IDENTIFICATION IN A CLASS OF NONPARAMETRIC SIMULTANEOUS EQUATIONS MODELS

By

Steven T. Berry and Philip A. Haile

March 2011
Revised November 2013

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

November 19, 2013

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 some of the available tradeoffs between conditions 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.

1 Introduction

Economic theory typically produces systems of equations that characterize equilibrium outcomes that might be observable to empirical researchers. The classical supply and demand framework is the most familiar of such models, but systems of simultaneous equations arise in a wide variety of contexts in which multiple agents interact or a single agent makes interrelated choices. The identifiability of such models is therefore a fundamental question for a wide range of topics in empirical economics. Early work on identification treated systems of simultaneous equations as a primary focus.[1] For example, Fisher's (1966) monograph, entitled *The Identification Problem in Econometrics*, considered only identification of simultaneous equations models with the explanation (p. *vii*), "Because the simultaneous equation context is by far the most important one in which the identification problem is encountered, the treatment is restricted to that context."[2]

Although there has been substantial recent interest in the identification of nonparametric economic models that feature endogenous regressors and nonseparable errors, there remain remarkably few results for fully simultaneous systems. A general nonparametric simultaneous equations 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,[3] this system of equations can be written in its "residual" form

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

---

[1] Many prominent examples can be found in Cowles Commission Monograph 10 and Cowles Foundation Monograph 14.

[2] See also the discussion in Manski (1995).

[3] See, e.g., Palais (1959), Gale and Nikaido (1965), and Berry, Gandhi, and Haile (2013) for conditions that can be used to show invertibility in different contexts.

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

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

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.[4] Matzkin (2010) develops an estimation approach for such models, focusing

---

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

on the case in which each function δ_j is linear in X_j (with coefficient normalized to 1), and provides some additional identification results.[5] 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). Matzkin's analysis relied substantial new machinery—primarily, a new characterization of observational equivalence— and proved identification by contradiction. We start by showing neither is necessary: we offer a constructive proof using a standard change-of-variables technique. We also show that the model is overidentified. We then move to the main contribution of the paper, which focuses on the case in which $g_j(Z, X_j)$ is linear in X_j (as in Matzkin (2010)). We show that in this case there is a range of sufficient conditions that trade off assumptions on the support of X and restrictions on the joint density of U. We first show that Matzkin's (2008, 2010) large support assumption can be dropped if one modifies the density restriction. In fact, for a large class of density functions, the support of X can be arbitrarily small. We then show that one can also go to the opposite extreme: if one retains 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 whose satisfaction is verifiable.

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

---

[5] 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).

implied that traditional exclusion restrictions would identify the model when U is independent of Z. Benkard and Berry (2006) showed that this claim is incorrect, leaving uncertain the nonparametric identifiability of fully simultaneous models.

A major breakthrough in this literature was Matzkin (2008).[6] 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 $J = 2$. The last of these easily generalizes to $J > 2$. To our knowledge this was the first result demonstrating identification in a fully simultaneous nonparametric model with nonseparable errors. More recently, Matzkin (2010), while focused on estimation, has included constructive identification results for a model that could be extended to that we consider. Like us, she considers identification using a combination of restrictions on the support of X and on the joint density f_U.

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

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

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 $g_j\left(Z, X_j\right)$,[7] (4) generalizes (5) by dropping the requirement of a monotonic transformation function and, more fundamental, allowing a vector of outcomes Y to enter each unknown transformation function.

---

[6] See also Matzkin (2007).

[7] 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).

4

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

$$Y_1 = m_1(Y_2, Z, X_1, U_1)$$
$$Y_2 = m_2(Z, X_1, X_2, U_2)$$

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.

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*

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

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 $\delta_j \left(Z, X_j, U_j \right)$. If each function γ_j is invertible (e.g., strictly increasing) in $\delta_j \left(Z, X_j, U_j \right)$ 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*

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

where $g \left(X_t \right) = \left(g_1 \left(X_{1t} \right) \ \cdots \ g_J \left(X_{Jt} \right) \right)'$, $P_t \in \mathbb{R}^J$ are the prices of products $1, \dots, J$, $X_t \in \mathbb{R}^J$ 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.

6

Prices are determined through oligopoly competition, yielding a reduced form pricing equation

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

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 $h\left(Z_t \right) = \left(h_1\left(Z_{1t} \right) \cdots h_J\left(Z_{Jt} \right) \right)'$, with each h_j strictly increasing. Berry and Haile (2013) 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 (2013) and Berry and Haile (2013) show that the system can be inverted, yielding a $2J \times 2J$ system of equations

$$g_j \left(X_{jt} \right) + \xi_{jt} = \sigma_j^{-1} \left(S_t, P_t \right)$$
$$h_j \left(Z_{jt} \right) + \eta_{jt} = \pi_j^{-1} \left(S_t, P_t \right)$$

where $S_t = \left(S_{1t}, \ldots, S_{Jt} \right)$, $P_t = \left(P_{1t}, \ldots, P_{Jt} \right)$. This system takes the form of (4). Berry and Haile (2013) show that identification of the unknown functions in this system implies identification of demand, marginal costs, all structural errors, and the reduced form for equilibrium prices.

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 $Q = F\left(Y, U \right)$, where $Y \in \mathbb{R}^J$ is a vector of inputs and $U \in \mathbb{R}^J$ is a vector of unobserved factor-specific 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*

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

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 can be 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 $g(X) = (g_1(X_1), \ldots, g_J(X_J))'$. We let $\mathcal{X} = \text{int}(\text{supp}(X))$ and $\mathcal{Y} = \text{int}(\text{supp}(Y))$.

We maintain the following assumptions on the model throughout.

Assumption 1. (a) g is differentiable, with $\partial g_j(x_j)/\partial x_j > 0$ for all j, x_j;

(b) r is one-to-one on $\mathcal{Y}$, differentiable on $\mathcal{Y}$, and has nonsingular Jacobian matrix

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

for $y \in \mathcal{Y}$;

(c) U is independent of X and has positive joint density function f_U on $\mathbb{R}^J$.

The following result documents two useful implications of Assumption 1.

Lemma 1. *Under Assumption 1, (i) $\forall y \in \mathcal{Y}$, $supp(X|Y = y) = supp(X)$; and (ii) $\forall x \in \mathcal{X}$, $supp(Y|X = x) = supp(Y)$.*

Proof. Both claims follow immediately from (9) and the assumption that U is independent of X with support $\mathbb{R}^J$. $\square$

For some results we will strengthen the smoothness assumption on r, allowing us to exploit the following result.

Lemma 2. *Let Assumption 1 hold and suppose that $r \in \mathcal{C}^1$. Then $\mathcal{Y}$ is path-connected.*

Proof. Because r is one-to-one, continuously differentiable, and has nonzero Jacobian determinant, it has a continuous inverse r^{-1} on $\mathcal{Y}$ such that $Y = r^{-1}(g(X) + U)$. Since $supp(U|X) = \mathbb{R}^J$, the result 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 impose three standard normalizations.[8] First, observe that all relationships between (Y, X, U) would be unchanged if for some constant κ_j, $g_j(X_j)$ were replaced by $g_j(X_j) + \kappa_j$ while $r_j(Y)$ is replaced by $r_j(Y) + \kappa_j$. Thus, without loss, for an arbitrary $y^0 \in \mathcal{Y}$ we set

$$r_j(y^0) = 0 \quad \forall j. \tag{10}$$

Given this restriction, we still require normalizations on the location and scale of the unobservables U_j, as usual.[9] Since (9) would continue to hold if both sides were multiplied by a nonzero constant, we normalize the scale of U_j by taking an arbitrary $x^0 \in \mathcal{X}$ and setting

$$\frac{\partial g_j(x_j^0)}{\partial x_j} = 1 \quad \forall j. \tag{11}$$

And since (9) would be unchanged if $g_j(X_j)$ were replaced by $g_j(X_j) + \kappa_j$ for some constant κ_j while U_j is replaced by $U_j - \kappa_j$, we fix the location of U_j by setting

$$g_j(x_j^0) = 0 \quad \forall j. \tag{12}$$

3.3 Change of Variables

All of our arguments below start with the standard strategy of relating the joint distribution (or density) of observables to the that of the unobservables U.[10] Let $\phi(y, x)$ denote the (observable) conditional density of $Y|X$ evaluated at $y \in \mathcal{Y}$, $x \in \mathcal{X}$. This density exists

---

[8]We follow Horowitz (1982, p. 168-169), who makes equivalent normalizations in his semiparametric single-equation version of our model. 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.

[9]Often these restrictions are without loss as well, although one can imagine applications in which the location and/or scale of U_j has economic meaning.

[10]See, e.g., Koopmans (1945) and Hurwicz (1950).

under the conditions above and can be expressed as

$$\phi(y, x) = f_U(r(y) - g(x)) |J(y)| \tag{13}$$

or, equivalently

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

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

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). This relies on additional regularity conditions, as well as conditions on the support of $g(X)$ and on the joint density f_U.[11]

Assumption 2. f_U is differentiable, and r is twice differentiable.

Assumption 3. $\operatorname{supp}(g(X)) = \mathbb{R}^J$.

Assumption 4. $\exists \bar{u} \in \mathbb{R}^J$ such that $\frac{\partial f_U(\bar{u})}{\partial u_j} = 0 \ \forall j$.

Assumption 5. 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 1–5, the model (r, g, f_U) is identified.*

---

[11]We allow $J > 2$ although this does not change the argument, as observed by Matzkin (2010). Our Assumption 5 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, including multivariate normals. Matzkin (2010), by imposing the additional restriction $g_j(x_j) = x_j \forall j$, allows one to replace "for almost all $\hat{u}_j$" with "for some $\hat{u}_j$" with only minor adjustment to the proof. The same is true of our proof. The regularity conditions we employ here slightly weaken those assumed in Matzkin (2008, 2010).

Proof. Differentiating (14), 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

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

For every $y \in \mathcal{Y}$, Assumptions 3 and 4 imply that there exists $\bar{x}(y)$ such that

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

From (15) and $\frac{\partial g_j(x_j)}{\partial x_j} > 0$,

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

Since $\frac{\partial \ln \phi(y,x)}{\partial x_j}$ is known for all $y \in \mathcal{Y}, x \in \mathcal{X}$, $\bar{x}(y)$ may be treated as known for all $y \in \mathcal{Y}$. Further, by (16),

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

so we can rewrite (17) as

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

Take an arbitrary (j, x_j) and observe that with (18) and $U \perp\!\!\!\perp X$, Assumptions 3 and 5 imply

that for almost all y there exists $\hat{x}^j (y, x_j) \in \mathcal{X}$ such that $\hat{x}^j_j (y, x_j) = x_j$ and

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

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

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

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

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

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

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 (10) and (12) and Lemma 2, we then obtain identification of the functions g_j and r_j. Identification of f_u then follows from (9). $\qquad\square$

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

Remark 1. *Under Assumptions 1–5, 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

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

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

5   Identification without Large Support

In this section and the next, we impose linearity of each function g_j.

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

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).

We drop Assumptions 2–5 and instead assume the following.[12]

Assumption 7. $r \in C^1$.

Assumption 8. $\mathcal{X}$ is nonempty.

Assumption 9. (i) f_U is twice 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.

Assumption 7 weakens the smoothness condition on r required for Theorem 1. Assumption 8 replaces the large support assumption with a requirement that the support have nonempty interior. Assumption 9 requires that the log density have nonsingular Hessian matrix at points $u^* = r(y) - x$ reachable through the support of X. A strong sufficient condition is that $\partial^2 \ln f_U\left(u\right)/\partial u \partial u'$ be nonsingular almost everywhere; in that case, the support of X can be arbitrarily small. This sufficient condition for Assumption 9 is satisfied by many standard joint probability distributions. For example, it holds under when $\frac{\partial^2 \ln f_U(u)}{\partial u \partial u'}$ is negative definite almost everywhere—a property of the multivariate normal (see the Appendix)

---

[12]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

$$\begin{bmatrix} \frac{\partial^2 \Psi(z)}{\partial z_1 \partial z_1} & \cdots & \frac{\partial^2 \Psi(z)}{\partial z_J \partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \Psi(z)}{\partial z_1 \partial z_J} & \cdots & \frac{\partial^2 \Psi(z)}{\partial z_J \partial z_J} \end{bmatrix}.$$

and many other log-concave densities (see, e.g., Bagnoli and Bergstrom (2005) and Cule, Samworth, and Stewart (2010)). Examples of densities that violate this sufficient condition for Assumption 9 are those with flat (uniform) or log-linear (exponential) regions.

Theorem 2. *Under Assumptions 1 and 6–9, the model* (r, f_U) *is identified.*

Proof. Differentiation of (14) with respect to x_j and then y_k gives (after setting $g_j(x_j) = x_j$)

$$\frac{\partial^2 \ln \phi(y, x)}{\partial x_j \partial y_k} = \sum_\ell -\frac{\partial^2 \ln f_u(r(y) - x)}{\partial u_j \partial u_\ell} \frac{\partial r_\ell(y)}{\partial y_k} \quad \forall y, x, k, \ell. \tag{23}$$

Differentiating (14) with respect to x_j and then x_ℓ gives

$$\frac{\partial^2 \ln \phi(y, x)}{\partial x_j \partial x_\ell} = \frac{\partial^2 \ln f_u(r(y) - x)}{\partial u_j \partial u_\ell}$$

so that (23) can be rewritten

$$\frac{\partial^2 \ln \phi(y, x)}{\partial x_j \partial y_k} = \sum_\ell -\frac{\partial^2 \ln \phi(y, x)}{\partial x_j \partial x_\ell} \frac{\partial r_\ell(y)}{\partial y_k} \quad \forall y, x, k, \ell.$$

In matrix form, this yields

$$A(x, y) = B(x, y) \ J(y)$$

where $A(x, y) = \frac{\partial^2 \ln \phi(y, x)}{\partial x \partial y'}$, $B(x, y) = -\frac{\partial^2 \ln \phi(y, x)}{\partial x \partial x'}$. $A(x, y)$ and $B(x, y)$ are known for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Assumption 9 ensures that for almost all y, $B(x, y)$ is invertible at a point $x = x^*(y)$, giving identification of $J(y)$ and, thus, $\frac{\partial r_j(y)}{\partial y_k}$ for all $j, k, y \in \mathcal{Y}$. Identification of $r(y)$ then follows as in Theorem 1, using the boundary condition (10). Identification of f_U then follows from the equations $U_j = r_j(Y) - X_j$. $\qquad \square$

This result offers a trade-off between assumptions on the support of X and restrictions on the density f_U. At one extreme, Assumption 9 holds with arbitrarily small support for X when $\partial^2 \ln f_U(u) / \partial u \partial u'$ is nonsingular almost everywhere (see the discussion above). At the opposite extreme, with large support for X, Assumption 9 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

15

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

Maintaining the assumed linearity of each function g_j, the trade-off illustrated above can be taken to the opposite extreme: under the large support condition of Matzkin (2008) there is no need for a restriction on the joint density f_U.[13]

Theorem 3. *Under Assumptions 1, 3, and 6, the model (r, f_u) is identified.*

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

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

from (13) we obtain

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

Thus the value of $f_U \left(r\left(y\right) - x \right)$ is uniquely determined by the observables for all $x \in \mathbb{R}^J$, $y \in \mathcal{Y}$. Since

$$\int_{\tilde{x}_j \geq x_j, \tilde{x}_{-j}} f_U \left(r\left(y\right) - \hat{x} \right) \ d\hat{x} = F_{U_j} \left(r_j\left(y\right) - x_j \right) \tag{24}$$

the value of $F_{U_j} \left(r_j\left(y\right) - x_j \right)$ is identified for $x \in \mathbb{R}^J$, $y \in \mathcal{Y}$. By the normalization (11),

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

For any $y \in \mathcal{Y}$ we can then find the value $\overset{o}{x}\left(y\right)$ such that $F_{U_j} \left(r_j\left(y\right) - \overset{o}{x}\left(y\right) \right) = F_{U_j}\left(0\right)$, which reveals $r_j\left(y\right) = \overset{o}{x}\left(y\right)$. This identifies each function r_j on the support of Y. Identification of f_U then follows from the equations $U_j = r_j\left(Y\right) - X_j$. $\qquad \square$

---

[13] The argument used to show Theorem 3 was first used by Berry and Haile (2013) in combination with additional assumptions and arguments to demonstrate identification in models of differentiated products demand and supply.

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, but failure of this condition is testable.

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

$$
\begin{aligned}
x^0 &= \left(x_1^0, x_2^0\right), \quad x^2 = \left(x_1', x_2^0\right), \\
x^1 &= \left(x_1^0, x_2'\right), \quad x^3 = \left(x_1', x_2'\right), \quad x^5 = \left(x_1'', x_2'\right), \\
& \qquad x^4 = \left(x_1', x_2''\right), \quad x^6 = \left(x_1'', x_2''\right)
\end{aligned}
\tag{25}
$$

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

$$
A_{\ell k} = B_{\ell 1} \frac{\partial r_1(y)/\partial y_k}{\partial g_1\left(x_1^\ell\right)/\partial x_1} + B_{\ell 2} \frac{\partial r_2(y)/\partial y_k}{\partial g_2\left(x_2^\ell\right)/\partial x_2} + \frac{\partial}{\partial y_k}|J(y)| \quad k = 1, 2
\tag{26}
$$

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 (26) obtained at all ℓ, we obtain a system of fourteen linear equations in the fourteen unknowns

$$
\begin{aligned}
&\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j)/\partial x_j} && j, k = 1, 2;\ x_j \in \left(x_j^0, x_j', x_j''\right) \\
&\frac{\partial}{\partial y_k}|J(y)| && k = 1, 2.
\end{aligned}
\tag{27}
$$

These unknowns are identified if the 14×14 matrix

$$
\begin{bmatrix}
B_{01} & 0 & B_{02} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & B_{01} & 0 & B_{02} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & B_{12} & 0 & B_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & B_{12} & 0 & B_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
B_{21} & 0 & 0 & 0 & 0 & 0 & B_{22} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & B_{21} & 0 & 0 & 0 & 0 & 0 & B_{22} & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & B_{31} & 0 & B_{32} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & B_{31} & 0 & B_{32} & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & B_{41} & 0 & 0 & 0 & 0 & 0 & B_{42} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & B_{41} & 0 & 0 & 0 & 0 & 0 & B_{42} & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & B_{52} & 0 & B_{51} & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{52} & 0 & B_{51} & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{61} & 0 & B_{62} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{61} & 0 & B_{62} & 0 & 1
\end{bmatrix}
\tag{28}
$$

representing the known coefficients of the linear system (26) has full rank. This holds iff the determinant

$$
\begin{aligned}
(&B_{12}B_{31}B_{42}B_{51}B_{22}B_{01} - B_{12}B_{31}B_{62}B_{51}B_{22}B_{01} - B_{12}B_{31}B_{42}B_{61}B_{22}B_{01} \\
&+B_{21}B_{02}B_{42}B_{61}B_{52}B_{31} + B_{42}B_{61}B_{52}B_{31}B_{12}B_{01} - B_{42}B_{61}B_{52}B_{31}B_{12}B_{21} \\
&+B_{12}B_{51}B_{62}B_{41}B_{22}B_{01} - B_{21}B_{02}B_{11}B_{32}B_{42}B_{51} + B_{21}B_{02}B_{11}B_{32}B_{62}B_{51} \\
&-B_{21}B_{02}B_{32}B_{51}B_{62}B_{41} - B_{32}B_{51}B_{62}B_{41}B_{12}B_{01} + B_{32}B_{51}B_{62}B_{41}B_{12}B_{21} \\
&-B_{21}B_{02}B_{11}B_{42}B_{61}B_{52} + B_{21}B_{02}B_{11}B_{32}B_{42}B_{61})^2
\end{aligned}
\tag{29}
$$

is nonzero. With (17) and our normalizations, knowledge of $\frac{\partial |J(y)|}{\partial y_k}$ and $\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j^0)/\partial x_j}$ for all y,

18

j, and k leads to identification of the model following the arguments above. Thus, we can state the following proposition.

Proposition 4. *Let Assumption 1 hold and suppose that for almost all $y \in \mathcal{Y}$ there exist points $x^0, x^1, \ldots, x^6$ with the structure (25) such that $x^\ell \in supp(X|Y = y) \ \forall \ell = 0, 1, \ldots, 6$, and such that (29) is nonzero. Then the model (r, g, f_U) is identified.*

Our approach here exploits linearity of the system (26) in the ratios $\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j^\ell)/\partial x_j}$ 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 $\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j^0)/\partial x_j}$ at each y, j, k to identify the model, yet the values of $\frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j^\ell)/\partial x_j}$ for $\ell \neq 0$ are also directly obtained by solving (26). 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 (27) are determined by just 10 unknown values

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

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 for simultaneous equations models 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.

Appendix: The Multivariate Normal

Matzkin (2010) includes two new identification result (Theorems 4.1 and 4.2) that do not require large support for X. Like our Theorem 2, these results require a combination of support and density restrictions, with the required support dependent on the true density.

We show here that with multivariate normal errors, the density restriction we use in Theorem 2 is satisfied by the normal. In fact, the normal satisfies a much stronger condition that allows Theorem 2 to deliver identification when X has arbitrarily small support. In contrast, with a multivariate normal, Matzkin's density requirements fail regardless of the support of X.

Matzkin's (2010) results concern a simplified version of the model with only 2 equations and an instrument in only one equation:

$$
\begin{aligned}
U_1 &= r_1(Y) \\
U_2 &= r_2(Y) + X.
\end{aligned}
$$

We discuss her conditions for identification of this model, not those one might use to extend her argument to the full model.

Matzkin's density condition for her Theorem 4.1 is the following:

Assumption 4.5. The density, f_U of (U_1, U_2) is such that for all u_1, there exists at least one value $u_2^*(u_1)$ such that
$$\frac{\partial^2 \log f_U (u_1, u_2^* (u_1))}{\partial u_2 \partial u_2} = 0.$$

At any such value $\frac{\partial^2 \log f_U (u_1, u_2^* (u_1))}{\partial u_1 \partial u_2} \neq 0$.

For her Theorem 4.2, a different combination of density and support conditions is used. The density restriction is:

Assumption 4.5′. The density, f_U of (U_1, U_2) is such that for all u_1, there exist distinct values $u_2^*(u_1)$ and $u_2^{**}(u_1)$ such that

$$\frac{\partial \log f_U (u_1, u_2^* (u_1))}{\partial u_2} = \frac{\partial \log f_U (u_1, u_2^{**} (u_1))}{\partial u_2} = 0.$$

At any such values, $\frac{\partial \log f_U (u_1, u_2^* (u_1))}{\partial u_1} \neq \frac{\partial \log f_U (u_1, u_2^{**} (u_1))}{\partial u_1}$.

It is immediate that this latter condition fails when f_U is a multivariate normal density where, given any u_1 there is a unique u_2 such that $\frac{\partial \log f_U (u_1, u_2)}{\partial u_2} = 0$ (the unique maximizer of the likelihood for U_2 given $U_1 = u_1$). Thus we consider only her Assumption 4.5 below.

Suppose
$$f_U (u) = (2\pi)^{-\frac{J}{2}} |\Sigma|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (u - \mu)' \Sigma^{-1} (u - \mu) \right]$$

where Σ is a nonsingular covariance matrix. This implies

$$\frac{\partial^2 \log f_U (u)}{\partial u \partial u'} = -\Sigma^{-1}.$$

Remark 2. *There is no point $u \in \mathbb{R}^J$ at which $\frac{\partial^2 \log f_U (u)}{\partial u_j \partial u_j} = 0$ for any j. Thus, Assumption 4.5 fails.*

Remark 3. *Since $-\Sigma^{-1}$ has inverse $-\Sigma$, our Assumption 9 holds.*

Note also that since Σ is a nonsingular covariance matrix, it is positive definite. This means that $\frac{\partial^2 \log f_U(u)}{\partial u \partial u'}$ is everywhere negative definite. As noted in the text, this allows Theorem 2 to deliver identification with arbitrarily small support for X.

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 (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81, 2087–2111.

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

——— (2013): "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).

FISHER, F. M. (1966): *The Identification Problem in Econometrics*. Robert E. Krieger Publishing, Huntington, NY.

GALE, D., AND H. NIKAIDO (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

HOOD, W. C., AND T. C. KOOPMANS (eds.) (1953): *Studies in Econometric Method, Cowles Foundation Monograph 10*. Yale University Press, New Haven.

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

HURWICZ, L. (1950): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10*, ed. by T. C. Koopmans, pp. 245–257. John Wiley and Sons, New York.

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

KOOPMANS, T. C. (1945): "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, 40, 448–466.

——— (ed.) (1950): *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10.* John Wiley and Sons, New York.

MANSKI, C. (1995): *Identification Problems in the Social Sciences.* Harvard U. Press, Cambridge, MA.

MATZKIN, R. L. (2007): "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

——— (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.

PALAIS, R. S. (1959): "Natural Operations on Differential Forms," *Transactions of the American Mathematical Society*, 92, 125–141.

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.