Control Functions in Nonseparable Simultaneous Equations Models[1]

Richard Blundell[2]
UCL & IFS
and
Rosa L. Matzkin[3]
UCLA

June 2013

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 paper, 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 the editor and referees, Whitney Newey, Arthur Lewbel, Yong Hyeon Yang, and participants at the Cowles-Cemmap conference in "Advancing Applied Microeconometrics," (Beijing, August 2010), the Cemmap-Northwestern conference in honor of Joel Horowitz (UCL, July 2011), and the California Econometrics Conference (USC, September 2011)) for helpful comments. The usual disclaimer applies.

[2] Blundell would like to thank the ESRC Centre for the Mircoeconomic Analysis of Public Policy at IFS for finacial support.

[3] Matzkin would like to gratefully acknowledge support from NSF Grant SES-1062090.

1. Introduction

Economic models of agent's optimization problems or of interactions among agents often exhibit simultaneity. It is well known that any function in which an explanatory variable is partly determined by the dependent variable of the function, can not be identified without additional information. Typically this additional information is provided by observable exogenous variables or functional structures.

Consider the identification and estimation of the derivative of the unknown function m^1 with respect to the scalar y_2 in the structural model

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

where m^1 is strictly increasing in the scalar ε_1 and where it is suspected or known that y_2 is itself a function of the scalar y_1. One approach to identify this derivative proceeds by using an observable scalar instrument, x, independent of ε_1 and functional dependent with y_2. Newey and Powell (1989, 2003), Darrolles, Florens, and Renault (2002), Ai and Chen (2003), and Hall and Horowitz (2003), Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007) and Chen and Pouzo (2012) follow this instrumental variable approach. Identification requires additional conditions on the relationship between y_2 and x.[4] Estimation requires dealing with the ill-posed inverse problem.

Another approach involves describing the source of simultaneity, by spec-

[4] See Chen, Chernozhukov, Lee, and Newey (2011) for the most up to date identification results for these models. See also Han and Ridder (2010) regarding identification of models using conditional moment restrictions.

ifying some function m^2 and a scalar unobservable ε_2, such that

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

where m^2 is strictly increasing with respect to ε_2 and where x is independent of $(\varepsilon_1, \varepsilon_2)$. Identification can then be analyzed in terms of conditions on the structural system composed by (m^1, m^2) and the distributions of $(\varepsilon_1, \varepsilon_2)$ and x. In this discussion all variables are scalars. Pointwise estimation of the derivative of m^1 with respect to y_2 can be performed without facing ill-posed inverse problems. Roehrig (1988), Benkard and Berry (2004, 2006), and Matzkin (2005, 2008, 2010a) follow this approach. Identification and estimation require additional restrictions on either the structural functions (m^1, m^2) or the density of $(\varepsilon_1, \varepsilon_2)$.

The control function approach assumes that the simultaneous system can be expressed in the triangular form

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

where the function s is strictly increasing in the unobservable scalar η and where x is independent of (ε_1, η).[5] Identification of various features of this triangular model has been widely studied, under different sets of assumptions, including Newey, Powell, and Vella (1999), Chesher (2003), Florens,

[5]Heckman (1978) references the Telser (1964) 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 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.

Heckman, Meghir and Vytlacil (2008), Imbens and Newey (2009), Torgovitsky (2011), and D'Haultfuille and Fevrier (2012), among others. Conditions for pointwise identification and estimation of the derivatives of m^1 have been derived and again avoid an ill-posed inverse problem.

Although the control function approach is attractive because of its simplicity, it requires the condition that the simultaneous system can be expressed in a triangular form. The question we aim to answer is the following: Suppose that we are 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 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)?

If (S) and (T) are observationally equivalent, then the average structural function of m^1, defined by Blundell and Powell (2003) as

$$G(y_2) = \int m^1(y_2, \varepsilon_1) \quad f_{\varepsilon_1}(\varepsilon_1)\ d\varepsilon_1$$

can be derived from the distribution of (Y_1, Y_2, X) as

$$G(y_2) = \int E\left(Y_1 | Y_2 = y_2, T = t\right)\ f_T(t)\ dt$$

where

$$T = F_{Y_2|X}(Y_2).$$

The local average response function, defined by Altonji and Matzkin (2001, 2005) as

$$\beta(y_2) = \int \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2}\ f_{\varepsilon_1|Y_2=y_2}(\varepsilon_1)\ d\varepsilon_1$$

can be derived from the distribution of (Y_1, Y_2, X) as

$$\beta(y_2) = \int \frac{\partial E\left(Y_1 | Y_2 = y_2, T = t\right)}{\partial y_2}\ f_{T|Y_2=y_2}(t)\ dt$$

where T is as defined above. The quantile structural function, defined by Imbens and Newey (2003) for the $\tau-$th quantile of ε_1, $q_{\varepsilon_1}(\tau)$, as

$$r^1(y_1, y_2) = \Pr\left(m^1\left(Y_2, q_{\varepsilon_1}(\tau)\right) \leq y_1 | Y_2 = y_2\right)$$

can be derived from the distribution of (Y_1, Y_2, X) by

$$r^1(y_1, y_2) = \int \Pr\left(Y_1 \leq y_1 | Y_2 = y_2, T = t\right)\ f_T(t)\ dt$$

The derivative of m^1 at (y_2, ε_1) for $\varepsilon_1 = r^1(y_1, y_2)$, can be derived from the distribution of (Y_1, Y_2, X), following Chesher (2003), by

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \left[\frac{\partial F_{Y_1|Y_2=y_2,T=t}(y_1)}{\partial y_1}\Big|_{t=F_{Y_2|X=x}(y_2)}\right]^{-1}\left[\frac{\partial F_{Y_1|Y_2=y_2,T=t}(y_1)}{\partial y_2}\Big|_{t=F_{Y_2|X=x}(y_2)}\right]$$

or, in terms of the distribution of the observable variables,

$$\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2} =$$

$$-\left[\frac{\partial F_{Y_1|Y_2=y_2,X=x}(y_1)}{\partial y_1}\right]^{-1}\left[\frac{\partial F_{Y_1|Y_2=y_2,X=x}(y_1)}{\partial y_2}\right]$$

$$+\left[\frac{\partial F_{Y_1|Y_2=y_2,X=x}(y_1)}{\partial y_1}\right]^{-1}\left[\frac{\partial F_{Y_1|Y_2=y_2,X=x}(y_1)}{\partial x}\right]\left[\left(\frac{\partial F_{Y_2|X=x}(y_2)}{\partial x}\right)\right]^{-1}\left[\left(\frac{\partial F_{Y_2|X=x}(y_2)}{\partial y_2}\right)\right]$$

In what follows we first define a new property of functions, *control function separability.* This is a condition states that y_1 is weakly separable from x in the structural inverse function r^2. 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. This condition is satisfied by simultaneous linear models with additive errors. In nonlinear models, this condition should be checked, since it is a strong assumption and if it is not satisfied, estimation using a control function approach may be severely inconsistent. Roughly, the condition states that the structural unobservable random term, ε_2, in the second equation of the simultaneous equations system can be represented as a function of the reduced form unobservable random term, η, in the second equation of the triangular system and the unobservable random term, ε_1 in the first equation of the simultaneous equations system. An example of a utility function whose system of demand functions satisfies control function separability is presented in Section 4. In Section 5 we describe 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

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

with $y_1, y_2, x, \varepsilon_1, \varepsilon_2 \in R$, 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 monotone in ε_1; and for all values (y_1, x) of (Y_1, X), the function m^2 is strictly monotone in ε_2.*

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

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

The technical assumptions S.1-S.3 could be partially relaxed at the cost of making the presentation more complex. Depending on the object of interest, the continuous differentiability in Assumption S.1 may be satisfied only on particular neighborhood of the support of the variables. The independence condition in Assumption S.2 may be satisfied only conditionally on some external variable. For example, if for a variable Z, the conditional density of (Y_1, Y_2, X) given $Z = z$ is identified, and if $(\varepsilon_1, \varepsilon_2)$ is independent of X conditional on $Z = z$, then our results can be extended to such situation. In many situations, the continuity of the densities of (Y_1, Y_2) and of $(\varepsilon_1, \varepsilon_2)$ given $X = x$ may be required to hold only on some neighborhoods of the supports of the variables.

Assumption S.3 is a weakening of the full support condition in Matzkin (2008). 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 value for (y_1, y_2), given any $X = x$. In other words, these assumptions guarantee the existence of a 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 Lemma 1 below.

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

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

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 $(\varepsilon_1, \varepsilon_2)$

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

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 $x, \varepsilon_1, \varepsilon_2$, $\partial h^1\left(x,\varepsilon_1,\varepsilon_2\right)/\partial \varepsilon_1 > 0$ and $\partial h^2\left(x,\varepsilon_1,\varepsilon_2\right)/\partial \varepsilon_2 > 0$. This completes the proof of Lemma 1.//

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

Definition: *Let* $(\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X})$ *denote the support of* (y_1, y_2, x). *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* $q : R^2 \to R$ *and* $v : R^2 \to R$ *such that for all* $(y_1, y_2, x) \in (\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X})$,

(a)

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

(b) q *is strictly increasing in its first argument, and*

(c) v *is strictly increasing in its first argument.*

Control function separability is weaker than the standard triangular specification, where the unobservable variable in the second structural equation, ε_2, is determined only by y_2 and x. Control function separability allows this unobservable variable, ε_2, to be determined by y_1 as well as by y_2 and x. However, the way in which y_1 affects the value of ε_2 is very limited. The dependent variable y_1 in the first structure equation can determine the value of the unobservable variable ε_2 in the second equation only through a function of (y_1, y_2), which is the same function through which (y_1, y_2) determine the value of ε_1. In other words, in control function separability, y_1 determines the value of ε_2 only indirectly, through the effect of y_1 on ε_1 and that of ε_1 on ε_2.

2.1. The triangular model and observational equivalence

We will consider triangular models of the form

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

with $y_1, y_2, x, \varepsilon_1, \eta \in R$, 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 monotone in* ε_1; *and for all values of* x, *the function* s *is strictly monotone 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

$$\begin{aligned} f_{\varepsilon_1,Y_2|X=x}(\varepsilon_1, y_2; T) &= f_{Y_1,Y_2|X=x}\left(m^1(y_2,\varepsilon_1), y_2; T\right)\left|\frac{\partial m^1(y_2,\varepsilon_1)}{\partial \varepsilon_1}\right| \\ &\text{and} \\ f_{\varepsilon_1,Y_2|X=x}(\varepsilon_1, y_2; S) &= f_{Y_1,Y_2|X=x}\left(m^1(y_2,\varepsilon_1), y_2; S\right)\left|\frac{\partial m^1(y_2,\varepsilon_1)}{\partial \varepsilon_1}\right|. \end{aligned}$$

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

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

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

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

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

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

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 of the uniqueness of equilibrium. At equilibrium, conditional on $X = x$, (y_1, y_2) is mapped into $(\varepsilon_1, \varepsilon_2) = (r^1(y_1, y_2), r^2(y_1, y_2, x))$, and $(\varepsilon_1, \varepsilon_2)$ is mapped into $(h^1(x, \varepsilon_1, \varepsilon_2), h^2(x, \varepsilon_1, \varepsilon_2))$. In other words, along the curve of all the values (y_1', y_2') for which $\varepsilon_1 = r^1(y_1', y_2')$, the value of $r^2(y_1', y_2', x)$ is equal to ε_2 only when $(y_1', y_2') = (y_1, y_2)$. Similarly, when ε_1 and y_2 are given, the only values of ε_2' and y_1' for which

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

are $(\varepsilon_2', y_1') = (\varepsilon_2, y_1) = (r^2(y_1, y_2, x), m^1(y_2, \varepsilon_1))$. But then, because of the strict monotonicity of $h^2(x, \varepsilon_1, \varepsilon_2')$ in ε_2', it must be that, given ε_1 and y_2, $\varepsilon_2 = \widetilde{h}(x, \varepsilon_1, y_2) = r^2(y_1, y_2, x)$. Substituting $y_1 = m^1(y_2, \varepsilon_1)$, we get that $\widetilde{h}(x, \varepsilon_1, y_2) = r^2(m^1(y_2, \varepsilon_1), y_2, x)$, which shows that the third equality is satisfied. 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 $\Pr(Y_2 \le y_2 | \varepsilon_1 = r^1(y_1, y_2), X = x)$

from Model (T) and from Model (S), we can conclude that for all y_2, x, ε_1

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

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

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

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 in $[\eta_L,\eta_U]$ and maps $[\eta_L,\eta_U]$ onto $[0,1]$. Hence, $(T1.1)$ 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 $f_{Y_1,Y_2|X=x}(y_1,y_2) > 0$

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

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

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

This implies that r^2 is weakly separable into $r^1(y_1, y_2)$ and a function of (y_2, x), 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 $\widetilde{s}$. We can then 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 functions $q : R^2 \to R$ and $v : R^2 \to R$ such that for all $(y_1, y_2, x) \in (\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X})$,

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

where on $(\mathcal{Y}_2 \times \mathcal{X})$, q is strictly increasing in its first argument, and for each for each $(y_1, y_2, x) \in (\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X})$, $\quad v(q(y_{2,}x), r^1(y_1, y_2))$ is strictly increasing in $q(y_{2,}x)$.

Let $\varepsilon_1 = r^1(y_1, y_2)$ and $\overline{\eta} = q(y_2, x)$. 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 $\widetilde{v}$ 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,

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

Since $\overline{\eta}$ is a function of $\left(\varepsilon_1, \varepsilon_2\right)$, Assumption S.2 implies Assumption T.2. Since also

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

it follows that

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

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

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

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

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

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

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

where $r^1_{y_1} = r^1_{y_1}\left(y_1, y_2\right)$ defines the partial derivative of r^1 with respect to its first argument. Similarly $r^2_{y_2} = r^2_{y_2}\left(y_1, y_2, x\right)$, $r^1_{y_2} = r^1_{y_2}\left(y_1, y_2\right)$, 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

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

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

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

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.

Alternative characterizations

An alternative characterization of systems where one of the functions can be estimated using a control function approach can be given in terms of the derivatives of the functions of Models (T) and (S). Let $r_x^2 = \partial r^2\left(y_1, y_2, x\right)/\partial x$, $r_{y_1}^2 = \partial r^2\left(y_1, y_2, x\right)/\partial y_1$, and $r_{y_2}^2 = \partial r^2\left(y_1, y_2, x\right)/\partial y_2$ denote the derivatives of r^2, $s_x = \partial s\left(y_2, x\right)/\partial x$ and $s_{y_2} = \partial s\left(y_2, x\right)/\partial y_2$ 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. The following theorem, whose proof is presented in the Appendix, provides one such characterization.

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,

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

In the Appendix, we show that in terms of the derivatives of the inverse system of structural equations of Model (S), the condition in Theorem 2 can be expressed as

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

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

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

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

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

Our arguments show that 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.[6]

[6]Kasy (2010) also highlights the one-dimensional distribution condition 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.

4. An example

We next provide two examples in the context of a consumer optimization problem with unobserved heterogeneity. In the first example, the utility is such that the first order conditions for maximization satisfy control function separability. In the second example, control function separability is not satisfied. In both examples, the utility function, $U(\varepsilon_1, \varepsilon_2, y_1, y_2, y_3)$, for goods y_1, y_2, and y_3, and for consumer with unobservable taste $(\varepsilon_1, \varepsilon_2)$ has the recursive structure:

$$U(\varepsilon_1, \varepsilon_2, y_1, y_2, y_3) = V(\varepsilon_1, \varepsilon_2, y_2) + W(\varepsilon_1, y_1, y_2) + y_3$$

where the unknown functions $V(\varepsilon_1, \varepsilon_2, \cdot)$ and $W(\varepsilon_1, \cdot, \cdot)$ are such that $U(\varepsilon_1, \varepsilon_2, \cdot, \cdot, y_3)$ is strictly increasing and strictly concave in (y_1, y_2). The typical consumer, characterized by $(\varepsilon_1, \varepsilon_2)$, chooses the quantity of (y_1, y_2, y_3) by maximizing $U(\varepsilon_1, \varepsilon_2, y_1, y_2, y_3)$ subject the the linear budget constraint, $x_1y_1 + x_2y_2 + y_3 \leq x_3$, where x_1 and x_2 denote the prices of, respectively, one unit of y_1 and y_2 and where x_3 denotes the income of the consumer The price of a unit of the third good is normalized to 1. Since $U(\varepsilon_1, \varepsilon_2, \cdot, \cdot, \cdot)$ is strictly increasing in (y_1, y_2, y_3), the optimal quantities will satisfy the budget constraint with equality. Substituting $y_3 = x_3 - y_1x_1 - y_2x_2$, we get that the optimal quantities of (y_1, y_2) must maximize the unconstrained function

$$\begin{aligned}
& \overline{U}(\varepsilon_1, \varepsilon_2, y_1, y_2, x_1, x_2, x_3) \\
= \; & V(\varepsilon_1, \varepsilon_2, y_2) + W(\varepsilon_1, y_1, y_2) - y_1x_1 - y_2x_2 + x_3
\end{aligned}$$

To provide an example where control function separability is satisfied, we let $u(\cdot)$ denote a strictly increasing and strictly concave function, and specify

$$V(\varepsilon_1, \varepsilon_2, y_2) = (\varepsilon_1 + \varepsilon_2)\ \ u(y_2)$$

and

$$W(\varepsilon_1, y_1, y_2) = \varepsilon_1\ \ \log(y_1 - u(y_2))$$

The first order conditions with respect to y_1 and y_2 for maximization of

$$\begin{aligned} &\overline{U}(\varepsilon_1, \varepsilon_2, y_1, y_2, x_1, x_2, x_3) \\ = \quad &(\varepsilon_1 + \varepsilon_2)\ \ u(y_2) + \varepsilon_1\ \ \log(y_1 - u(y_2)) - y_1 x_1 - y_2 x_2 + x_3 \end{aligned}$$

are

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

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

The Hessian of the objective function is

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

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, $\varepsilon_1/\left(y_1 - u\left(y_2\right)\right) = x_1$ and $\left(\varepsilon_1 + \varepsilon_2 - \left(\varepsilon_1/\left(y_1 - u\left(y_2\right)\right)\right)\right) u'\left(y_2\right) = x_2$, the objective function is strictly concave at values of (y_1, y_2) that satisfy the First Order Conditions as long as $x_1 > 0$ and $x_2 > 0$. Since the demanded quantities for (y_1, y_2) are optimal, they satisfy the simultaneous system of equations given by the First Order Conditions. To express those conditions in the form of an indirect system of equations, Note that from (5.1), we get

$$\text{(5.3)} \qquad \varepsilon_1 = \left[y_1 - u\left(y_2\right)\right]\ x_1$$

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

$$\text{(5.4)} \qquad \left[\left(\varepsilon_1 + \varepsilon_2\right) - x_1\right] u'\left(y_2\right) = x_2$$

Hence,

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

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

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

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

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.

To modify the example so that control function separability is not satisfied, suppose that V is specified as above but W is instead given by

$$W(\varepsilon_1, y_1, y_2) = \varepsilon_1 \; w(y_1, y_2)$$

for a strictly increasing and strictly concave function w. The First Order conditions for optimization become

(5.5) $\frac{\partial}{\partial y_1}$: $\qquad \varepsilon_1 \; w_{y_1}(y_1, y_2) - x_1 = 0$

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

The system of simultaneous equations that (y_1, y_2) satisfy can then be expressed as

$$(5.7) \qquad \varepsilon_1 = \frac{x_1}{w_{y_1}(y_1, y_2)}$$

and

$$(5.8) \qquad \varepsilon_2 = \frac{x_2}{u'(y_2)} - \varepsilon_1 \left[1 + \frac{w_{y_2}(y_1, y_2)}{u'(y_2)} \right]$$

Unless $-w_{y_2}(y_1, y_2) / w_{y_1}(y_1, y_2) = u'(y_2)$, as in the previous example, then in general, control function separability will not be satisfied.

5. Simultaneity in Latent Variables

Our results can be applied to models with simultaneity in continuous latent variables. These are models with "no structural shifts" in Heckman (1978), which do not allow for explanatory variables that are dummy endogenous variables. More specifically, our results can be applied in simultaneous equations models specified as

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

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

$$\begin{aligned} y_1 &= T_1(y_1^*, y_2^*) \\ y_2 &= T_2(y_1^*, y_2^*) \end{aligned} \qquad (2)$$

Suppose that m^1 and m^2 satisfy assumptions S1,...,S5 and also control function separability. Then, by Theorem 1, (1) can be written equivalently as the triangular model

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

satisfying T.1,...T.4. Identification in the model

$$\begin{aligned} y_1 &= T_1\left(y_1^*, y_2^*\right) \\ y_2 &= T_2\left(y_1^*, y_2^*\right) \end{aligned} \tag{3}$$

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

can then be analyzed using known techniques for models with latent variables and triangularity.

To provide an example, consider the binary response model with simultaneity

$$\begin{aligned} y_1^* &= g(y_2) - \varepsilon_1 \\ y_2 &= \Lambda\left(\beta\ y_1^* + \gamma\ x\ + \varepsilon_2\right) \end{aligned}$$

$$y_1 = \begin{cases} 1 \quad if \qquad y_1^* \geq 0 \\ \\ 0 \qquad\quad otherwise \end{cases}$$

where only the random vector (y_1, y_2, x), whose support is $\{0,1\} \times R^2$, is observed, Λ is strictly increasing, and where the functions g and Λ and the parameters β and γ are unknown. Usually in this model one is interested in objects such as the probability that $y_1 = 1$ when ε_1 is assumed to be distributed with its marginal distribution (Blundell and Powell (2004)) or the derivative of the probability that $y_1 = 1$ with respect to y_2 when the distribution of ε_1 conditional on y_2 is kept unchanged (Altonji and Matzkin (2005)). In an empirical application in Blundell and Powell (2004), y_1^* is the hours of work of one of the spouses in a household and y_2 is the income of the other spouse. The family makes the joint decision of (y_1^*, y_2).

Assume that the function $\Lambda^{-1}(y_2) \; - \beta \; g(y_2)$ is strictly increasing in y_2, and that Assumptions S1,...,S5 are satisfied.. Then, the system

$$\begin{aligned} y_1^* &= g(y_2) - \varepsilon_1 \\ y_2 &= \Lambda\left(\beta \; y_1^* + \gamma \; x \; + \varepsilon_2\right) \end{aligned}$$

satisfies control function separability. To see this, note that the random variable ε_2 in this system is determined by the function

$$\varepsilon_2 = \left[\Lambda^{-1}(y_2) \; - \beta \; g(y_2) \; - \gamma \; x\right] + \; \beta \; \left(g(y_2) - y_1^*\right),$$

which is separable into the function $[\Lambda^{-1}(y_2) \; - \beta \; g(y_2) \; - \gamma \; x]$ and the function $\beta \; \left(g(y_2) - y_1^*\right)$. It then follows by Theorem 1, that the system is observationally equivalent to a triangular system

$$\begin{aligned} y_1^* &= g(y_2) - \varepsilon_1 \\ y_2 &= s(x, \eta) \end{aligned}$$

satisfying assumptions T1,...,T4. Given the distribution of η, or of (η, y_2) (Matzkin (2003), Imbens and Newey (2009)) one can obtain the probability that $y_1 = 1$ when ε_1 is distributed with its marginal distribution as

$$F_{\varepsilon_1}\left(g\left(y_2\right)\right) = \int \; \Pr\left(y_1 = 1 | y_2, \eta\right) \; f_\eta\left(\eta\right) \; d\eta$$

The derivative with respect to y_2 of the probability that $y_1 = 1$ when the conditional distribution of ε_1 stays fixed, can be calculated as

$$\frac{\partial F_\delta\left(g(y_2)\right)}{\partial y_2} = \int \; \frac{\partial \Pr\left(y_1 = 1 | y_2, \eta\right)}{\partial y_2} \; f\left(\eta | y_2\right) \; d\eta.$$

where δ denotes a random term having the same distribution as that of ε_1 conditional on y_2. (See Blundell and Powell (2004) and Altonji and Matzkin (2005) for details.)

Alternatively, one can employ our results to identify g or its derivative with respect to y_2, separately from the distribution of ε_1, by first identifying the distribution of the latent variables and then proceeding with that distribution as if the latent variables were observed. For single equations binary response models, with all the explanatory variables being independent of the unobservable ε_1, the conditions under which such procedure can be done were given in Cosslett (1983) for linear in parameters g and in Matzkin (1992) for nonparametric g. Matzkin (1992) showed pointwise identification under shape restrictions on g, such as homogeneity of degree one or additivity, and without requiring large support conditions. Matzkin and Newey (1993) used the conditions in Matzkin (1992) to develop an estimator which followed the two step procedure. Still under independence conditions, Briesch, Chintaguna, and Matzkin (1997, 2009) considered binary response models with nonparametric random functions g. Identification in more general limited dependent variable models and under conditional independence, rather than full independence, and with linear in parameters g, were developed in Lewbel (2000), where the power of having a special regressor, conditionally independent of the unobservables given the other regressors, and with large support was shown. A large literature developed since then using special regressors.

For models with simultaneity

$$\begin{aligned} y_1^* &= m^1\left(y_2^*, \varepsilon_1\right) \\ y_2^* &= m^2\left(y_1^*, x, \varepsilon_2\right). \end{aligned} \tag{4}$$

where instead of observing $\left(y_1^*, y_2^*\right)$, one observes a transformation, $\left(y_1, y_2\right)$, of (y_1^*, y_2^*) defined by a known vector function (T_1, T_2),

$$\begin{aligned} y_1 &= T_1(y_1^*, y_2^*) \\ y_2 &= T_2(y_1^*, y_2^*) \end{aligned} \tag{5}$$

Matzkin (2012) showed identification using additional regressors (w_1, w_2) satisfying the conditions as in Matzkin (1992). Specifically, the vector (w_1, w_2, x) was assumed to be distributed independently of $(\varepsilon_1, \varepsilon_2)$, (w_1, w_2) entered the functions in known ways, and restrictions on the support of the continuous (w_1, w_2) conditional on x had only the effect of restricting the set of values (y_1, y_2, x) at which the functions or its derivatives were identified. Identification followed the two step procedure. This required attaching to each y_1^* and y_2^* one of the continuous regressors, w_1 and w_2, in a known way. Specifically, Matzkin (2012) assumed that the additional regressors (w_1, w_2) are observed, and that the simultaneous model is weakly separable into known functions $b_1(y_1^*, w_1)$ and $b_2(y_2^*, w_2)$. In other words, the model was assumed to be

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

with the restriction that for some unknown functions $\overline{m}^1$ and $\overline{m}^2$ and known functions $b_1(y_1^*, w_1)$ and $b_2(y_2^*, w_2)$

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

The restrictions on the support of (w_1, w_2, x) depended on the range of the known transformations (T_1, T_2) and on the elements of $\overline{m}^1$ and $\overline{m}^2$ that one is interested in identifying. For example, if y_2^* is observed, so that $y_2 = y_2^*$, or in other words,

$$T_2\left(y_1^*, y_2^*\right) = y_2$$

then w_2 is not needed. Or, if y_2^* is only observed when it is positive, the support of w_2 may only be the half real line. The most important feature is that the system can be expressed as

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

where the distribution of $\left(\bar{b}_1, \bar{b}_2, x\right) = \left(b_1\left(y_1^*, w_1\right), b_2\left(y_2^*, w_2\right), x\right)$ is continuous. Since the mapping between $\left(m^1, m^2\right)$ and $\left(\overline{m}^1, \overline{m}^2\right)$ is known, identification of pointwise features in $\left(m^1, m^2\right)$ can be obtained from identification of analogous features in $\left(\overline{m}^1, \overline{m}^2\right)$.

To provide some insight to the methods, consider first a single equation binary response model. Suppose that (w_1, y_2) is independent of ε_1, the conditional distribution of w_1 is continuously distributed given $y_2 = 0$, and $g(0) = 0$. Matzkin (1992, Example 3 in Section 5) shows identification of g in the model

$$y_1^* = w_1 + g(y_2) - \varepsilon_1$$
$$y_1 = \begin{cases} 1 \quad if \qquad y_1^* \geq 0 \\ \\ 0 \qquad\qquad otherwise \end{cases}$$

by applying her Theorem 1. In her theorem, the support of w_1 is not required to be unbounded. As mentioned above, the effect of a smaller support is to restrict the set of values of y_2 at which the function g is identified. Identification of F_{ε_1} follows because

$$\Pr\left(y_1 = 1 | w_1, y_2 = 0\right) = F_{\varepsilon_1}\left(w_1\right)$$

while, once F_{ε_1} is identified in the relevant support, $g(y_2)$ is identified, given any value of w_1, by

$$g(y_2) = F_{\varepsilon}^{-1}\left(\Pr\left(y_1 = 1 | w_1, y_2\right)\right) - w_1.$$

Consider now the binary response model with simultaneity and random function g^1

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

$$y_1 = \begin{cases} 1 \quad if \quad y_1^* \geq 0 \\ \\ 0 \qquad otherwise \end{cases}$$

where (w_1, x) is independent of $(\varepsilon_1, \varepsilon_2)$. An example of such model is where y_1^* is the utility of one of the spouses from working, y_2 is the work income of the other spouse, and w_1 is an exogenous variable that increases the utility y_1^* for one of the spouses but decreases the amount of work income of the other. Denote $\bar{b}_1 = y_1^* - w_1$. Assume that g^1 is invertible in ε_1 and g^2 is invertible in ε_2. Then, for some functions r^1 and r^2,

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

Note that $\bar{b}_1$ takes the place of y_1 in the model with continuous dependent variables. The only problem is that the distribution of $\left(\bar{b}_1, y_2\right)$ is not directly observed. To identify the distribution of distribution of $\left(\bar{b}_1, y_2\right)$, Matzkin (2012) follows the argument in Lewbel (2000). Assume that (x, w_1) 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, $\left(\overline{b}_1, y_2\right)$ 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 $\left(\overline{b}_1, y_2\right)$ conditional on X is identified. The analysis of the system

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

when the distribution of $\left(\overline{b}_1, y_2, x\right)$ is identified is analogous to the analysis of the system

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

when the distribution of (y_1, y_2, x) is given.

Suppose that the system (g^1, g^2) satisfies control function separability, in the sense that it can be expressed as

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

for some functions v and s, each strictly increasing in its first argument, then one can identify and estimate g^1 using a control function approach.

6. Conclusions

In this paper 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 defined 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 showed that this a strong condition, 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 provided 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 a heterogeneous consumer optimization problem that satisfies control function separability, was presented. By slightly modifying the example, we have shown the restrictiveness of the control function separability condition. 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.

Appendix A

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, ε_1

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

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

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

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

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

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

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

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.//

Alternative expression for $(T2.1)$

Instead of characterizing observationally equivalence in terms of the derivatives of the functions m^1 and r^2, as in $(T2.1)$,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

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

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

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

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

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

or

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

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 $\left(y_1, y_2\right)$.

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)

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

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.

BLUNDELL, R. and J. L. POWELL (2004) "Endogeneity in Semiparametric Binary Response Models", *Review of Economic Studies*, Vol; 71(3) No, 248, 581-913, July.

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 (2009) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics,* forthcoming.

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

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

CHEN, X, and D. POUZO (2012), “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments”, *Econometrica,* 80, 277-321.

D’HAULTFUILLE, X. and P. FEVRIER (2012) "Identification of Nonseparable Models with Endogeneity and Discrete Instruments," mimeo.

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," *Econometrica*, 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," Review of Economics and Statistics (forthcoming).

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.

HAUSMAN, J.A. and W. NEWEY (2011), "Nonparametric Consumer Surplus Estimation with General Heterogeneity," presented at the Cemmap-Northwestern Conference in Honor of Joel Horowitz.

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. (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. (2009) “Estimation of Nonparametric Models with Simultaneity," mimeo, UCLA.

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

MATZKIN, R.L. and W NEWEY (1993) "Kernel Estimation for Nonparametric Limited Dependent Variable Models," Mimeo, Northwestern University.

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,* 71(5), 1565-1578.

NEWEY, W., J. POWELL and F. VELLA (1999), “Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica,* 67(3), 565–603.

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.

TORGOVITSKY, A. (2012) "Identification of Nonseparable Models with General Instruments," mimeo, Northwestern University.