x\right)\right|$ is the Jacobian determinant of the vector function $r = \left(r^1, r^2\right)$ with respect to $\left(y_1, y_2\right)$.

Note that differentiating both sides of the above equation with respect to y_1, we get the following expression, only in terms of the derivatives of the inverse system of structural equations of Model (S)

$$\frac{\partial}{\partial y_1}\left(\frac{r^1_{y_1}\left(y_1, y_2\right)\ r^2_x\left(y_1, y_2, x\right)}{\left|r_y\left(y_1, y_2, x\right)\right|}\right) = 0$$

7.2 Characterization in terms of the Reduced Form Functions

An alternative characterization, which follows from the proof of Theorem 1, is in terms of the reduced form functions. Suppose we ask when the function

$$y_2 = m^2\left(y_1, x, \varepsilon_2\right)$$

can be used to derive a control function to identify the function m^1, where

$$y_1 = m^1\left(y_2, \varepsilon_1\right).$$

Our arguments show that the control function approach can be used *if and only if* the reduced form function, $h^2\left(x, \varepsilon_1, \varepsilon_2\right)$, for y_2 can be expressed as a function of x and a function of $\left(\varepsilon_1, \varepsilon_2\right)$. That is the control function approach can be used *if and only if*, for some functions $\widetilde{q}$ and $\widetilde{v}$

$$h^2\left(x, \varepsilon_1, \varepsilon_2\right) = \widetilde{q}\left(x, \widetilde{v}\left(\varepsilon_1, \varepsilon_2\right)\right)$$

Note that while the sufficiency of such a condition is obvious, the necessity, which follows from Theorem 1, had not been previously known.[4]

[4]Kasy (2010) also highlights the one-dimensional distribution conditional on the reduced form h^2 but does not relate this to restrictions on the structure of the simultaneous equation system (S) which is our primary objective.

8 References

AI, C. and X. CHEN (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1843.

BENKARD, C.L. and S. BERRY (2004) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), Cowles Foundation Discussion Paper #1482.

BENKARD, C.L. and S. BERRY (2006) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), *Econometrica, 74, 5.*

BLUNDELL, R. and J. L. POWELL (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress,* Volume II, edited by M. Dewatripont, L.P. Hansen, and S.J.Turnovsky, Cambridge University Press, Cambridge, U.K.

BRIESCH, R., P. CHINTAGUNTA, and R.L. MATZKIN (1997) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," mimeo, Northwestern University.

BRIESCH, R., P. CHINTAGUNTA, and R.L. MATZKIN (2010) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics,* Vol. 28, No. 2

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

BROWN, D.J. and R.L. MATZKIN (1998) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," Cowles Foundation Discussion Paper #1175.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2007): "Instrumental Variable Identification and Estimation of Nonseparable Models" *Journal of Econometrics*, 139(1), 4-14.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, Vol. 71, No. 5.

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2002), "Nonparametric Instrumental Regression," mimeo, IDEI, Toulouse.

FLORENS, J. P., J. J. HECKMAN, C. MEGHIR, and E. VYTLACIL (2008). "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Economet-*

rica, vol. 76(5), 1191-1206, 09

GALE, D. and H. NIKAIDO (1965), "The Jacobian Matrix and Global Univalence of Mappings," *Math. Annalen* 159, 81-93.

HAHN, J. and G. RIDDER (2010), "Conditional Moment Restrictions and Triangular Simultaneous Equations," forthcoming in the Review of Economics an Statistics.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," Center for Microdata Methods and Practice, WP 102/03.

HAUSMAN, J.A. (1983) "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics,* Vol. I, edited by Z. Griliches and M.D. Intriligator.

HECKMAN, J.J. and R. ROBB (1985) "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, edited by J.J. Heckman and B. Singer, Econometric Society Monograph 10, Cambridge University Press, Cambridge, U.K.

HECKMAN, J.J. (1978) "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica,* Vol. 46, No. 4, 931-959.

IMBENS, G.W. AND W.K. NEWEY (2009) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, Vol. 77, No. 5, September.

KASY, M. (2010) "Identification in Triangular Systems using Control Functions," mimeo, Berkeley, April.

LEWBEL, A. (2000) "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity and Instrumental Variables," *Journal of Econometrics,* 97, 145-177.

MATZKIN, R.L. (1992) "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold Crossing Models," Econometrica, Vol. 60, No. 2, p. 239-270.

MATZKIN, R.L. (2005) "Identification in Nonparametric Simultaneous Equations Models," mimeo, Northwestern University.

MATZKIN, R.L. (2007) "Nonparametric Identification," Chapter 73 in Handbook of Econometrics, Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

MATZKIN, R.L. (2008) "Identification in Nonparametric Simultaneous Equations," Econometrica, Vol. 76, No. 5, 945-978.

MATZKIN, R.L. (2010a) "Estimation of Nonparametric Models with Simultaneity," mimeo, UCLA.

MATZKIN, R.L. (2010b) “Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," mimeo, UCLA.

NEWEY, W. and J. POWELL (1989) “Instrumental Variables Estimation of Nonparametric Models,” mimeo.

NEWEY, W. and J. POWELL (2003) “Instrumental Variables Estimation of Nonparametric Models,” Econometrica,.

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica,* 56:2, 433-447.

TELSER, Lester G. (1964) "Iterative Estimation of a Set of Linear Regression Equations." Journal of the American Statistical Association, 59(307), 845-62.