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

Richard Blundell
UCL

and

Rosa L. Matzkin
UCLA

January 2010
Preliminary

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 under which this procedure can be used in nonlinear and nonparametric simultaneous equations has thus far been unknown. We define a new property of functions called *control function separability.* This is shown to provide a complete characterization of the 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 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. The usual disclaimer applies. Rosa L. Matzkin gratefully acknowledges as well support from NSF.

1. **Introduction**

A standard situation in applied econometrics is where one is interested in estimating a 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 and where is a nonseparable nonlinear function of y_2 and the unobservable ε_1. Additionally there is an observable variable x, which might be used as an instrument for the estimation of m^1. That is, 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) 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

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 in Chesher (2003) and Imbens and Newey (2003), who consider the structural model

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

with x independent (or locally independent) of $(\varepsilon_1, \varepsilon_2)$.

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

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

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 model

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 2SLS. Dhrymes (1970, equation 4.3.57) shows that the 2SLS coefficients can be obtained by a leastsquares 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.

- (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 model

- (R)

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

with x independent of (ε_1, η)?

In what follows we first a new property of functions, control function separability. We then show that this property completely characterizes systems of simultaneous equations where a function of interest can be estimated using a control function. We also provide conditions in terms of the derivatives of the two functions in the system.

2. **Assumptions and Definitions**

We will use the standard definition of observational equivalence:

Definition: *Model (S) is observationally equivalent to model (R) iff*

$$f_{y_1,y_2|x}(y_1, y_2; S) = f_{y_1,y_2|x}(y_1, y_2; R).$$

We will make the following assumptions:

Assumption 1 (monotonicity): *For all values of* y_2*, the function* m^1 *is strictly monotone in* ε_1*; and for all values of* (y_1, ε_2)*, the function* m^2 *is strictly monotone in* ε_2*.*

Assumption 2 (uniqueness): *For all values of* $(x, \varepsilon_1, \varepsilon_2)$*, there exist unique values of* (y_1, y_2) *satisfying (S).*

These assumptions guarantee the existence of two common alternative representations of (S). Given the structural direct system of equations (m^1, m^2), where

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

we define, as usual, the structural inverse system of equations (r^1, r^2) as the system that maps the vector of observable variables into the vector of unobservable variables

$$\begin{aligned} \varepsilon^1 &= r^1(y_1, y_2) \\ \varepsilon^2 &= r^2(y_1, y_2, x). \end{aligned}$$

We define the reduced form system of equations as the system that maps the exogenous variables into the endogenous variables

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

Assumption 1 guarantees the existence of (r^1, r^2) and Assumption 2 guarantees the existence of (h^1, h^2).

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* r^2 *is separable into* r^1 *and a function of* y_2 *and* x*. That is, for some functions* v *and* $\widetilde{s}$ *and all* (y_1, y_2, x),

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

3. Characterization of Observational Equivalence and Control Function Separability

Our characterization theorem is the following:

Theorem 1: *Model (S) is observationally equivalent to Model (R) 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 (R). Then, for all y_1, y_2, x,

$$f_{y_1,y_2|x}(y_1, y_2; S) = f_{y_1,y_2|x}(y_1, y_2; R).$$

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 joint density of (ε_1, y_2, x), under R and S is

$$f_{\varepsilon_1,y_2,x}(\varepsilon_1, y_2, x; R) = f_{y_1,y_2,x}\left(m^1(y_2, \varepsilon_1), y_2, x; R\right)\left|\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}\right|$$

$$f_{\varepsilon_1,y_2,x}(\varepsilon_1, y_2, x; S) = f_{y_1,y_2,x}\left(m^1(y_2, \varepsilon_1), y_2, x; S\right)\left|\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}\right|$$

It follows then that observational equivalence implies that

$$f_{\varepsilon_1,y_2,x}(\varepsilon_1, y_2, x; R) = f_{\varepsilon_1,y_2,x}(\varepsilon_1, y_2, x; S)$$

This implies that

$$f_{y_2|\varepsilon_1,x}(y_2; R) = f_{y_2|\varepsilon_1,x}(y_2; S)$$

Hence, observational equivalence between Model (S) and Model (R) implies that for all y_2, ε_1, x,

$$(T1.1) \qquad f_{y_2|\varepsilon_1,x}(y_2; R) = f_{y_2|\varepsilon_1,x}(y_2; S)$$

$(T1.1)$ implies that the distribution of y_2 conditional on ε_1 and x, generated by either (S) of (R) must be the same. In (R),

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

where the second equality holds 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

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

In (S),

$$\begin{aligned} \Pr(Y_2 \leq y_2 | \varepsilon_1, x) &= \Pr(s(x, \eta) \leq y_2 | x, \varepsilon_1) \\ &= \Pr(\eta \leq v(y_2, x) | x, \varepsilon_1) \\ &= F_{\eta|\varepsilon_1}(v(y_2, x)) \end{aligned}$$

Hence, for all y_2, x, ε_1

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

In particular, for any ε_1,

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

Note that the distribution of ε_2 conditional on ε_1 can be expressed as a nonparametric function $G(\varepsilon_2, \varepsilon_1)$, of two arguments. Analogously, the distribution of η conditional on ε_1 can be expressed as a nonparametric function $H(\eta, \varepsilon_1)$. Hence,

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

implies that for some functions G and H, both strictly increasing in their first arguments,

$$G\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right),\varepsilon_1\right)=H\left(s\left(y_2,x\right),\varepsilon_1\right)$$

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

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

Since G is strictly increasing in its first argument, it has an inverse, $\widetilde{G}$, conditional on the value of $r^1\left(y_1,y_2\right)$. This implies that

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

Hence, r^2 must be weakly separable into a function of (y_2,x) and on $r^1\left(y_1,y_2\right)$. Since H and $\widetilde{G}$ are both strictly increasing with respect to their first argument, r^2 must be strictly increasing into the function of $\left(y_2,x\right)$. We have shown that $(T1.1)$, and hence also the observational equivalence between (R) and (S), implies that $r^2(y_1,y_2,x)$ is separable into $r^1\left(y_1,y_2\right)$ and a function of $\left(y_2,x\right)$. In other words, we have shown that observational equivalence between Model (R) 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 (R), we let $\varepsilon_1=r^1\left(y_1,y_2\right)$ and $\eta=\widetilde{s}\left(y_2,x\right)$. Control function separability then implies that

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

for some function $\widetilde{v}$ that is strictly increasing in η. Hence, for some function $\widetilde{V}$,

$$\eta = \widetilde{V}(\varepsilon_2, \varepsilon_1)$$

where $\widetilde{V}$ is the inverse of $\widetilde{v}$ with respect to its first argument. Since $\eta = \widetilde{s}(y_2, x)$,

$$\widetilde{s}(y_2, x) = \eta = \widetilde{V}(\varepsilon_2, \varepsilon_1)$$

Since $\widetilde{s}$ is strictly increasing with respect to y_2, it follows that

$$y_2 = s(x, \eta) = s\left(x, \widetilde{V}(\varepsilon_2, \varepsilon_1)\right)$$

Since also

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

it follows that

$$y_2 = h^2(x, \varepsilon_1, \varepsilon_2) = s\left(x, \widetilde{V}(\varepsilon_1, \varepsilon_2)\right)$$

In other words, the reduced form for y_2 in Model (S) must be weakly separable into the unobservables of the structural system, $(\varepsilon_1, \varepsilon_2)$. 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

$$\begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= h^2(x, \varepsilon_1, \varepsilon_2) = s\left(x, \widetilde{V}(\varepsilon_1, \varepsilon_2)\right) \end{aligned}$$

In other words, the reduced form function for y_2 must be separable into $(\varepsilon_1, \varepsilon_2)$. Letting $\eta = \widetilde{V}(\varepsilon_1, \varepsilon_2)$, this implies that Model (S) and Model (R) are observationally equivalent.//

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.

An alternative characterization, which follows from the proof of Theorem 1, of systems where the function

$$\begin{aligned} y_2 &= \\ &\quad m^2\left(y_1, x, \varepsilon_2\right) \end{aligned}$$

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

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

is in terms of the reduced form function for y_2. The control function approach can be used if and only if the reduced form function, $h^2(x, \varepsilon_1, \varepsilon_2)$, for y_2 can be expressed as a function of x and a function of $(\varepsilon_1, \varepsilon_2)$. That is the control function approach can be used if and only if, for some functions s and $\widetilde{v}$

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

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

4. **Characterization in terms of Derivatives**

When the functions are differentiable, 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 (R) 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, $s_x = \partial s(y_2, x)/\partial x$ and $s_{y_2} = \partial s(y_2, x)/\partial y_2$ denote the derivatives of 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: *Model (S) is observationally equivalent to Model (R) if and only if for all* x, y_1, y_2,

$$\frac{r_x^2}{r_{y_1}^2\, m_{y_2}^1 + r_{y_2}^2} = \frac{s_x}{s_{y_2}}$$

Proof of Theorem 2: As in the proof of Theorem 1, observational equivalence between Model (R) and Model (S) implies that for all y_2, x, ε_1

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

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

$$\begin{aligned} f_{\varepsilon_2|\varepsilon_1}\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)\right)\ \left(r^2_{y_1}\ m^1_{y_2}\ +\ r^2_{y_2}\right) &= f_{\eta|\varepsilon_1}\left(s\left(y_2,x\right)\right)\ s_{y_2} \\ f_{\varepsilon_2|\varepsilon_1}\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)\right)\ r^2_x &= f_{\eta|\varepsilon_1}\left(s\left(y_2,x\right)\right)\ s_x \end{aligned}$$

where $f_{\varepsilon_2|\varepsilon_1} = f_{\varepsilon_2|\varepsilon_1}\left(r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)\right),\ f_{\eta|\varepsilon_1} = f_{\eta|\varepsilon_1}\left(s\left(y_2,x\right)\right),\ \ f_{\eta|\varepsilon_1} = f_{\eta|\varepsilon_1}\left(s\left(y_2,x\right)\right)$, and where, as defined above, $r^2_{y_1} = \partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial y_1$, $r^2_{y_2} = \partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial y_2$, $r^2_x = \partial r^2\left(m^1\left(y_2,\varepsilon_1\right),y_2,x\right)/\partial x$, $m^1_{y_2} = \partial m^1\left(y_2,\varepsilon_1\right)/\partial y_2$, $s_{y_2} = \partial s\left(y_2,x\right)/\partial y_2$, and $s_x = \partial s\left(y_2,x\right)/\partial x$.
Taking ratios, we get that

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

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

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

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 $b(y_2, x, \varepsilon_1)$ is a transformation of $s(y_2, x)$. Let $t(\cdot, \cdot, \varepsilon_1) : R \to R$ denote such a transformation. Then, for all y_2, x,

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

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

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

Hence, $(T2.1)$ implies control function separability. This implies, by Theorem 1, that Model (R) and Model (S) are observationally equivalent.//

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 reduced form functions. Differentiating with respect to y_1 and y_2 the identity

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

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

$$\frac{r^2_x}{r^2_{y_1}\; m^1_{y_2} \; + \; r^2_{y_2}} = \frac{s_x}{s_{y_2}}$$

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

$$\frac{r^1_{y_1}(y_1, y_2)\;\; r^2_x(y_1, y_2, x)}{r^1_{y_1}(y_1, y_2)\;\; r^2_{y_2}(y_1, y_2, x) - r^1_{y_2}(y_1, y_2)\;\; r^2_{y_1}(y_1, y_2, x)} = \frac{s_x(y_2, x)}{s_{y_2}(y_2, x)}$$

or

$$\frac{r^1_{y_1}(y_1,y_2)\ \ r^2_x(y_1,y_2,x)}{|r_y(y_1,y_2,x)|} = \frac{s_x(y_2,x)}{s_{y_2}(y_2,x)}$$

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

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)

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

5. 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 for situations where the specification in this example is realistic, one can estimate the structural equation using a contol function approach.

The objective function in our example is specified as

$$V(y_1,y_2,x_1,x_2,x_3) = (\varepsilon_1+\varepsilon_2)\ \ u(y_2) + \varepsilon_1\ \log(y_1 - u(y_2)) - y_1x_1 - y_2x_2 + x_3$$

This could be the objective function of a consumer choosing demand for three products, (y_1, y_2, y_3) subject to a linear budget contraint, $x_1 y_1 + x_2 y_2 + y_3 \leq x_3$, 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

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

$$(5.2)\ \frac{\partial}{\partial y_2}: \qquad (\varepsilon_1 + \varepsilon_2)\, u'(y_2) - u'(y_2) \frac{\varepsilon_1}{(y_1 - u(y_2))} - x_2 = 0$$

The Hessian of the objective function is

$$\begin{bmatrix} \frac{-\ \varepsilon_1}{(y_1 - u(y_2))^2} & \frac{\varepsilon_1\ u'(y_2)}{(y_1 - u(y_2))^2} \\ \frac{\varepsilon_1\ u'(y_2)}{(y_1 - u(y_2))^2} & (\varepsilon_1 + \varepsilon_2 - x_1)\, u''(y_2) \end{bmatrix}$$

At the values of (y_1, y_2) that satisfy the First Order conditions, where $\varepsilon_1 / [y_1 - u(y_2)] = x_1$, this Hessian is

$$= \begin{bmatrix} -(y_1 - u(y_2))^{-1} & u'(y_2)\ (y_1 - u(y_2))^{-1} \\ u'(y_2)\ (y_1 - u(y_2))^{-1} & (\varepsilon_1 + \varepsilon_2 - x_1)\, u''(y_2) \end{bmatrix}$$

Hence, the second order conditions for maximization, requiring that the Hessian is negative definite, will be satisfied as long as

$$(\varepsilon_1 + \varepsilon_2 - x_1) > \left| \frac{[u'(y_2)]^2}{[y_1 - u(y_2)]\, u''(y_2)} \right| = \left| \frac{\frac{\partial \log[y_1 - u(y_2)]}{\partial y_2}}{\frac{\partial \log[u'(y_2)]}{\partial y_2}} \right|$$

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,

$$\begin{aligned} \varepsilon_2 &= \frac{p_2}{u'(y_2)} - y_1 + u(y_2) + 1 \\ &= \left(\frac{p_2}{u'(y_2)} + 1 \right) - (y_1 - u(y_2)) \end{aligned}$$

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

$$\begin{aligned} \varepsilon_1 &= [y_1 - u(y_2)]\ x_1 \\ \varepsilon_2 &= \left(\frac{p_2}{u'(y_2)} + 1 \right) - (y_1 - u(y_2)) \end{aligned}$$

satisfy control function separability. The system of *reduced form functions*, which can then be estimated using a control function for nonseparable models, as in Chesher (2003) and Imbens and Newey (2003) is

$$
\begin{aligned}
y_1 &= u(y_2) + \frac{\varepsilon_1}{x_1} \\
y_2 &= (u')^{-1}\left(\frac{x_2}{\varepsilon_1 + \varepsilon_2 - x_1}\right)
\end{aligned}
$$

5. Conclusions

We have provided a conclusive answer to the question of when is it 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, is presented.

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.

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.

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 (2004): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," mimeo, Department of Economics, M.I.T.

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.

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 (2003) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," mimeo, M.I.T.

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.

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.