IDENTIFICATION IN A CLASS OF NONPARAMETRIC SIMULTANEOUS EQUATIONS MODELS

By

Steven T. Berry and Philip A. Haile

March 2011
Revised April 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1787R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Identification in a Class of Nonparametric Simultaneous Equations Models*

Steven T. Berry
Yale University
Department of Economics
Cowles Foundation
and NBER

Philip A. Haile
Yale University
Department of Economics
Cowles Foundation
and NBER

April 5, 2011

Abstract

We consider identification in a class of nonseparable nonparametric simultaneous equations models introduced by Matzkin (2008). These models combine standard exclusion restrictions with a requirement that each structural error enter through a "residual index" function. We provide constructive proofs of identification under several sets of conditions, demonstrating tradeoffs between restrictions on the support of the instruments, restrictions on the joint distribution of the structural errors, and restrictions on the form of the residual index function.

*We received helpful comments from Alex Torgovitzky. We also thank Zhentao Shi for capable research assistance and the National Science Foundation for financial support.

1 Introduction

There is substantial recent interest in the identification of nonparametric economic models that feature endogenous regressors and nonseparable errors. For a simultaneous equations setting, a general nonparametric model can be written

$$m_j(Y, Z, U) = 0 \quad j = 1, \ldots, J \tag{1}$$

where $J \geq 2$, $Y = (Y_1, \ldots, Y_J) \in \mathbb{R}^J$ are the endogenous variables, $U = (U_1, \ldots, U_J) \in \mathbb{R}^J$ are the structural errors, and Z is a vector of exogenous variables. Assuming m is invertible in U, this system of equations can be written in its "residual" form

$$U_j = \rho_j(Y, Z) \quad j = 1, \ldots, J. \tag{2}$$

Unfortunately, there are no known identification results for this fully general model, and most recent work has considered a triangular restriction of (1) that rules out many important economic applications.

In this paper we consider identification in a class of fully simultaneous models introduced by Matzkin (2008). These models take the form

$$m_j(Y, Z, \delta) = 0 \;\; j = 1, \ldots, J.$$

where $\delta = (\delta_1(Z, X_1, U_1), \ldots, \delta_J(Z, X_J, U_J))'$ and

$$\delta_j(Z, X_j, U_j) = g_j(Z, X_j) + U_j. \tag{3}$$

Here $X = (X_1, \ldots, X_J) \in \mathbb{R}^J$ are observed exogenous variables specific to each equation and each $g_j(Z, X_j)$ is assumed to be strictly increasing in X_j.

This formulation respects traditional exclusion restrictions in that X_j is excluded from equations $k \neq j$ (e.g., a "demand shifter" enters only the demand equation). However, it restricts (1) by requiring X_j and U_j to enter through a "residual index" $\delta_j(Z, X_j, U_j)$. If we again assume invertibility of m (now in δ—see the examples below), we obtain the analog of (2),

$$\delta_j(Z, X_j, U_j) = r_j(Y, Z)_j \qquad j = 1, \ldots, J$$

or, equivalently,

$$r_j(Y, Z) = g_j(Z, X_j) + U_j \qquad j = 1, \ldots, J. \tag{4}$$

Below we provide several examples of important economic applications in which this structure can arise.

Matzkin (2008, section 4.2) considered a two-equation model of the form (4) and showed that it is identified when X has large support and the joint density of U satisfies certain shape restrictions.[1] Matzkin (2010), relying on the same proof of identification, considers estimation of a restricted version of this model, where each function δ_j is linear in X_j (with coefficient normalized to 1).[2] We provide a further investigation of identification in this class of models under several alternative sets of conditions.

We begin with the model and assumptions of Matzkin (2008). We offer a constructive proof of identification and show that the model is overidentified. We then move to the main contribution of the paper, where we show that there is a trade-off between assumptions on the support of X, on the joint density of U, and on the functions $g_j(Z, X_j)$.[3] We first show that Matzkin's (2008, 2010) large support assumption can be dropped if one modifies the density restriction. Here we provide two results. The first (Theorem 2) leaves each $g_j(Z, X_j)$ fully nonparametric and requires only arbitrarily little variation in the "instruments" X. However, like Matzkin's results, it requires a global restriction on the density of U. The second result (Theorem 3) imposes the linear residual index structure of Matzkin (2010) but

[1] Precise statements of these restrictions and other technical conditions are given below.

[2] In Matzkin (2010) the index structure and restriction $g_j(X_j) = X_j$ follow from Assumption 3.2 (see also equation T.3.1).

[3] To our knowledge these results are all new with exception of Theorem 4, which was first shown by Berry and Haile (2010) for a system of simultaneous equations obtained from a model of differentiated products supply and demand.

allows trade-offs between the strength of the density restrictions and the variation in the instruments. We then show (Theorem 4) that one can take this trade-off to the opposite extreme: under the linear index structure, by retaining the large support assumption, all restrictions on the joint density can be dropped. Finally, we explore an alternative rank condition for which we lack sufficient conditions on primitives, but which could in principle be checked in applications.

All our proofs are constructive; i.e., they provide a mapping from the observables to the functions that characterize the model. Constructive proofs can make clear how observable variation reveals the economic primitives of interest. They may also suggest possible estimation approaches, although that is a topic we leave for future work.

Prior Results for Nonparametric Simultaneous Equations Brown (1983), Roehrig (1988), Brown and Matzkin (1998), and Brown and Wegkamp (2002) have previously considered identification of simultaneous equations models, assuming one structural error per equation and focusing on cases where the structural model (1) can be inverted to solve for the "residual equation" (2). A claim made in Brown (1983) and relied upon by the others implied that traditional exclusion restrictions would identify the model when U is independent of Z. Benkard and Berry (2006) recently showed that this claim is incorrect, leaving uncertain the nonparametric identifiability of fully simultaneous models.

For models of the form (2) with U independent of Z, Matzkin (2008) provided a new characterization of observational equivalence and showed how this could be used to prove identification in several special cases. These included a linear simultaneous equations model, a single equation model, a triangular (recursive) model, and a fully simultaneous nonparametric model (her "supply and demand" example) of the form (4) with $J = 2$. To our knowledge, the last of these is the only prior result demonstrating identification in a fully simultaneous nonparametric model with nonseparable errors.

Relation to Transformation Models The model (4) considered here can be interpreted as a generalization of the transformation model to a system of simultaneous equations. The usual (single-equation) semiparametric transformation model (e.g., Horowitz (1996)) takes the form

$$t(Y_j) = Z_j\beta + U_j \tag{5}$$

where $Y_i \in \mathbb{R}$, $U_i \in \mathbb{R}$, and the unknown transformation function t is strictly increasing. In addition to replacing $Z_j\beta$ with $g_j(Z, X_j)$,[4] (4) generalizes (5) by (a) allowing a vector of outcomes Y to enter the unknown transformation function, (b) dropping the requirement of a monotonic transformation function, and (c) allowing most exogenous variables (all besides X) to enter the fully nonparametric transformation functions r_j.

Relation to Triangular Models Much recent work has focused on models with a triangular (recursive) structure (see, e.g., Chesher (2003), Imbens and Newey (2009), and Torgovitsky (2010)). A two-equation version of the triangular model is

$$\begin{aligned} Y_1 &= m_1(Y_2, Z, X_1, U_1) \\ Y_2 &= m_2(Z, X_1, X_2, U_2) \end{aligned}$$

with U_2 a scalar monotonic error and with X_2 excluded from the first equation. In a supply and demand system, for example, Y_1 might be the quantity of the good, with Y_2 being its price. The first equation would be the structural demand equation, in which case the second equation would be the reduced-form equation for price, with X_2 as a supply shifter excluded from demand. However, in a supply and demand context—as in many other traditional simultaneous equations settings—the triangular structure is difficult to reconcile with economic theory. Typically both the demand error and the supply error will enter the reduced form for price. Thus, one obtains a triangular model only in the special case that the two structural errors monotonically enter the reduced form for price through a single index.

[4] A recent paper by Chiappori and Komunjer (2009) considers a nonparametric version of the single-equation transformation model. See also the related paper by Berry and Haile (2009).

The triangular framework therefore requires that at least one of the reduced-form equations feature a monotone index of the all original structural errors. This is an index assumption that is simply different from the index restriction of the model we consider. Our structure arises naturally from a fully simultaneous structural model with a nonseparable residual index; the triangular model will be generated by other kinds of restrictions on the functional form of simultaneous equations models. Examples of simultaneous models that do reduce to a triangular system can be found in Benkard and Berry (2006), Blundell and Matzkin (2010) and Torgovitsky (2010). Blundell and Matzkin (2010) have recently provided a necessary and sufficient condition for the simultaneous model to reduce to the triangular model, pointing out that this condition is quite restrictive.

Outline We begin with some motivating examples in section 2. Section 3 then completes the setup of the model. Our main results are presented in sections 4 through 6, followed by our exploration of a rank condition in section 7.

2 Examples

Example 1. *Consider a nonparametric version of the classical simultaneous equations model, where the structural equations are given by*

$$Y_j = \Gamma_j \left(Y_{-j}, Z, X_j, U_j\right) \quad j = 1, \ldots, J.$$

Examples include classical supply and demand models or models of peer effects. The residual index structure is imposed by requiring

$$\Gamma_j \left(Y_{-j}, Z, X_j, U_j\right) = \gamma_j \left(Y_{-j}, Z, \delta_j \left(Z, X_j, U_j\right)\right) \quad \forall j$$

where $\delta_j \left(Z, X_j, U_j\right) = g_j \left(Z, X_j\right) + U_j$*. This model features nonseparable structural errors but requires them to enter the nonseparable nonparametric function* Γ_j *through the index*

$\delta_j(Z, X_j, U_j)$. *If each function* γ_j *is invertible (e.g., strictly increasing) in* $\delta_j(Z, X_j, U_j)$ *then one obtains (4) from the inverted structural equations by letting* $r_j = \gamma_j^{-1}$. *Identification of the functions* r_j *and* g_j *implies identification of* Γ_j.

Example 2. *Consider a nonparametric version of the Berry, Levinsohn, and Pakes (1995) model of differentiated products markets. Market shares of each product* j *in market* t *are given by*

$$S_{jt} = \sigma_j(P_t, g(X_t) + \xi_t) \tag{6}$$

where $g(X_t) = (g_1(X_{1t}) \ \cdots \ g_J(X_{Jt}))'$, $P_t \in \mathbb{R}^J$ *are the prices of products* $1, \ldots, J$, $X_t \in \mathbb{R}^J$ *is a vector of product characteristics (all other observables have been conditioned out), and* $\xi_t \in \mathbb{R}^J$ *is a vector of unobserved characteristics associated with each product* j *and market* t. *Prices are determined through oligopoly competition, yielding a reduced form pricing equation*

$$P_{jt} = \pi_j(X_t, g(X_t) + \xi, h(Z_t) + \eta_t) \quad j = 1, \ldots, J \tag{7}$$

where $Z_t \in \mathbb{R}^J$ *is a vector of observed cost shifters associated with each product (other observed cost shifters have been conditioned out), and* $\eta_t \in \mathbb{R}^J$ *is a vector of unobserved cost shifters. Parallel to the demand model,* h *takes the form* $h(Z_t) = (h_1(Z_{1t}) \ \cdots \ h_J(Z_{Jt}))'$, *with each* h_j *strictly increasing. Berry and Haile (2010) show that this structure follows from a nonparametric random utility model of demand and standard oligopoly models of supply under appropriate residual index restrictions on preferences and costs. Unlike Example 1, here the structural equations specify each endogenous variable (*S_{jt} *or* P_{jt}*) as a function of multiple structural errors. Nonetheless, Berry, Gandhi, and Haile (2011) and Berry and Haile (2010) show that the system can be inverted, yielding a* $2J \times 2J$ *system of equations*

$$\begin{aligned} g_j(X_{jt}) + \xi_{jt} &= \sigma_j^{-1}(S_t, P_t) \\ h_j(Z_{jt}) + \eta_{jt} &= \pi_j^{-1}(S_t, P_t) \end{aligned}$$

where $S_t = (S_{1t}, \ldots, S_{Jt})$, $P_t = (P_{1t}, \ldots, P_{Jt})$. *This system takes the form of (4). Berry* *and Haile (2010) show that identification of the functions σ_j^{-1} and π_j^{-1} for all j allows identification of demand, marginal costs, and the mode of imperfect competition among firms.*

Example 3. *Consider identification of a production function in the presence of unobserved shocks to the marginal product of each input. Output is given by $Q = F(Y, U)$, where $Y \in \mathbb{R}^J$ is a vector of inputs and $U \in \mathbb{R}^J$ is a vector of unobserved productivity shocks. Let P and W denote the (exogenous) prices of the output and inputs, respectively. The observables are (Q, P, W, Y). With this structure, input demand is determined by a system of first-order conditions*

$$p\frac{\partial F(y,u)}{\partial y_j} = w_j \quad j = 1, \ldots, J \tag{8}$$

whose solution can be written

$$y_j = \eta_j(p, w, u) \quad j = 1, \ldots, J.$$

Observe that the reduced form for each Y_j depends on the entire vector of shocks U. The index structure is imposed by assuming that each structural error U_j enters as a multiplicative shock to the marginal product of the associated input, i.e.,

$$\frac{\partial F(y,u)}{\partial y_j} = f_j(y)\, u_j$$

for some function f_j. The first-order conditions (8) then take the form (after taking logs)

$$\ln(f_j(y)) = \ln\left(\frac{w_j}{p}\right) - \ln(u_j) \quad j = 1, \ldots, J.$$

which have the form of our model (4). The results below will imply identification of the functions f_j and, therefore, the realizations of each U_j. Since Q is observed, this implies identification of the production function F.

3 Model

3.1 Setup

The observables are (Y, X, Z). The exogenous observables Z, while important in applications, add no complications to the analysis of identification. Thus, from now on we drop Z from the notation. All assumptions and results should be interpreted to hold conditional on a given value of Z.

Stacking the equations in (4), we then consider the model

$$r(Y) = g(X) + U \tag{9}$$

where $g = (g_1, \ldots, g_J)'$ and each g_j is a strictly increasing continuously differentiable function of X_j. We let $\mathcal{X} = \text{int}(\text{supp}(X))$, require $\mathcal{X} \neq \emptyset$, and assume that the cumulative distribution of X is strictly increasing on $\mathcal{X}$. We let $\mathcal{Y} = \text{int}(\text{supp}(Y))$. We assume r is twice continuously differentiable and one-to-one. The latent random variables U are independent of X and have a continuously differentiable joint density f_U with support $\mathbb{R}^J$. Finally, we assume that the determinant $|J(y)|$ of the Jacobian matrix

$$J(y) = \begin{bmatrix} \frac{\partial r_1(y)}{\partial y_1} & \cdots & \frac{\partial r_1(y)}{\partial y_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_J(y)}{\partial y_1} & \cdots & \frac{\partial r_J(y)}{\partial y_J} \end{bmatrix}$$

is nonzero for all $y \in \mathcal{Y}$.

Some useful implications of these assumptions are summarized in the following lemma.

Lemma 1. *(i) $\forall y \in \mathcal{Y}$, supp$(X|Y = y) =$supp$(X)$; (ii) $\forall x \in \mathcal{X}$,supp$(Y|X = x) =$supp$(Y)$; (iii) $\mathcal{Y}$ is path-connected.*

Proof. Part (i) follows from (9) and the assumption that U is independent of X with support $\mathbb{R}^J$. Because r is one-to-one, continuously differentiable, and has nonzero Jacobian determinant, it has a continuous inverse r^{-1} such that $Y = r^{-1}(g(X) + U)$. Since supp$(U|X) = \mathbb{R}^J$, part (ii) follows immediately while part (iii) follows from the fact that the image of a path-connected set (here $\mathbb{R}^J$) under a continuous mapping is path-connected. $\square$

3.2 Normalizations

We make two types of normalizations without loss.[5] First, we normalize the location and scale of the unobservables U_j. To do this, we use (9), take an arbitrary $x^0 \in \mathcal{X}$ and $y^0 \in \mathcal{Y}$ (recalling part (i) of Lemma 1) and set

$$\begin{aligned} r_j\left(y^0\right) - g_j(x_j^0) &= 0 \quad \forall j && (10) \\ \frac{\partial g_j\left(x_j^0\right)}{\partial x_j} &= 1 \quad \forall j. && (11) \end{aligned}$$

Second, since adding a constant κ_j to both sides of (9) would leave all relationships between (Y, X, U) unchanged, we can normalize the location of one of the functions r_j or g_j for each j. We therefore set

$$r_j\left(y^0\right) = 0 \quad \forall j. \tag{12}$$

With (10), this implies

$$g_j\left(x_j^0\right) = 0 \quad \forall j. \tag{13}$$

3.3 Change of Variables

All of our arguments below start with the standard strategy of relating the joint density of observables to the joint distribution of the unobservables U. Let $\phi\left(y, x\right)$ denote the (observable) conditional density of $Y|X$ evaluated at $y \in \mathcal{Y}$, $x \in \mathcal{X}$. This density exists under the conditions above and can be expressed as

$$\phi\left(y, x\right) = f_U\left(r\left(y\right) - g(x)\right)\left|J(y)\right|. \tag{14}$$

[5] Alternatively we could follow Matzkin (2008), who makes no normalizations in her supply and demand example, instead showing that the derivatives of r and g are identified up to scale.

We treat $\phi(y,x)$ as known for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

Taking logs of (14) and differentiating, we obtain

$$\frac{\partial \ln \phi(y,x)}{\partial x_j} = -\frac{\partial \ln f_U(r(y)-g(x))}{\partial u_j}\frac{\partial g_j(x_j)}{\partial x_j} \tag{15}$$

$$\frac{\partial \ln \phi(y,x)}{\partial y_k} = \sum_j \frac{\partial \ln f_U(r(y)-g(x))}{\partial u_j}\frac{\partial r_j(y)}{\partial y_k} + \frac{\partial \ln |J(y)|}{\partial y_k}. \tag{16}$$

Substituting (15) into (16) gives

$$\frac{\partial \ln \phi(y,x)}{\partial y_k} = \sum_j -\frac{\partial \ln \phi(y,x)}{\partial x_j}\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j)/dx_j} + \frac{\partial \ln |J(y)|}{\partial y_k}. \tag{17}$$

4 A Constructive Proof of Matzkin's Result

We begin by providing a constructive proof of the identification result in Matzkin (2008, section 4.2), which relies on the following additional assumptions.[6]

Assumption 1. $\text{supp}(g(X)) = \mathbb{R}^J$.

Assumption 2. $\exists \bar{u} \in \mathbb{R}^J$ such that $\frac{\partial f_U(\bar{u})}{\partial u_j} = 0\ \forall j$.

Assumption 3. For all j and almost all $\hat{u}_j \in \mathbb{R}$, $\exists\ \hat{u}_{-j} \in \mathbb{R}^{J-1}$ such that for $\hat{u} = (\hat{u}_j, \hat{u}_{-j})$, $\frac{\partial f_U(\hat{u})}{\partial u_j} \neq 0$ and $\frac{\partial f_U(\hat{u})}{\partial u_k} = 0\ \forall k \neq j$.

Theorem 1. *Under Assumptions 1–3, the model (r, g, f_U) is identified.*

Proof. For every $y \in \mathcal{Y}$, Assumptions 1 and 2 imply that there exists $\bar{x}(y)$ such that

$$\frac{\partial f_U(r(y) - g(\bar{x}(y)))}{\partial u_j} = 0 \quad \forall j.$$

[6] We allow $J > 2$ although this does not change the argument, as observed by Matzkin (2010). Our Assumption 3 is weaker than its analog in Matzkin (2008), which uses the quantifier "for all $\hat{u}_j$" instead of "for almost all $\hat{u}_j$." We interpret the weaker version as implicit in Matzkin (2008). The stronger version would rule out many standard densities; for example, with a standard gaussian distribution, $\frac{\partial f_u(\hat{u})}{\partial u_j} = 0$ for all $\hat{u}_{-j}$ when $\hat{u}_j = 0$.

With (15), the maintained hypothesis $\frac{\partial g_j(x_j)}{\partial x_j} > 0$ implies

$$\frac{\partial f_U\left(r\left(y\right)-g\left(x\right)\right)}{\partial u_j}=0 \;\text{ iff }\; \frac{\partial \ln \phi\left(y,x\right)}{\partial x_j}=0 \tag{18}$$

so $\bar{x}\left(y\right)$ may be treated as known for all $y \in \mathcal{Y}$. Further, by (16),

$$\frac{\partial \ln \phi\left(y,\bar{x}\left(y\right)\right)}{\partial y_k}=\frac{\partial \ln\left|J(y)\right|}{\partial y_k}$$

so we can rewrite (17) as

$$\frac{\partial \ln \phi\left(y,x\right)}{\partial y_k}-\frac{\partial \ln \phi\left(y,\bar{x}\left(y\right)\right)}{\partial y_k}=\sum_j -\frac{\partial \ln \phi\left(y,x\right)}{\partial x_j}\frac{\partial r_j\left(y\right)/\partial y_k}{\partial g_j\left(x_j\right)/\partial x_j}. \tag{19}$$

Take an arbitrary (j,x_j) and observe that with (18) and $U \perp\!\!\!\perp X$, Assumptions 1 and 3 imply that for almost all y there exists $\hat{x}^j\left(y,x_j\right) \in \mathbb{R}^J$ such that $\hat{x}^j_j\left(y,x_j\right)=x_j$ and

$$\frac{\partial \ln \phi\left(y,\hat{x}^j\left(y,x_j\right)\right)}{\partial x_j} \neq 0 \tag{20}$$

$$\frac{\partial \ln \phi\left(y,\hat{x}^j\left(y,x_j\right)\right)}{\partial x_k} = 0 \quad \forall k \neq j. \tag{21}$$

Since the derivatives $\frac{\partial \ln \phi(y,x)}{\partial x_\ell}$ are observed, the points $\hat{x}^j\left(y,x_j\right)$ can be treated as known. Taking $x_j = x_j^0$, (11), (19) and (21) yield

$$\frac{\partial \ln \phi\left(y,\hat{x}^j\left(y,x_j^0\right)\right)}{\partial y_k}-\frac{\partial \ln \phi\left(y,\bar{x}\left(y\right)\right)}{\partial y_k}=-\frac{\partial \ln \phi\left(y,\hat{x}^j\left(y,x_j^0\right)\right)}{\partial x_j}\frac{\partial r_j\left(y\right)}{\partial y_k} \quad k=1,\ldots,J.$$

By (20) and continuity of $\frac{\partial r_j(y)}{\partial y_k}$, these equations identify $\frac{\partial r_j(y)}{\partial y_k}$ for all j,k, and $y \in \mathcal{Y}$. Now fix Y at an arbitrary value $\tilde{y} \in \mathcal{Y}$. For any j and $x_j \neq x_j^0$, (19) and (21) yield

$$\frac{\partial \ln \phi\left(\tilde{y},\hat{x}^j\left(\tilde{y},x_j\right)\right)}{\partial y_k}-\frac{\partial \ln \phi\left(\tilde{y},\bar{x}\left(\tilde{y}\right)\right)}{\partial y_k}=-\frac{\partial \ln \phi\left(\tilde{y},\hat{x}^j\left(y,x_j\right)\right)}{\partial x_j}\frac{\partial r_j\left(\tilde{y}\right)/\partial y_k}{\partial g_j\left(x_j\right)/dx_j} \quad k=1,\ldots,J. \tag{22}$$

By (20), (22) uniquely determines $\partial g_j(x_j)/dx_j$ as long as the known value $\frac{\partial r_j(\tilde{y})}{\partial y_k}$ is nonzero for some k. This is guaranteed by the maintained assumption $|J(y)| \neq 0 \ \forall y \in \mathcal{Y}$. Thus, $\frac{\partial g_j(x)}{\partial x_j}$ is identified for all j and $x \in \mathcal{X}$. With the boundary conditions (12) and (13) and part (iii) of Lemma 1, we then obtain identification of the functions g_j and r_j. Identification of f_u then follows from (9). $\square$

The argument also makes clear that the model is overidentified, since the choice of $\tilde{y}$ before (22) was arbitrary.

Remark 1. *Under Assumptions 1–3, the model is testable.*

Proof. Solving (22) for $\partial g_j(x_j)/dx_j$ at $\tilde{y} = y'$ and at $\tilde{y} = y''$, we obtain the overidentifying restrictions

$$\frac{\frac{\partial \ln \phi\left(y', \hat{x}^j(y', x_j)\right)}{\partial x_j} \frac{\partial r_j(y')}{\partial y_k}}{\frac{\partial \ln \phi(y', \hat{x}^j(y', x_j))}{\partial y_k} - \frac{\partial \ln \phi(y', \bar{x}(y'))}{\partial y_k}} = \frac{\frac{\partial \ln \phi\left(y'', \hat{x}^j(y'', x_j)\right)}{\partial x_j} \frac{\partial r_j(y'')}{\partial y_k}}{\frac{\partial \ln \phi(y'', \hat{x}^j(y'', x_j))}{\partial y_k} - \frac{\partial \ln \phi(y'', \bar{x}(y''))}{\partial y_k}}$$

for all j, k, x_j and $y', y'' \in \mathcal{Y}$. $\square$

5 Identification without Large Support

A large support assumption (Assumption 1 above) is not essential. Drop Assumptions 1–3 and instead assume the following.[7]

Assumption 4. For all j, $\text{supp}(X_j)$ is convex.

Assumption 5. f_U is twice continuously differentiable, with $\frac{\partial^2 \ln f_U(u)}{\partial u \partial u'}$ nonsingular almost everywhere.

[7] For a twice differentiable function Ψ on $\mathbb{R}^J$, we use the notation $\frac{\partial^2 \Psi(z)}{\partial z \partial z'}$ to denote the matrix $\begin{bmatrix} \frac{\partial^2 \Psi(z)}{\partial z_1 \partial z_1} & \cdots & \frac{\partial^2 \Psi(z)}{\partial z_J \partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \Psi(z)}{\partial z_1 \partial z_J} & \cdots & \frac{\partial^2 \Psi(z)}{\partial z_J \partial z_J} \end{bmatrix}$.

Assumption 4 requires a weak notion of connected support for X, but allows this support to be arbitrarily small.[8] Assumption 5 is a density restriction satisfied by many standard joint probability distributions. One sufficient condition is that $\frac{\partial^2 \ln f_U(u)}{\partial u \partial u'}$ be negative definite almost everywhere—a restriction on the class of log-concave densities (see, e.g., Bagnoli and Bergstrom (2005) and Cule, Samworth, and Stewart (2010)). Examples of densities that violate this condition are those with flat or log-linear regions.

Theorem 2. *Under Assumptions 4 and 5, the model (r, g, f_U) is identified.*

Proof. For any $y \in \mathcal{Y}$, differentiating (16) at x^0 gives

$$\frac{\partial^2 \ln \phi\left(y, x^0\right)}{\partial y_k \partial x_\ell} = \sum_j -\frac{\partial^2 \ln f_u\left(r\left(y\right) - g\left(x^0\right)\right)}{\partial u_j \partial u_\ell} \frac{\partial r_j\left(y\right)}{\partial y_k} \quad \forall k, \ell. \tag{23}$$

Further, differentiating (15) at x^0 gives

$$\frac{\partial^2 \ln f_u\left(r\left(y\right) - g\left(x^0\right)\right)}{\partial u_j \partial u_\ell} = \frac{\partial^2 \ln \phi\left(y, x^0\right)}{\partial x_j \partial x_\ell}. \tag{24}$$

Thus, we obtain

$$A = B\ C$$

where $A = \frac{\partial^2 \ln \phi\left(y, x^0\right)}{\partial y \partial x'}$, $B = -\frac{\partial^2 \ln \phi\left(y, x^0\right)}{\partial x \partial x'}$, and $C = J\left(y\right)$. The matrices A and B are known and, given (24) and Assumption 5, B is invertible for almost all y. This gives identification of $\frac{\partial r_j(y)}{\partial y_k}$ for all j, k and $y \in \mathcal{Y}$. To complete the proof, observe that with each $\frac{\partial r_j(y)}{\partial y_k}$ known, evaluating (17) at x^0 identifies $\frac{\partial \ln |J(y)|}{\partial y_k}$ for all $k, y \in \mathcal{Y}$. So (17) can be rearranged as

$$D = -E\ F$$

[8] If $\text{supp}(X_j)$ were instead the union of two or more intervals, the argument below would still prove identification of r, of $\partial g_j\left(x_j\right)/\partial x_j$ for all j and x_j, and of each g_j on one of the intervals (that containing x_j^0). Identification of g_j on each additional interval would hold up to an additional unknown location parameter. This partial identification would be sufficient to answer some types of questions.

where $D = \frac{\partial}{\partial y}\left[\ln \phi(y,x) - \ln |J(y)|\right]$ and is known, $E = J(y)'$ and is known, and

$$F = \begin{pmatrix} \frac{\partial \ln \phi(y,x)}{\partial x_1} \frac{1}{\partial g_1(x_1)/\partial x_1} \\ \vdots \\ \frac{\partial \ln \phi(y,x)}{\partial x_J} \frac{1}{\partial g_J(x_J)/\partial x_J} \end{pmatrix}.$$

Since $\left|J(y)'\right|$ is nonzero, each element $\frac{\partial \ln \phi(y,x)}{\partial x_j} \frac{1}{\partial g_j(x_j)/\partial x_j}$ of the matrix F is uniquely determined at every $x \in \mathcal{X}, y \in \mathcal{Y}$. This implies that $\frac{\partial g_j(x_j)}{\partial x_j}$ is identified for all $x_j \in$supp(X_j) as long as for each such x_j the (known) value of $\frac{\partial \ln \phi(y,(x_j,x_{-j}))}{\partial x_j}$ is nonzero for some y and x_{-j}. To confirm this, take any x_j and any x_{-j} and suppose to the contrary that $\frac{\partial \ln \phi(y,(x_j,x_{-j}))}{\partial x_j} = 0$ for all y. Then by (15), $\frac{\partial \ln f_U(r(y)-g(x))}{\partial u_j} = 0$ for all y. This requires $\frac{\partial}{\partial y_k}\left(\frac{\partial \ln f_U(r(y)-g(x))}{\partial u_j}\right) = 0$ for all y, k, i.e.,

$$\sum_{\ell=1}^{J} \frac{\partial^2 \ln f_U(r(y) - g(x))}{\partial u_j \partial u_\ell} \frac{\partial r_\ell(y)}{\partial y_k} = 0 \quad \forall y, k.$$

Stacking these J equations, we obtain

$$J(y)' z = 0$$

where

$$z = \begin{pmatrix} \frac{\partial^2 \ln f_U(r(y)-g(x))}{\partial u_j \partial u_1} \\ \vdots \\ \frac{\partial^2 \ln f_U(r(y)-g(x))}{\partial u_j \partial u_J} \end{pmatrix}.$$

Since $J(y)'$ is full rank, this requires $z = 0$ for all y, which is ruled out by Assumption 5. This contradiction implies that $\frac{\partial g_j(x_j)}{\partial x_j}$ is identified for all j, x_j. The remainder of the proof then follows that for Theorem 1, using the boundary conditions (12) and (13) with Assumption 4. □

As with the assumptions of Theorem 1, Assumptions 4 and 5 lead to overidentification.

Remark 2. *Under Assumptions 4 and 5, the model is testable.*

Proof. In the final step (beginning with "To confirm…") of the proof of Theorem 2, we demonstrated that $\frac{\partial \ln \phi(y,(x_j,x_{-j}))}{\partial x_j}$ is nonzero for some y given any $x = (x_j, x_{-j})$, leading to identification of $\frac{\partial g_j(x_j)}{\partial x_j}$. Letting $\frac{\partial \tilde{g}_j(x_j;x_{-j})}{\partial x_j}$ be the value of $\frac{\partial g_j(x_j)}{\partial x_j}$ implied when $x = (x_j, x_{-j})$, we obtain the testable restrictions

$$\frac{\partial \tilde{g}_j\left(x_j; x'_{-j}\right)}{\partial x_j} = \frac{\partial \tilde{g}_j\left(x_j; x''_{-j}\right)}{\partial x_j} \quad \forall j, x_j, x'_{-j}, x''_{-j}.$$

□

We can weaken the global restriction on the density (Assumption 5) if we assume that each function g_j is known up to scale or, equivalently, that $g_j\left(x_j\right) = x_j \beta_j$.

Assumption 6. $g_j\left(x_j\right) = x_j \beta_j \ \forall j, x_j$.

Assumption 7. (i) f_U is twice continuously differentiable; and (ii) for almost all $y \in \mathcal{Y}$ there exists $x^*\left(y\right) \in \mathcal{X}$ such that the matrix $\frac{\partial^2 \ln f_U(r(y)-g(x^*(y)))}{\partial u \partial u'}$ is nonsingular.

With Assumption 6 we are still free to make the scale normalization (11); thus, without further loss we set $\beta_j = 1 \ \forall j$. The restricted model we consider here is then identical to that studied in Matzkin (2010). Assumption 7 weakens Assumption 5 by requiring invertibility of the matrix $\frac{\partial^2 \ln f_U(u)}{\partial u \partial u'}$ only at one (unknown) point in supp$(U|Y = y)$.

Theorem 3. *Under Assumptions 6 and 7 the model* (r, f_U) *is identified.*

Proof. Differentiation of (16) gives (after setting $g_j\left(x_j\right) = x_j$)

$$\frac{\partial^2 \ln \phi\left(y, x\right)}{\partial y_k \partial x_\ell} = \sum_j - \frac{\partial^2 \ln f_u\left(r\left(y\right) - x\right)}{\partial u_j \partial u_\ell} \frac{\partial r_j\left(y\right)}{\partial y_k} \quad \forall y, x, k, \ell.$$

Assumption 7 ensures that for almost all y, $\frac{\partial^2 \ln f_U(r(y)-x)}{\partial u \partial u'}$ is invertible at a point $x = x^*\left(y\right)$, giving identification of $\frac{\partial r_j(y)}{\partial y_k}$ for all $j, k, y \in \mathcal{Y}$. Identification of $r\left(y\right)$ then follows as in Theorem 1, using the boundary condition (12). Identification of f_U then follows from the equations $U_j = r_j(Y) - X_j$. □

This result offers a trade-off between assumptions on the support of X and restrictions on the density f_U. At one extreme, Assumption 7 holds with arbitrarily little variation in X when f_U satisfies Assumption 5. At the opposite extreme, with large support for X, Assumption 7 holds when there is a single point u^* at which $\frac{\partial^2 \ln f_U(u^*)}{\partial u \partial u'}$ is nonsingular. Between these extremes are cases in which $\frac{\partial^2 \ln f_U(u)}{\partial u \partial u'}$ is nonsingular in a neighborhood (or set of neighborhoods) that can be reached for any value of Y through the available variation in X.

6 Identification without Density Restrictions

The trade-off illustrated above can be taken to the opposite extreme. If we restrict attention to linear residual index functions by requiring $g_j(x_j) = x_j \beta_j$, then under the large support condition of Matzkin (2008) there is no need for any restriction on the joint density f_U. The following result was first given in Berry and Haile (2010) for a class of models of demand and supply in differentiated products oligopoly markets.

Theorem 4. *Under Assumptions 1 and 6, the model (r, f_u) is identified.*

Proof. Recall that we have normalized $\beta_j = 1 \; \forall j$ without loss. Since

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_U(r(y) - x) \; dx = 1,$$

from (14) we obtain

$$f_U(r(y) - x) = \frac{\phi(y, x)}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y, t) \; dt}.$$

Thus the value of $f_U(r(y) - x)$ is uniquely determined by the observables for all (y, x). Since

$$\int_{\tilde{x}_j \geq x_j, \tilde{x}_{-j}} f_U(r(y) - \hat{x}) \; d\hat{x} = F_{U_j}(r_j(y) - x_j) \tag{25}$$

the value of $F_{U_j}(r_j(y) - x_j)$ is identified for every (y, x). By the normalization (11),

$$F_{U_j}\left(r_j\left(y^0\right) - x_j^0\right) = F_{U_j}(0).$$

For any y we can then find the value $\overset{o}{x}(y)$ such that $F_{U_j}\left(r_j(y) - \overset{o}{x}(y)\right) = F_{U_j}(0)$, which reveals $r_j(y) = \overset{o}{x}(y)$. This identifies each function r_j. Identification of f_U then follows as in the previous results. □

The restricted model considered here is identical to that considered by Matzkin (2010). We have retained her large support condition but dropped all restrictions on derivatives of f_U. Thus, this result provides an even stronger foundation for estimation of this type of model, using the methods proposed in Matzkin (2010) or others.

7 A Rank Condition

Here we explore an alternative invertibility condition that is sufficient for identification and may allow additional trade-offs between the support of X and the properties of the joint density f_U. Like the classical rank condition for linear models (or completeness conditions for nonparametric models—e.g., Newey and Powell (2003) or Chernozhukov and Hansen (2005)) the condition we obtain is not easily derived from primitives. However, in principle it could be checked in applications.

For simplicity, we restrict attention here to the case $J = 2$. Fix $Y = y$ and consider seven values of X,

$$\begin{aligned} &x^0 = (x_1^0, x_2^0), \quad x^2 = (x_1', x_2^0), \\ &x^1 = (x_1^0, x_2'), \quad x^3 = (x_1', x_2'), \quad x^5 = (x_1'', x_2'), \\ &\qquad\qquad\qquad\quad\; x^4 = (x_1', x_2''), \quad x^6 = (x_1'', x_2'') \end{aligned} \tag{26}$$

where x^0 is as in (11), and $x_j'' \neq x_j' \neq x_j^0$. For $\ell \in \{0, 1, \ldots, 6\}$, rewrite (17) as

$$A_{\ell k} = B_{\ell 1} \frac{\partial r_1(y)/\partial y_k}{\partial g_1(x_1^\ell)/\partial x_1} + B_{\ell 2} \frac{\partial r_2(y)/\partial y_k}{\partial g_2(x_2^\ell)/\partial x_2} + \frac{\partial}{\partial y_k}|J(y)| \quad k = 1, 2 \tag{27}$$

where

$$
\begin{aligned}
A_{\ell k} &= \frac{\partial \ln \phi\left(y, x^{\ell}\right)}{\partial y_k} \\
B_{\ell j} &= \frac{\partial \ln \phi\left(y, x^{\ell}\right)}{\partial x_j}.
\end{aligned}
$$

$A_{\ell k}$ and $B_{\ell j}$ are known. Stacking the equations (27) obtained at all ℓ, we obtain a system of fourteen linear equations in the fourteen unknowns

$$
\begin{array}{cc}
\dfrac{\partial r_j\left(y\right)/\partial y_k}{\partial g_j\left(x_j\right)/\partial x_j} & j,k=1,2;\ x_j \in \left(x_j^0, x_j', x_j''\right) \\
\dfrac{\partial}{\partial y_k}\left|J\left(y\right)\right| & k=1,2.
\end{array}
\tag{28}
$$

These unknowns are identified if the 14 $\times$14 matrix

$$
\begin{bmatrix}
B_{01} & 0 & B_{02} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & B_{01} & 0 & B_{02} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & B_{12} & 0 & B_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & B_{12} & 0 & B_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
B_{21} & 0 & 0 & 0 & 0 & 0 & B_{22} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & B_{21} & 0 & 0 & 0 & 0 & 0 & B_{22} & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & B_{31} & 0 & B_{32} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & B_{31} & 0 & B_{32} & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & B_{41} & 0 & 0 & 0 & 0 & 0 & B_{42} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & B_{41} & 0 & 0 & 0 & 0 & 0 & B_{42} & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & B_{52} & 0 & B_{51} & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{52} & 0 & B_{51} & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{61} & 0 & B_{62} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{61} & 0 & B_{62} & 0 & 1
\end{bmatrix}
\tag{29}
$$

representing the known coefficients of the linear system (27) has full rank. This holds iff the determinant

$$
\begin{aligned}
&(B_{12}B_{31}B_{42}B_{51}B_{22}B_{01} - B_{12}B_{31}B_{62}B_{51}B_{22}B_{01} - B_{12}B_{31}B_{42}B_{61}B_{22}B_{01} \\
&+B_{21}B_{02}B_{42}B_{61}B_{52}B_{31} + B_{42}B_{61}B_{52}B_{31}B_{12}B_{01} - B_{42}B_{61}B_{52}B_{31}B_{12}B_{21} \\
&+B_{12}B_{51}B_{62}B_{41}B_{22}B_{01} - B_{21}B_{02}B_{11}B_{32}B_{42}B_{51} + B_{21}B_{02}B_{11}B_{32}B_{62}B_{51} \\
&-B_{21}B_{02}B_{32}B_{51}B_{62}B_{41} - B_{32}B_{51}B_{62}B_{41}B_{12}B_{01} + B_{32}B_{51}B_{62}B_{41}B_{12}B_{21} \\
&-B_{21}B_{02}B_{11}B_{42}B_{61}B_{52} + B_{21}B_{02}B_{11}B_{32}B_{42}B_{61})^2
\end{aligned}
\tag{30}
$$

is nonzero. With (17) and our normalizations, knowledge of $\frac{\partial |J(y)|}{\partial y_k}$ and $\frac{\partial r_j(y)/\partial y_k}{\partial g_j\left(x_j^0\right)/\partial x_j}$ for all y, j, and k leads to identification of the model following the arguments above. Thus, we can state the following proposition.

Proposition 5. *Suppose that for almost all $y \in \mathcal{Y}$ there exist points $x^0, x^1, \ldots, x^6$ with the structure (26) such that $x^\ell \in$supp$(X|Y = y)$ $\forall \ell = 0, 1, \ldots, 6$, and such that (30) is nonzero. Then the model (r, g, f_U) is identified.*

Our approach here exploits linearity of the system (27) in the ratios $\frac{\partial r_j(y)/\partial y_k}{\partial g_j\left(x_j^\ell\right)/\partial x_j}$ in order to provide a rank condition that is sufficient for identification, despite the highly nonlinear model. Two observations should be made, however. One is that we have not used all the information available from the seven values of X; in particular, we used only $\frac{\partial}{\partial y_k} |J(y)|$ and $\frac{\partial r_j(y)/\partial y_k}{\partial g_j\left(x_j^0\right)/\partial x_j}$ at each y, j,k to identify the model, yet the values of $\frac{\partial r_j(y)/\partial y_k}{\partial g_j\left(x_j^\ell\right)/\partial x_j}$ for $\ell \neq 0$ are also directly obtained by solving (27). This provides a set of overidentifying restrictions and suggests that it may be possible to obtain identification under weaker conditions. Second, at each value of y the 14 linear unknowns in (28) are determined by just 10 unknown values

$$\begin{array}{ll} \dfrac{\partial r_j\left(y\right)}{\partial y_k} & j,k=1,2 \\ \partial g_j\left(x_j\right)/\partial x_j & j=1,2;\ x_j\in\left(x_j',x_j''\right) \\ \dfrac{\partial}{\partial y_k}\left|J\left(y\right)\right| & k=1,2. \end{array}$$

Although conditions for invertibility of a nonlinear system are much more difficult to obtain, this again suggests overidentification, at least in some cases.

8 Conclusion

Simultaneous equations models play an important role in many economic applications. Unfortunately, identification results have been limited almost exclusively to parametric models or to settings admitting a recursive structure.

We have examined the identifiability of a class of nonparametric nonseparable simultaneous equations models with a residual index structure first explored by Matzkin (2008). The model incorporates standard exclusion restrictions and a requirement that each structural error enter the system through an index that also depends on the corresponding instrument. This is a significant restriction, but one that allows substantial generalization of standard functional form restrictions in a variety of economic contexts. With this structure, nonparametric identification can be obtained in a fully simultaneous system despite the challenges pointed out by Benkard and Berry (2006). Indeed, we have provided constructive proofs of identification for this model under several alternative sets of sufficient conditions, illustrating trade-offs between the assumptions one places on the support of instruments, on the joint density of the structural errors, and on the form of the residual index.

References

Bagnoli, M., and T. Bergstrom (2005): "Log-Concave Probability and its Applications," *Economic Theory*, 26, 445–469.

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

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60(4), 889–917.

Berry, S. T., A. Gandhi, and P. A. Haile (2011): "Connected Substitutes and Invertibility of Demand," Discussion paper, Yale University.

Berry, S. T., and P. A. Haile (2009): "Identification of a Nonparametrc Generalized Regression Model with Group Effects," Discussion paper, Yale University.

——— (2010): "Identification in Differentiated Products Markets Using Market Level Data," Discussion paper, Yale University.

Blundell, R., and R. L. Matzkin (2010): "Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models," Discussion Paper CWP28/10, cemmap.

Brown, B. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51(1), 175–96.

Brown, D. J., and R. L. Matzkin (1998): "Estimation of Nonparametric Functions in Simultaneous Equations Models with Application to Consumer Demand," Working paper, Yale Univeristy.

Brown, D. J., and M. H. Wegkamp (2002): "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica*, 70(5), pp. 2035–2051.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

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

Chiappori, P.-A., and I. Komunjer (2009): "Correct Specification and Identifiction of Nonparametric Transformation Models," Discussion paper, University of California San Diego.

Cule, M., R. Samworth, and M. Stewart (2010): "Maximum Likelihood Estimation of a Multidimensional Log-Concave Density," *Journal of the Royal Statistical Society, Series B*, 72, 545–607 (with discussion).

Horowitz, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103–137.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Matzkin, R. L. (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

——— (2010): "Estimation of Nonparametric Models with Simultaneity," Discussion paper, UCLA.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation in Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametic Models," *Econometrica*, 56(2), 433–47.

Torgovitsky, A. (2010): "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity," Discussion paper, Yale.