Identification of Nonparametric Simultaneous Equations Models with a Residual Index Structure*

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 9, 2014

Abstract

We present new results on the identifiability of a class of nonseparable nonparametric simultaneous equations models introduced by Matzkin (2008, 2010). These models combine standard exclusion restrictions with a requirement that each structural error enter through a "residual index." We provide three general results nesting a variety of special cases that permit tradeoffs between exogenous variation required of instruments and restrictions on the joint density of the structural errors. Among these corollaries are results with no density restriction and results allowing instruments with arbitrarily small support. Matzkin's (2008, 2010) identification results for this class of models are also obtained as special cases. We show that our primary conditions for identification are verifiable (i.e., their satisfaction or failure is identified) and that the maintained assumptions of the model imply falsifiable restrictions.

*This paper supersedes our 2011 and 2013 working papers, entitled "Identification in a Class of Nonparametric Simultaneous Equations Models," which included special cases of results shown here. Some of these results grew out of our related work on differentiated products markets and benefited from the comments of audiences at several university seminars, the 2008 World Congress of the Game Theory Society, 2008 LAMES, 2009 Econometrics of Demand Conference, 2009 FESAMES 2010 Guanghua-CEMMAP-Cowles Advancing Applied Microeconometrics Conference, 2010 French Econometrics Conference, 2011 LAMES/LACEA, and 2012 ESEM. We also received helpful comments from Xiaohong Chen, Elie Tamer, Alex Torgovitzky, and seminar/conference audiences at the LSE, 2013 Econometrics of Demand Conference, 2014 SITE Summer Conference, 2014 NASM of the Econometric Society, and 2014 IAAE Conference. We thank the National Science Foundation for financial support.

1 Introduction

Economic theory typically produces systems of equations characterizing the outcomes observable to empirical researchers. The classical supply and demand model is a canonical example, but systems of simultaneous equations arise in many economic contexts in which multiple agents interact or a single agent makes multiple interrelated choices. The identifiability of simultaneous equations models is therefore an important question for a wide range of topics in empirical economics. This paper provides new results on the nonparametric identifiability of such models.

Early work on (parametric) identification in econometrics treated systems of simultaneous equations as a primary focus. Prominent examples can be found among the contributions to Koopmans (1950) and Hood and Koopmans (1953). Fisher's (1966) monograph, *The Identification Problem in Econometrics*, illustrates the extent of this focus: he considered only identification of simultaneous equations models, explaining (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."

Nonparametric identification, on the other hand, has remained a significant challenge. Despite substantial recent interest in identification of nonparametric economic models with endogenous regressors and nonseparable errors, there remain remarkably few such results for fully simultaneous systems. A general representation of a nonparametric simultaneous equations model (more general than we will allow) could be written as

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

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 set of exogenous variables. Assuming m is invertible in

U,[1] 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.

We consider identification in a class of fully simultaneous nonparametric models introduced by Matzkin (2008). These models, which consider the case of continuous endogenous variables Y, 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 (instruments) specific to each equation, and each $g_j(Z, X_j)$ is assumed 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 the more general model (1) by requiring X_j and U_j to enter the full nonparametric nonseparable function m_j through a "residual index" $\delta_j(Z, X_j, U_j)$. If we again assume invertibility of m (now in δ), one obtains the analog of (2),

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

or, equivalently,

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

[1] 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.

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 U is independent of X, $(g_1(X_1), \ldots, g_J(X_J))$ has large support, and the joint density of U satisfies certain restrictions. Matzkin (2010) develops an estimation approach for such models, focusing on the case in which each index function δ_j is linear in X_j (with coefficient normalized to 1), again assuming large support for X and restrictions on the joint density to ensure identification.[2] Matzkin (2010) also provided identification results for related models that could be extended to the model considered here. These results also require a combination of conditions on the support of the "instruments" X and on the joint density of the structural errors.[3]

We provide new identification results for this class model models, generalizing those in Matzkin (2008) and Matzkin (2010). We present three theorems nesting a wide variety of special cases that permit tradeoffs between the support of the instruments X and the restrictions placed on the joint density of the structural errors U. At one extreme, if one maintains the large support condition, identification of the model in Matzkin (2010) (where the residual index function is linear) holds without any restriction on the joint density. Alternatively, one can relax the large support condition by placing restrictions on the joint density that hold for a wide range of known density functions. Among these results are cases in which the support of X can be arbitrarily small. The identification results for nonparametric fully simultaneous models in Matzkin (2008) and Matzkin (2010) are also obtained as special cases of our results. All of our identification results are constructive. We show that our primary sufficient conditions for identification are *verifiable*—i.e., their satisfaction or failure is identified—and that the maintained assumptions defining the model imply falsifiable restrictions.

[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] We provide additional discussion of Matzkin's results below.

1.1 Relation to the Literature

Simultaneous Equations Brown (1983), Roehrig (1988), Brown and Matzkin (1998), and Brown and Wegkamp (2002) 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) showed that this claim is incorrect, leaving uncertain the nonparametric identifiability of fully simultaneous models.

A major breakthrough in this literature came in Matzkin (2008).[4] 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 with $J = 2$ (her "supply and demand" example) of the form (4). The last of these easily generalizes to $J > 2$ and was, to our knowledge, the first result demonstrating identification in a fully simultaneous nonparametric model with nonseparable errors. As already noted, Matzkin's sufficient conditions combine a large support condition with a restriction on the joint density of the structural errors.

More recently, Matzkin (2010), while focused on estimation, has included additional identification results for the case in which the residual index function is linear. Her main identification result follows Matzkin (2008), although the restriction to a linear residual index allows her to partially relax the density restriction used previously. Matzkin (2010) also offers constructive partial identification results for a two-equation model that could be extended to show identification of the full model we consider.[5] These results require joint restrictions on the support of the instruments and shape of the joint density; the latter,

[4] See also Matzkin (2007).

[5] We refer here to the results in her section 4.1. In an Appendix we show that such extensions, at least as we can anticipate them, would also be special cases of our theorems.

however, rule out some natural densities, including the multivariate normal (see Berry and Haile (2013)). Our work builds on that of Matzkin (2008) and Matzkin (2010) by showing that identification can be obtained in these classes of models under more general conditions than previously recognized.

Transformation Models The model (4) considered here can be interpreted as a generalization of the transformation model to a simultaneous system. 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 $g_j\left(Z, X_j\right)$, (4) generalizes (5) by dropping the requirement of a monotonic transformation function and, more fundamental, allowing a vector of outcomes Y to enter the unknown transformation function.[6]

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

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

where U_2 is a scalar error entering m_2 monotonically and X_2 is excluded from the first equation. This structure often arises in a program evaluation setting, where Y_2 might denote a non-random treatment and Y_1 an outcome of interest. To contrast this model with a fully simultaneous system, consider the classical supply and demand system, where Y_1 might be the quantity of the good and Y_2 its price. Supposing that the first equation is the structural

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

demand equation, the second equation would be the reduced form for price (quantity does not appear on the right), 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. 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.

This "problem" of having either multiple endogenous variables or multiple structural errors in each equation is the key challenge in a fully simultaneous system. The triangular framework requires that at least one of the reduced-form equations depend on the structural errors only through a scalar index that enters monotonically. This is an index assumption, although one that is different from that 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 (2014) and Torgovitsky (2010). Blundell and Matzkin (2014) provide a necessary and sufficient condition for the simultaneous model to reduce to the triangular model, pointing out that this condition is quite restrictive.

Completeness We consider a fully specified model of all endogenous variables and seek restrictions on the primitives sufficient to imply identification. An alternative approach is to combine conditional moment conditions with completeness assumptions (Lehmann and Scheffe, 1950, 1955) sufficient for identification. This is a natural approach in many settings, and in Berry and Haile (2014) we have shown how the identification arguments of Newey and Powell (2003) or Chernozhukov and Hansen (2005) can be adapted to an example of the class of models considered here.[7] However, independent of general concerns one might

[7]The identification results for nonparametric regression models are not directly applicable because the structural functions m_j take multiple structural errors as arguments. Nonetheless, the extension is straight-

have with the interpretability of completeness conditions, such conditions may be particularly unsatisfactory in a simultaneous equations setting. A simultaneous equations model already specifies the structure that generates the joint distribution of the endogenous variables, exogenous variables, and structural errors. A high-level assumption like completeness implicitly places further restrictions on the model, although the nature of these restrictions is typically unclear.[8]

1.2 Outline

In what follows, we first step back to provide some motivating examples in section 2. Section 3 then completes the setup of the model. In section 4 we consider identification of the model when the residual index is linear, as in Matzkin (2010). Section 5 provides extensions to the case of a nonlinear index.

2 Examples

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

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

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* $\delta_j \left(Z, X_j, U_j\right)$*,*

forward.

[8] Recent work on this issue includes D'Haultfoeuille (2011) and Andrews (2011).

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 identification of a production function in the presence of unobserved shocks to the marginal product of each input. Output is given by* $Q = F\left(Y, \tilde{U}\right)$, *where* $Y \in \mathbb{R}_+^J$ *is a vector of input quantities and* $\tilde{U} \in \mathbb{R}^J$ *is a vector of latent factor-specific productivity shocks.*[9] *Let* P *and* W *denote the (exogenous) prices of the output and inputs, respectively. The observables are* (Q, P, W, Y). *With this structure, cost-minimizing input demand is determined by a system of first-order conditions*

$$p\frac{\partial F\left(y, \tilde{u}\right)}{\partial y_j} = w_j \quad j = 1, \ldots, J \tag{6}$$

whose solution can be written

$$y_j = \eta_j\left(p, w, \tilde{u}\right) \quad j = 1, \ldots, J.$$

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.,*

$$\frac{\partial F\left(y, \tilde{u}\right)}{\partial y_j} = f_j\left(y\right)\tilde{u}_j$$

for some function f_j. *The first-order conditions (6) then take the form (after taking logs)*

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

which have the form of our model (4) after defining $X_j = \ln\left(\frac{W_j}{P}\right)$, $U_j = \ln\left(\tilde{U}_j\right)$. *The results below will imply identification of the functions* f_j *and, therefore, the realizations of each* U_j.

[9] Alternatively, one can derive the same structure from a model with a Hicks-neutral productivity shock and factor-specific shocks for $J - 1$ of the inputs.

Since Q is observed, this implies identification of the production function F.

Example 3. *Consider an equilibrium model of an imperfectly competitive market for differentiated products. Demand (expressed, e.g., in levels or shares) for each product $j = 1, \ldots, J$ in market t is given by*

$$S_{jt} = \sigma_j \left(P_t, \delta_t^d\right) \tag{7}$$

where $P_t \in \mathbb{R}^J$ is the price vector and δ_t^d is a vector of demand indexes

$$\delta_{jt}^d = g_j\left(X_{jt}\right) + \xi_{jt} \quad j = 1, \ldots, J$$

with $X_{jt} \in \mathbb{R}$ and $\xi_{jt} \in \mathbb{R}$ reflecting, respectively observed and unobserved demand shifters (all other observed demand shifters have been conditioned out, treating them fully flexibly). Prices are determined through oligopoly competition, yielding a reduced form for equilibrium prices with the form

$$P_{jt} = \pi_j \left(\delta_t^d, \kappa_t^s\right) \quad j = 1, \ldots, J \tag{8}$$

where κ_t^s is a vector of cost indexes

$$\kappa_{jt}^s = h_j\left(W_{jt}\right) + \omega_{jt} \quad j = 1, \ldots, J.$$

Here $W_{jt} \in \mathbb{R}$ and $\omega_{jt} \in \mathbb{R}$ denote, respectively, observed and unobserved cost shifters (all other observed cost shifters have been conditioned out). The functions g_j and h_j are all assumed to be strictly increasing. Following Berry, Gandhi, and Haile (2013) and Berry and Haile (2014), one can show that (a) this structure is obtained from many standard models of differentiated products demand and oligopoly supply under appropriate residual index restrictions on preferences and costs and (b) the system can be inverted, yielding a $2J \times 2J$ system of equations

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

where $S_t = (S_{1t}, \ldots, S_{Jt})$, $P_t = (P_{1t}, \ldots, P_{Jt})$. *This system takes the form of (4).*

3 The Model

3.1 Setup

The observables are (Y, X, Z), with $X \in \mathbb{R}^J$, $Y \in \mathbb{R}^J$, and $J \geq 2$. The exogenous observables Z, while important in applications, add no complications to the analysis of identification. Thus, from now on we condition on an arbitrary value of Z and drop it from the notation. As usual, this treats Z in a fully flexible way, although all assumptions should be interpreted to hold conditional on Z. Stacking the equations in (4), we then consider the model

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

where $r(Y) = (r_1(Y), \ldots, r_J(Y))'$ and $g(X)$ denotes $(g_1(X_1), \ldots, g_J(X_J))'$. Here the function r maps $\mathbb{Y} \subseteq \mathbb{R}^J$ to $\mathbb{R}^J$. We let $\mathbb{X}$ =int(supp (X)).

Assumption 1. (i) $\mathbb{X}$ is nonempty;
(ii) g is continuously differentiable, with $\partial g_j(x_j)/\partial x_j > 0$ for all j, x_j;
(iii) U is independent of X and has positive joint density f on $\mathbb{R}^J$;
(iv) r is injective, differentiable, and has nonsingular Jacobian matrix

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

for all $y \in \mathbb{Y}$.

Part (i) rules out purely discrete X but is otherwise mild. Part (ii) combines an important monotonicity restriction with a differentiability requirement imposed for convenience. The primary role of parts (iii) and (iv) is to allow us to attack the identification problem using a standard change of variables approach, relating the joint density of observables to that of the structural errors. However, the following result documents three additional useful implications.

Lemma 1. *Under Assumption 1, (a)* $\forall y \in \mathbb{Y}$*, supp*$(X|Y=y)=$*supp*(X)*; (b)* $\forall x \in \mathbb{X}$*, supp*$(Y|X=x)=$*supp*(Y)*; and (c)* $\mathbb{Y}$ *is open and connected*

Proof. With (9), part (iii) of Assumption 1 immediately implies (a) and (b). Part (iv) of Assumption 1 then implies that r has a continuous inverse $r^{-1}:\mathbb{R}^J \to \mathbb{R}^J$. Connectedness of $\mathbb{Y}$ follows from the fact that the continuous image of a connected set (here $\mathbb{R}^J$) is connected. Since r^{-1} is continuous and injective and $r^{-1}\left(\mathbb{R}^J\right)=\mathbb{Y}$, Brouwer's invariance of domain theorem implies that $\mathbb{Y}$ is open. □

3.2 Normalizations

We impose three standard normalizations.[10] 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 $r_j(Y)$ were replaced by $r_j(Y)+\kappa_j$. Thus, without loss, for an arbitrary point $\dot{y}\in\mathbb{Y}$ and an arbitrary vector $\tau=(\tau_1,\ldots,\tau_J)$ we will set

$$r_j(\dot{y})=\tau_j \quad \forall j. \tag{10}$$

Similarly since, even with (10), (9) would be unchanged if, for every j, $g_j(X_j)$ were replaced by $g_j(X_j)+\kappa_j$ for some constant κ_j while U_j were replaced by $U_j-\kappa_j$, we set

$$g_j(\dot{x}_j)=\dot{x}_j \quad \forall j. \tag{11}$$

[10] We follow Horowitz (2009, pp. 215–216), who makes equivalent normalizations in his semiparametric single-equation version of our model. His exclusion of an intercept is the implicit analog of our location normalization (11). Alternatively we could follow Matzkin (2008), who makes no normalizations in her supply and demand example and shows only that the derivatives of r and g are identified up to scale.

This fixes the location of each U_j, but we must still choose its scale.[11] In particular, since (9) would continue to hold if both sides were multiplied by a nonzero constant, we normalize the scale of each U_j by taking an arbitrary $\dot{x} \in \mathbb{X}$ and setting

$$\frac{\partial g_j(\dot{x}_j)}{\partial x_j} = 1 \quad \forall j. \tag{12}$$

Given (11), we will find a convenient choice of τ (recall (10)) to be

$$\tau_j = \dot{x}_j \quad \forall j$$

so that

$$r_j(\dot{y}) - g_j(\dot{x}_j) = 0 \quad \forall j. \tag{13}$$

3.3 Identifiability, Verifiability, and Falsifiability

As usual, we will say that a structural feature is *identified* if it is uniquely determined by the joint distribution of the observables. The primary structural features of interest here are the functions r, f and, when we allow a nonlinear index function, g. However, we will also be interested in binary features reflecting whether key assumptions hold. We will say that a condition $\mathcal{A}$ is *verifiable* if the value of the indicator $1\{\mathcal{A}\}$ is identified. Thus, when a condition is verifiable, its satisfaction or failure is an identified feature. Our primary sufficient conditions for identification (those beyond the model setup and maintained regularity conditions) are verifiable.

Finally, we will say that a model is *falsifiable* it implies a nontrivial verifiable condition. To be more precise, let $\mathcal{P}^*$ denote the set of all possible probability distributions on the observables (Y, X, Z) which are consistent with the maintained hypotheses that define a model. Let $\mathcal{P}$ denote the set of all possible (regardless of the model) probability distributions

[11] Typically the location and scale of the unobservables can be set arbitrarily without loss. However, there may be applications in which the location or scale of U_j has economic meaning. With this caveat, we follow the longstanding convention in the literature and refer to these restrictions as normalizations.

on (Y,X,Z) (thus $\mathcal{P}^* \subset \mathcal{P}$). For any $P \in \mathcal{P}$, let $1_P\{\mathcal{B}\}$ denote the indicator for satisfaction of a condition $\mathcal{B}$ when the observables have distribution P. We say that the model is falsifiable when there exists a verifiable condition $\mathcal{B}$ such that

$$\begin{aligned} 1_P\{\mathcal{B}\} &= 1 \quad \forall P \in \mathcal{P}^* \\ 1_P\{\mathcal{B}\} &= 0 \quad \exists P \in \mathcal{P}. \end{aligned}$$

Here $\mathcal{B}$ is a verifiable condition which must hold under the maintained assumptions of the model but which does not always hold then these assumptions fail.

Falsifiability is implied by verifiability, but not vice-versa.[12] Often falsifiability of a model is shown by demonstrating existence of overidentifying restrictions on observables, and this will be the case with our results below. A model that is falsifiable is sometimes said to be *testable* or to imply *testable restrictions.* We avoid this terminology because, just as identification does not imply existence of a consistent estimator, falsifiability (or verifiability) does not imply existence of a statistical test with power exceeding its size. We leave all matters of estimation and hypothesis testing for future work.

4 Identification with a Linear Index

We begin with the case in which each g_j is linear in X_j. With the normalizations above,[13] this yields the model

$$r_j(Y) = X_j + U_j \quad j = 1, \ldots, J. \tag{14}$$

By a standard change of variables, the conditional density of $Y = y$ given $X = x$ can then be written

[12]We are not aware of any prior use of the notion of verifiability in the econometrics literature although, as our definition makes clear, this is merely a particular case of identifiability. Our use of the term borrows from philosophy (e.g., Ayer (1936)), where similar distinctions between falsifiability and verifiability have sometimes been made for different purposes.

[13]Linearity of the index function does not alter the need for these normalizations or the extent to which they are without loss.

$$\phi(y|x) = f(r(y) - x)\,|J(y)|\,. \tag{15}$$

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

4.1 Identification without Density Restrictions

Our first result shows that if one maintains Matzkin's (2008, 2010) large support assumption, there is no need for any restriction on the joint density f.[14]

Theorem 1. *Let Assumption 1 hold and suppose $\mathbb{X} = \mathbb{R}^J$. Then r and f are identified.*

Proof. Since

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

(15) implies

$$|J\left(y\right)| = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi\left(y|x\right)\ dx$$

so that (again from (15)) we obtain

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

Thus the value of $f\left(r\left(y\right) - x\right)$ is uniquely determined by the observables for all $x \in \mathbb{R}^J$ and $y \in \mathbb{Y}$. Now let F_j denote the marginal CDF of U_j. Since

$$\int_{\hat{x}_j \geq x_j, \hat{x}_{-j}} f\left(r\left(y\right) - \hat{x}\right)\ d\hat{x} = F_j\left(r_j\left(y\right) - x_j\right) \tag{16}$$

the value of $F_j\left(r_j\left(y\right) - x_j\right)$ is identified for all $x_j \in \mathbb{R}$ and $y \in \mathbb{Y}$. By (11) and (13),

$$F_j\left(r_j\left(\dot{y}\right) - \dot{x}_j\right) = F_{U_j}\left(0\right).$$

[14] The argument used to show Theorem 1 was first used by Berry and Haile (2014) in combination with additional assumptions and arguments to demonstrate identification in models of differentiated products demand and supply.

For every $y \in \mathbb{Y}$ we can then find the value $\overset{o}{x}(y)$ such that $F_j\left(r_j(y) - \overset{o}{x}(y)\right) = F_j(0)$, which reveals $r_j(y) = \overset{o}{x}(y)$. This identifies each function r_j on $\mathbb{Y}$. Identification of f then follows from (14). □

Thus, given the maintained Assumption 1, large support for X is sufficient for identification of the model. Because this sufficient condition involves only the support of the observables X, its satisfaction (or failure) is also identified, giving the following result.

Remark 1. *The condition* $\mathbb{X} = \mathbb{R}^J$ *is verifiable.*

4.2 Identification Combining Support and Density Conditions

We now explore restrictions on f that will lead to several alternative identification results, including some permitting more limited (even arbitrarily small) support for X. We first develop a general result, then illustrate a variety of special cases. To permit an approach relying on differentiation of (15), we will add the following maintained hypotheses.

Assumption 2. (i) r is twice differentiable; (ii) f is continuously differentiable.

Taking logs of (15) and differentiating, we obtain

$$\frac{\partial \ln \phi(y|x)}{\partial x_j} = -\frac{\partial \ln f(r(y) - x)}{\partial u_j} \tag{17}$$

and

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = \sum_j \frac{\partial \ln f(r(y) - x)}{\partial u_j} \frac{\partial r_j(y)}{\partial y_k} + \frac{\partial \ln |J(y)|}{\partial y_k}. \tag{18}$$

Together (17) and (18) imply

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = \frac{\partial \ln |J(y)|}{\partial y_k} - \sum_j \frac{\partial \ln \phi(y|x)}{\partial x_j} \frac{\partial r_j(y)}{\partial y_k} \tag{19}$$

which we rewrite as

$$a_k(x, y) = d(x, y)' \, b_k(y) \tag{20}$$

where

$$
\begin{aligned}
a_k(x,y) &= \frac{\partial \ln \phi(y|x)}{\partial y_k} \\
d(x,y)' &= \left(1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \ldots, -\frac{\partial \ln \phi(y|x)}{\partial x_J}\right) \\
b_k(y) &= \left(\frac{\partial \ln |J(y)|}{\partial y_k}, \frac{\partial r_1(y)}{\partial y_k}, \ldots, \frac{\partial r_J(y)}{\partial y_k}\right)'.
\end{aligned}
$$

Here $a_k(x,y)$ and $d(x,y)$ are observable for all $x \in \mathbb{X}$, $y \in \mathbb{Y}$, whereas $b_k(y)$ involves unknown derivatives of the functions r_j.

The following observation provides a key result of this section.

Lemma 2. *Let Assumptions 1 and 2 hold. For a given $y \in \mathbb{Y}$, suppose there exists no nonzero vector $c = (c_0, c_1, \ldots, c_J)'$ such that*

$$d(x,y)' c = 0 \quad \forall x \in \mathbb{X}.$$

Then $\partial r(y)/\partial y_k$ is identified for all k.

Proof. Suppose the contrary. Then for some k there must exist distinct vectors $b_k^{(1)}(y)$ and $b_k^{(2)}(y)$ solving (20) for all $x \in \mathbb{X}$, i.e.,

$$
\begin{aligned}
a_k(x,y) &= d(x,y)' b_k^{(1)}(y) \quad \forall x \in \mathbb{X} \\
a_k(x,y) &= d(x,y)' b_k^{(2)}(y) \quad \forall x \in \mathbb{X}.
\end{aligned}
$$

Subtracting one equation from the other, we have

$$0 = d(x,y)' c \quad \forall x \in \mathbb{X}$$

for the nonzero vector $c = b_k^{(1)}(y) - b_k^{(2)}(y)$, contradicting the hypothesis of the Lemma. $\square$

Lemma 2 provides a sufficient condition for identification of the derivatives of r at a point y. This condition is equivalent to the requirement that there exist points $\tilde{\mathbf{x}} = (\tilde{x}^0, \ldots, \tilde{x}^J)$

with each $\tilde{x}^j \in \mathbb{X}$ such that the $(J+1) \times (J+1)$ matrix

$$D(\tilde{\mathbf{x}}, y) \equiv \begin{pmatrix} d(\tilde{x}^0, y)' \\ \vdots \\ d(\tilde{x}^J, y)' \end{pmatrix}$$

has full rank. This "rank condition" allows constructive identification of the derivative $\partial r(y)/\partial y_k$ as follows. For arbitrary $k \in \{1, \ldots, J\}$ let

$$A_k(\tilde{\mathbf{x}}, y) = \begin{pmatrix} a_k(\tilde{x}^0, y) \\ \vdots \\ a_k(\tilde{x}^J, y) \end{pmatrix}$$

and stack the equations obtained from (20) at each of the points $\tilde{x}^{(0)}, \ldots, \tilde{x}^{(J)}$, yielding

$$A_k(\tilde{\mathbf{x}}, y) = D(\tilde{\mathbf{x}}, y)\, b_k(y).$$

We then have the closed-form solution

$$b_k(y) = D(\tilde{\mathbf{x}}, y)^{-1} A_k(\tilde{\mathbf{x}}, y) \tag{21}$$

in terms of the observables $D(\tilde{\mathbf{x}}, y)$ and $A_k(\tilde{\mathbf{x}}, y)$.

Using (17), Assumption 3 documents this sufficient condition. Theorem 2 then shows that identification of the model follows easily.

Assumption 3. For almost all $y \in \mathbb{Y}$ there is no $c = (c_0, c_1, \ldots, c_J)' \neq 0$ such that

$$\left(1, \frac{\partial \ln f(r(y) - x)}{\partial x_1}, \ldots, \frac{\partial \ln f(r(y) - x)}{\partial x_J}\right) c = 0 \quad \forall x \in \mathbb{X}.$$

Theorem 2. *Let Assumptions 1, 2, and 3 hold. Then r and f are identified.*

Proof. By Lemma 2 and continuity of the derivatives of r, $\partial r_j(y)/\partial y_k$ is identified for all j,

k, and $y \in \mathbb{Y}$. Since $\mathbb{Y}$ is an open connected subset of $\mathbb{R}^J$, every pair of points in $\mathbb{Y}$ can be joined by a piecewise smooth (continuously differentiable) path in $\mathbb{Y}$.[15] With the boundary condition (10) and Lemma 1 (part (c)), identification of $r_j(y)$ for all y and j then follows from the fundamental theorem of calculus for line integrals. Identification of f then follows from (9). □

4.3 Verifiability and Falsifiability

Assumption 3 involves a joint restriction on the density f and the support of X. Because $\frac{\partial \ln f(r(y)-x)}{\partial u_j} = -\frac{\partial \ln \phi(y|x)}{\partial x_j}$ and $\frac{\partial \ln \phi(y|x)}{\partial x_j}$ is observable, this condition is verifiable.

Remark 2. *Assumption 3 is verifiable.*

Under another verifiable condition—that there exist two sets of points satisfying the rank condition above—the maintained assumptions of the model are falsifiable.

Remark 3. *Suppose that, for some $y \in \mathbb{Y}$, $\mathbb{X}$ contains two sets of points $\tilde{\mathbf{x}} = \left(\tilde{x}^0, \ldots, \tilde{x}^J\right)$ and $\tilde{\tilde{\mathbf{x}}} = \left(\tilde{\tilde{x}}^0, \ldots, \tilde{\tilde{x}}^J\right)$ such that (i) $\tilde{\mathbf{x}} \neq \tilde{\tilde{\mathbf{x}}}$ and (ii) $D(\tilde{\mathbf{x}}, y)$ and $D\left(\tilde{\tilde{\mathbf{x}}}, y\right)$ have full rank. Then the model defined by (14), Assumption 1, and Assumption 2 is falsifiable.*

Proof. By Lemma 2, $\partial r(y)/\partial y_k$ is identified for all k using only $\tilde{\mathbf{x}}$ or only $\tilde{\tilde{\mathbf{x}}}$. Letting $\partial r(y)/\partial y_k [\tilde{\mathbf{x}}]$ and $\partial r(y)/\partial y_k \left[\tilde{\tilde{\mathbf{x}}}\right]$ denote the implied values of $\partial r(y)/\partial y_k$, we obtain the verifiable restrictions $\partial r(y)/\partial y_k [\tilde{\mathbf{x}}] = \partial r(y)/\partial y_k \left[\tilde{\tilde{\mathbf{x}}}\right]$ for all k. □

Finally, we point out that Theorem 2 proves separate identification of the derivatives $\{\partial r(y)/\partial y_k\}_{k=1,\ldots,J}$ at all y and the derivatives $\frac{\partial \ln |J(y)|}{\partial y_k}$ for all k. Since knowledge of the former implies knowledge of the latter, under the assumptions of Theorem 2 we have the overidentifying falsifiable restrictions.

$$\frac{\partial}{\partial y_k} \left| \begin{pmatrix} \frac{\partial r_1(y)}{\partial y_1} & \cdots & \frac{\partial r_1(y)}{\partial y_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_J(y)}{\partial y_1} & \cdots & \frac{\partial r_J(y)}{\partial y_J} \end{pmatrix} \right| = \frac{\partial \ln |J(y)|}{\partial y_k} \quad \forall k \tag{22}$$

[15] See, e.g., Giaquinta and Modica (2007), Theorem 6.63.

This gives the following result.

Remark 4. *Under Assumption 3, the model defined by (14), Assumption 1, and Assumption 2 is falsifiable.*

4.4 Special Cases

Although Assumption 3 provides a sufficient condition for identification, it may be unsatisfying in the much same way that completeness conditions may be unsatisfying in a simultaneous equations framework. In particular, Assumption 3 does not make clear what types of support conditions and/or density restrictions will ensure that this condition holds. Here we now discuss several more explicit combinations of support conditions and/or density restrictions that will ensure satisfaction of Assumption 3, leading to several corollaries to theorem 2. This exploration will also enable us to make precise connections between our results and those in Matzkin (2008, 2010).

4.4.1 Critical Points and Tangencies

We first consider conditions inspired by the proofs in Matzkin (2008, 2010), although we will be able to offer milder sufficient conditions for identification. Begin with the case $J = 2$ and suppose that for almost all $y \in \mathbb{Y}$ there exist points $(x^0(y), x^1(y), x^2(y))$, each in $\mathbb{X}$, with the following properties:

$$\begin{aligned}
\frac{\partial \ln f\left(r(y) - x^0(y)\right)}{\partial u_j} &= 0 \quad \forall j \\
\frac{\partial \ln f\left(r(y) - x^1(y)\right)}{\partial u_1} &\neq 0 = \frac{\partial \ln f\left(r(y) - x^1(y)\right)}{\partial u_2} \\
\frac{\partial \ln f\left(r(y) - x^2(y)\right)}{\partial u_2} &\neq 0.
\end{aligned}$$

As defined here, the point $u^0(y) = r(y) - x^0(y)$ is a critical point of f. At $u^1(y) = r(y) - x^1(y)$, the derivative of f with respect to u_1 is nonzero, but that with respect to u_2 is zero. Thus, $u^1(y)$ has a graphical interpretation as a point of vertical tangency with a level

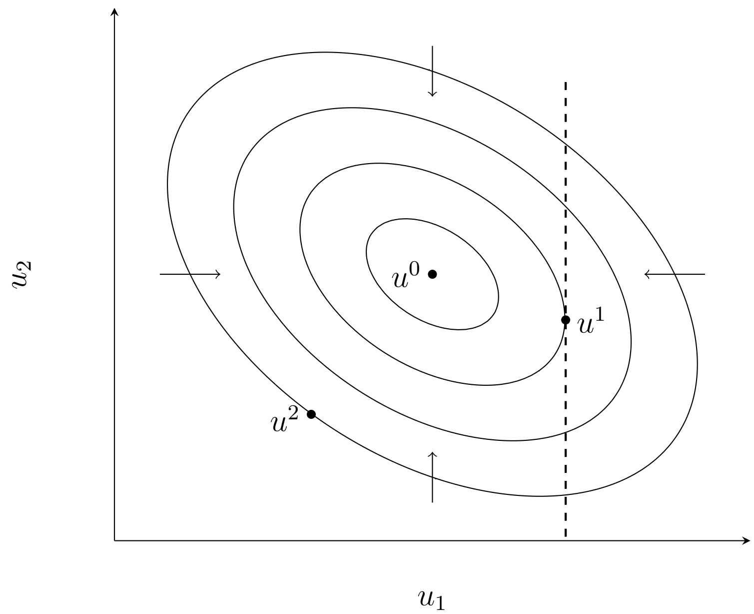


Figure 1: Level sets of a log density. The point u^0 is a local extremum, u^1 is a point of vertical tangency, and u^2 is any point where the log density has non-zero second derivative (i.e., not a point of vertical tangency).

set of f (see Figure 1). Finally $u^2(y) = r(y) - x^2(y)$ is any point such that the derivative of $f(u^2)$ with respect to u_2 is nonzero (i.e., not also a point of vertical tangency).

Letting $\tilde{\mathbf{x}} = (x^0(y), x^1(y), x^2(y))$, we then have

$$D(\tilde{\mathbf{x}}, y) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \frac{\partial \ln f\left(r(y) - x^1(y)\right)}{\partial u_1} & 0 \\ 1 & \frac{\partial \ln f\left(r(y) - x^2(y)\right)}{\partial u_1} & \frac{\partial \ln\left(r(y) - x^2(y)\right)}{\partial u_2} \end{pmatrix}$$

Because this matrix is triangular with nonzero diagonal terms, it is full rank, ensuring that Assumption 3 holds. There is, of course, an analogous construction using a horizontal tangency to a level set of f instead of a vertical tangency. Generalizing to $J \geq 2$ is straightforward and yields the following result.

Corollary 1. *Let Assumptions 1 and 2 hold and suppose that for almost all $y \in \mathbb{Y}$ there exist points $\left(x^0(y), x^1(y), \ldots, x^J(y)\right)$ in $\mathbb{X}$ such that (a) $\partial \ln f(r(y) - x^0(y))/\partial u_j = 0$ for* *all j, (b) $\partial \ln f(r(y) - x^j(y))/\partial u_j \neq 0$ for all j, and (c) the matrix*

$$\begin{pmatrix} \frac{\partial \ln f\left(r(y)-x^1(y)\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(r(y)-x^1(y)\right)}{\partial u_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f\left(r(y)-x^J(y)\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(r(y)-x^J(y)\right)}{\partial u_J} \end{pmatrix}$$

can be placed in triangular form through simultaneous permutation of rows and columns. Then r and f are identified.

In general, existence of the points $\left(x^0(y), x^1(y), \ldots, x^J(y)\right)$ required by Corollary 1 for all y involves a joint requirement on the density f and the support of X. For example, if f has many critical values in different neighborhoods admitting tangencies to its level sets, identification can be obtained from Corollary 1 with limited variation in X—enough for the support of $r(y) - X$ to cover one such neighborhood for almost all $y \in \mathbb{Y}$. Alternatively, with large support for X, Corollary 1 would require existence of only a single set of points $\left(u^0, u^1, \ldots, u^J\right)$ such that $\frac{\partial \ln f\left(u^0\right)}{\partial u_j} = 0$ for all j while

$$\begin{pmatrix} \frac{\partial \ln f\left(u^1\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(u^1\right)}{\partial u_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f\left(u^J\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(u^J\right)}{\partial u_J} \end{pmatrix} \tag{23}$$

is triangular (or, more generally, invertible).[16]

An even stronger restriction on f, but one satisfied by many densities on $\mathbb{R}^J$, is that f have at least one critical point and at least one point of tangency in *each* dimension (recall Figure 1). In that case the matrix in (23) will be diagonal. Combining this property with a large support condition yields the following special case of Corollary 1, which is also the result given by Matzkin (2010, Theorem 3.1).

---

[16] Of course, when X has large support, Theorem 1 provides identification under weaker conditions. However, whereas Theorem 1 uses a type of "identification at infinity" argument (see equation (16)), no such argument is used here. Rather, the large support would ensure only that the set of points at which the derivatives of r are identified includes the entire support of Y.

Corollary 2. *Let Assumptions 1 and 2 hold. Suppose that* $\mathbb{X} = \mathbb{R}^J$ *and that there exist points* $\left(u^0, u^1, \ldots, u^J\right)$*, each in* $\mathbb{R}^J$*, such that (a)* $\partial \ln f(u^0)/\partial u_j = 0$ *for all* j*, (b)* $\partial \ln f(u^j)/\partial u_j \neq 0$ *for all* j*, and (c) the matrix*

$$\begin{pmatrix} \frac{\partial \ln f\left(u^1\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(u^1\right)}{\partial u_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial \ln f\left(u^J\right)}{\partial u_1} & \cdots & \frac{\partial \ln f\left(u^J\right)}{\partial u_J} \end{pmatrix}$$

is diagonal. Then r *and* f *are identified.*

4.4.2 Second Derivatives

The results in the previous section ensured that the matrix

$$\begin{pmatrix} \frac{\partial \ln f\left(r(y)-x^1\right)'}{\partial u} \\ \vdots \\ \frac{\partial \ln f\left(r(y)-x^J\right)'}{\partial u} \end{pmatrix}$$

is invertible for some $\left(x^1, \ldots, x^J\right) \in \mathbb{X}^J$ by requiring points u at which $\frac{\partial \ln f(u)}{\partial u_j}$ is nonzero only for certain j. Obviously this is not necessary for Theorem 2 to apply. Given additional smoothness on f, we can provide a necessary and sufficient condition for the key requirement, Assumption 3, in terms of the second-derivative matrix

$$H_\phi\left(x, y\right) = \begin{pmatrix} \frac{\partial^2 \ln \phi(y|x)}{\partial x_1^2} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_J \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln \phi(y|x)}{\partial x_1 \partial x_J} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_J^2} \end{pmatrix}. \tag{24}$$

Lemma 3. *Let* f *be twice differentiable. For a given* $y \in \mathbb{Y}$*, there is a nonzero vector* $c = (c_0, c_1, \ldots, c_J)'$ *such that*

$$d\left(x, y\right)' c = 0 \quad \forall x \in \mathbb{X} \tag{25}$$

if and only if for the nonzero vector $\tilde{c} = (c_1, \ldots, c_J)'$

$$H_\phi(x, y)\,\tilde{c} = 0 \quad \forall x \in \mathbb{X}. \tag{26}$$

Proof. Recall that $d(x,y)' = \left(1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \ldots, -\frac{\partial \ln \phi(y|x)}{\partial x_J}\right)$. Suppose first that (25) holds for nonzero $c = (c_0, c_1, \ldots, c_J)$. Differentiating (25) with respect to x yields (26), with $\tilde{c} = (c_1, \ldots, c_J)'$. If $c_0 = 0$ then the fact that $c \neq 0$ implies $c_j \neq 0$ for some $j > 0$. If $c_0 \neq 0$, then because the first component of $d(x,y)$ is nonzero and $d(x,y)'\,c = 0$, we must have $c_j \neq 0$ for some $j > 0$. Thus (26) must hold for some nonzero $\tilde{c}$. Now suppose (26) holds for nonzero $\tilde{c} = (c_1, \ldots, c_J)'$. Take an arbitrary point x^0 and let $c_0 = \sum_{j=1}^{J} \frac{\partial \ln \phi\left(y|x^0\right)}{\partial x_j} c_j$ so that, for $c = (c_0, c_1, \ldots, c_J)'$, $d\left(x^0, y\right)' c = 0$ by construction. Since the first component of $d(x,y)$ equals 1 for all (x,y), (26) implies that $\frac{\partial}{\partial x_j}\left[d(x,y)'\;c\right] = 0$ for all j and every $x \in \mathbb{X}$. Thus (25) holds for some nonzero c. $\square$

This equivalence allows us to provide a sufficient condition for identification in terms of the Hessian matrices $\frac{\partial^2 \ln f(r(y)-x)}{\partial u \partial u'}$.

Corollary 3. *Let Assumptions 1 and 2 hold and assume that f is twice differentiable. Suppose that, for almost all $y \in \mathbb{Y}$, there is no nonzero J-vector $\tilde{c}$ such that*

$$\frac{\partial^2 \ln f\left(r(y) - x\right)}{\partial u \partial u'}\,\tilde{c} = 0 \quad \forall x \in \mathbb{X}.$$

Then r and f are identified.

Proof. From (17), $\frac{\partial^2 \ln \phi(y|x)}{\partial x_j \partial x_k} = \frac{\partial^2 \ln f(r(y)-x)}{\partial u_j \partial u_k}$. The result then follows from the definition (24), Lemma 3, and Theorem 2. $\square$

By this result, a sufficient condition for identification is that for each y, $\ln f(u)$ have nonsingular Hessian matrix at a point $u = r(y) - x$ reachable through the support of X. At one extreme, if X has large support it is sufficient that $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ be invertible at a single point. At an opposite extreme, if $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ be nonsingular almost everywhere, the support of X can be arbitrarily small. We state this second special case in the following result.

Corollary 4. *Let Assumptions 1 and 2 hold and assume that f is twice differentiable. If $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ is nonsingular almost everywhere, r and f are identified.*

The requirement that $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ be nonsingular almost everywhere is satisfied by many standard joint probability distributions. For example, it holds when $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ is negative definite almost everywhere—a property of the multivariate normal and many other log-concave densities (see, e.g., Bagnoli and Bergstrom (2005) and Cule, Samworth, and Stewart (2010)). Examples of densities that violate the requirement of Corollary 4 are those with flat (uniform) or log-linear (exponential) regions.

In some applications, it may be reasonable to assume that U_j and U_k are independent for all $k \neq j$. For example, in Example 2, it may be reasonable to assume independence between factor-specific productivity shocks. Under independence, $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ is diagonal. The following result, whose proof is immediate from Corollary 3, shows that this can allow identification with arbitrarily small $\mathbb{X}$ under only a mild additional restriction on f.

Corollary 5. *Let Assumptions 1 and 2 hold. Suppose that $f(u) = \prod_j f_j(u_j)$ for all u and that, for all j, $\frac{\partial^2 \ln f_j(u_j)}{\partial u_j^2}$ exists and is nonzero almost surely. Then r and f are identified.*

We pause to emphasize that although Corollaries 4–5 have provided sufficient conditions involving nonsingularity of $\frac{\partial^2 \ln f(u)}{\partial u \partial u'}$ at some points u, this is not necessary for one to obtain identification through Corollary 3. For a given pair (x, y), $\frac{\partial^2 \ln f(r(y)-x)}{\partial u \partial u'}$ is singular if and only if there exists a nonzero vector c such that $\frac{\partial^2 \ln f(r(y)-x)}{\partial u \partial u'} c = 0$. However, Corollary 3 shows that only when (for values of y with positive measure) the *same* vector c solves this equation *for every* $x \in \mathbb{X}$ does the sufficient condition for identification in Theorem 2 fail. When this happens, the columns of $\frac{\partial^2 \ln f(r(y)-x)}{\partial u \partial u'}$ do not merely exhibit linear dependence at each x: they exhibit the *same* linear dependence for *all* x. Thus *failure* identification through Corollary 3 requires strong restrictions on the joint density f.

4.4.3 Differenced Derivatives

Although we exploited the assumed twice-differentiability of f in the previous section, it is straightforward to extend our arguments to cases without this additional differentiability by replacing the matrix of second derivatives with differences of the first derivatives. To see this, suppose that (25) holds for some nonzero c. This implies that $d\,(y,x)'\,c$ is constant across all $x \in \mathbb{X}$; i.e., for any x and x' in $\mathbb{X}$,

$$[d\,(y,x) - d\,(y,x')]'\;c = 0.$$

Since the first component of $d\,(y,x) - d\,(y,x')$ is zero, this is equivalent to the condition

$$\begin{bmatrix} \frac{\partial \ln \phi(y|x)}{\partial x_1} - \frac{\partial \ln \phi(y|x')}{\partial x_1} \\ \vdots \\ \frac{\partial \ln \phi(y|x)}{\partial x_J} - \frac{\partial \ln \phi(y|x')}{\partial x_J} \end{bmatrix}' \tilde{c} = 0 \qquad \forall x \in \mathbb{X},\ x' \in \mathbb{X}. \tag{27}$$

Thus, only when there exists a nonzero vector $\tilde{c}$ satisfying (27) does the rank condition of Theorem 2 fail.

5 Identification with a Nonlinear Index

We now move to the more general model (9), where the change of variables (cf. equation (15)) now yields

$$\phi(y|x) = f(r(y) - g(x))\,|J\,(y)|$$

for all $y \in \mathbb{Y}$, $x \in \mathbb{X}$. Thus

$$\frac{\partial \ln \phi(y|x)}{\partial x_j} = -\frac{\partial \ln f(r(y) - g(x))}{\partial u_j}\frac{\partial g_j(x_j)}{\partial x_j} \tag{28}$$

and

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = \sum_j \frac{\partial \ln f(r(y) - g(x))}{\partial u_j} \frac{\partial r_j(y)}{\partial y_k} + \frac{\partial \ln |J(y)|}{\partial y_k} \tag{29}$$

which together imply

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = -\sum_j \frac{\partial \ln \phi(y|x)}{\partial x_j} \frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j)/\partial x_j} + \frac{\partial \ln |J(y)|}{\partial y_k}. \tag{30}$$

Relative to the linear index model considered in section 4, the complication introduced by the more general model (9) is the need to identify the index functions g_j. Our strategy for doing so will build on an important insight in Matzkin (2008) that again relies on critical points of f and tangencies to its level sets. We first develop a general sufficient condition involving joint restrictions on the density f and support of X. Although this requires some new concepts, we ultimately provide some simpler special cases. One of these, which corresponds to the conditions of Matzkin's (2008) result, combines large support for X and global restrictions on f. Another allows arbitrarily small support for X and essentially any density f with an isolated critical value that is a local min or max. Under any of these conditions, the results of section 4 extend immediately to the more general model.

5.1 Rectangle Regularity

We begin with some definitions.

Definition 1. A *J-dimensional rectangle* is a Cartesian product of J nonempty open intervals.

Definition 2. Let $\mathcal{M} \equiv \times_{j=1}^{J} \left(\underline{m}_j, \overline{m}_j\right)$ and $\mathcal{N} \equiv \times_{j=1}^{J} \left(\underline{n}_j, \overline{n}_j\right)$ denote two J-dimensional rectangles. $\mathcal{M}$ is *smaller than* $\mathcal{N}$ if $\overline{m}_j - \underline{m}_j \leq \overline{n}_j - \underline{n}_j$ for all j.

Definition 3. Given a J-dimensional rectangle $\mathcal{U} \equiv \times_{j=1}^{J} \left(\underline{u}_j, \overline{u}_j\right)$, the joint density f is *regular on $\mathcal{U}$* if (i) there exists $u^* \in \mathcal{U}$ such that $\frac{\partial f(u^*)}{\partial u_j} = 0$ for all j; and (ii) for all j and almost all $u'_j \in \left(\underline{u}_j, \overline{u}_j\right)$, there exists $\hat{u}\left(u'_j\right) \in \mathcal{U}$ satisfying

$$\hat{u}_j\left(u'_j\right) = u'_j$$

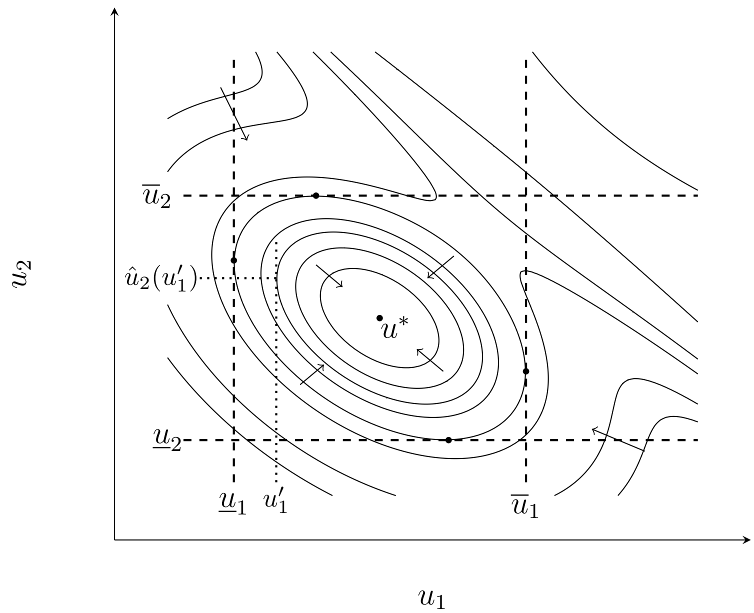
$$\frac{\partial f(\hat{u}(u'_j))}{\partial u_j} \neq 0 \quad \text{and}$$

$$\frac{\partial f(\hat{u}(u'_j))}{\partial u_k} = 0 \quad \forall k \neq j.$$

Definitions 1 and 2 are standard and provided here only to avoid ambiguity. Definition 3 introduces a particular notion of regularity for the density f. It requires that f have a critical value u^* in a rectangular neighborhood $\mathcal{U}$ in which the level sets of f are "nice" in a sense defined by part (ii). There, $\hat{u}(u'_j)$ has a geometric interpretation as a point of tangency between a level set of f and the $(J-1)$-dimensional plane $\left\{u \in \mathbb{R}^J : u_j = u'_j\right\}$.

Figure 2 provides an example in which $J = 2$ and u^* is a local extremum. There, within a neighborhood of u^* the level sets of f (or $\ln f$) are connected, smooth, and strictly increasing toward u^*. Therefore, each level set is horizontal at (at least) one point above u^* and one point below u^*. Similarly, each level set is vertical at least once each to the right and to the left of u^*. There are many J-dimensional rectangles on which the illustrated density is regular. One such rectangle is defined in the figure using a single level set. The upper limit $\overline{u}_2$ is defined by the largest horizontal point on this level set, while $\overline{u}_1$ is defined by the rightmost vertical point, and so forth. For each $u'_1 \in (\underline{u}_1, \overline{u}_1)$, the point $\hat{u}_2(u'_1)$ is the value of U_2 at a tangency between the vertical line $U_1 = u'_1$ and a level set of f closer to u^* than that defining $\mathcal{U}$. Since level sets within $\mathcal{U}$ are smooth, the tangency cannot be at a corner of the rectangle $\mathcal{U}$; therefore, $\underline{u}_2 < \hat{u}_2(u'_1) < \overline{u}_2$, implying $(u'_1, \hat{u}_2(u'_1)) \in \mathcal{U}$. Regularity on $\mathcal{U}$ requires that there exist a vertical tangency within $\mathcal{U}$ for (almost) all $u'_1 \in (\underline{u}_1, \overline{u}_1)$, as well as a horizontal tangency within $\mathcal{U}$ for (almost) all $u'_2 \in (\underline{u}_2, \overline{u}_2)$. The generalization to $J > 2$ is straightforward.

Figure 2: The solid curves are the level sets of a bivariate density (or log-density) with a "regular" hill leading up to a local maximum at u^*, but with a less useful shape in other areas. For each $u_1' \in (\underline{u}_1, \overline{u}_1)$ the point $\hat{u}_2(u_1')$ is the value of U_2 at a tangency between the vertical line $U_1 = u_1'$ and a level set.



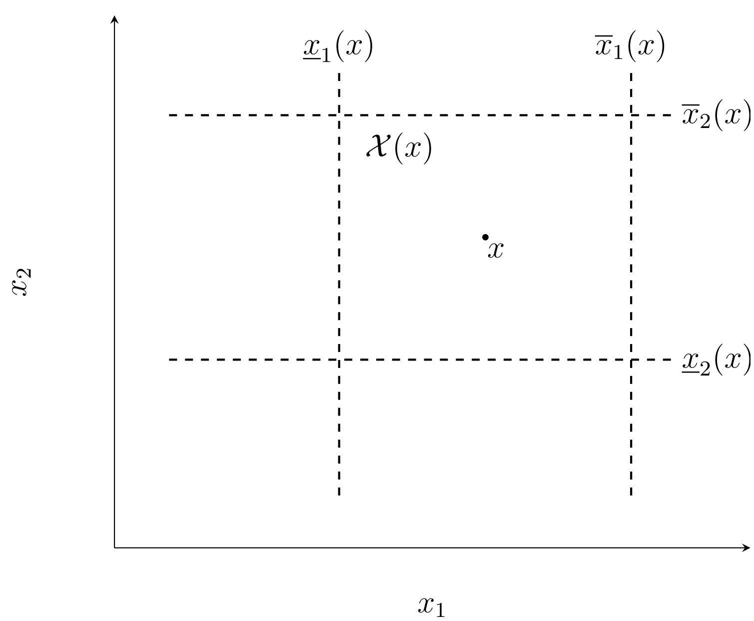


Figure 3: For arbitrary $x \in \mathcal{X}$, the rectangle $\mathcal{U} \equiv (\underline{u}_1, \overline{u}_1) \times (\underline{u}_2, \overline{u}_2)$ in Figure 2 is mapped to a rectangle $\mathcal{X}(x)$ by first defining y^* to satisfy $r_j(y^*) = g_j(x_j) + u_j^*$ for all j, then defining $\underline{x}(x)$ and $\overline{x}(x)$ by $r_j(y^*) = g_j\left(\underline{x}_j(x)\right) + \overline{u}_j = g_j(\overline{x}_j(x)) + \underline{u}_j$, thereby satisfying (31).

The following is our primary condition allowing identification of the index functions g_j.

Assumption 4. For all $x \in \mathbb{X}$ there is a J-dimensional rectangle $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x)\right) \subseteq \mathbb{X}$ containing x such that for (i) some u^* such that $\partial f(u^*)/\partial u_j = 0$ for all j and (ii) $\underline{u}_j(x)$ and $\overline{u}_j(x)$ defined by

$$\begin{aligned} \underline{u}_j(x) &= \quad u_j^* + g_j(x_j) - g_j(\overline{x}_j(x)) \\ \overline{u}_j(x) &= \quad u_j^* + g_j(x_j) - g_j\left(\underline{x}_j(x)\right) \end{aligned} \tag{31}$$

f is regular on $\mathcal{U}(x) = \times_j \left(\underline{u}_j(x), \overline{u}_j(x)\right)$.

Assumption 4 requires, for any given x, that f be regular on a rectangular neighborhood around a critical point u^* that maps through (9) to a rectangular neighborhood in $\mathbb{X}$ around x.[17] Because $\mathbb{X}$ is open, there exists a rectangle in $\mathbb{X}$ around every point $x \in \mathbb{X}$. Further, when $\mathbb{X}$ includes any rectangle $\mathcal{M}$, it also includes all smaller rectangles $\mathcal{X} \subset \mathcal{M}$. Thus, since $g(\mathcal{X})$ is a rectangle whenever $\mathcal{X}$ is, as long as f is regular on some rectangle that is not too big relative to the support of X around x, the set $\mathcal{X}(x)$ required by Assumption 4 is guaranteed to exist. Figure 3 illustrates, taking an arbitrary point x and the rectangle $\mathcal{U} = (\underline{u}_1, \overline{u}_1) \times (\underline{u}_2, \overline{u}_2)$ in Figure 2 and mapping them to the rectangle $\mathcal{X}(x)$.

Observe that although we write $\underline{u}_j(x)$ and $\overline{u}_j(x)$ in (31), the same rectangle $\times_j \left(\underline{u}_j, \overline{u}_j\right)$ may be used to construct $\mathcal{X}(x)$ for many (even all) values of x. This is because for every $x \in \mathbb{X}$ there must exist $y^*(x) \in \mathbb{Y}$ such that

$$r(y^*(x)) = g(x) + u^* \tag{32}$$

allowing construction of the required rectangle $\times_j \left(\underline{x}_j(x), \overline{x}_j(x)\right)$ from (31) with a single critical value u^* and with $\underline{u}_j(x) = \underline{u}_j$ and $\overline{u}_j(x) = \overline{u}_j$ for all x and j. Thus, Assumption 4 can be satisfied with only local variation in X.

[17] Starting from a rectangle $\times_j \left(\underline{u}_j, \overline{u}_j\right)$ on which f is regular, first define y^* by the equation $r(y^*) = g(x) + u^*$. Then define each $\underline{x}_j(x)$ and $\overline{x}_j(x)$ by $r(y^*) = g\left(\underline{x}_j(x)\right) + \overline{u}_j$. and $r(y^*) = g(\overline{x}_j(x)) + \underline{u}_j$.

5.1.1 Sufficient Conditions for Assumption 4

Here we provide two alternative sufficient conditions for Assumption 4. The first combines a large support condition on X with regularity of f on $\mathbb{R}^J$. This was the combination of conditions proposed by Matzkin (2008).[18] The second allows arbitrarily limited support for X and requires regularity only in sufficiently small rectangular neighborhoods around a critical point u^*. Outside such neighborhoods, f is unrestricted.

Remark 5. *Suppose that $g(\mathbb{X}) = \mathbb{R}^J$ and that f is regular on $\mathbb{R}^J$. Then Assumption 4 holds.*

Proof. Let $\mathcal{X}(x) = \times_j \left(g_j^{-1}(-\infty), g_j^{-1}(\infty)\right)$ for all x. Then by (31), $\mathcal{U}(x) = \mathbb{R}^J$, yielding the result. □

Definition 4. f satisfies *local rectangle regularity* if for every J-dimensional rectangle $\mathcal{M}$, there exists a J-dimensional rectangle $\mathcal{N}$ smaller than $\mathcal{M}$ such that f is regular on $\mathcal{N}$.

Remark 6. *Suppose that f satisfies local rectangle regularity. Then Assumption 4 holds.*

Proof. Take arbitrary $x \in \mathbb{X}$. Because $\mathbb{X}$ is open, it must contain a rectangle $\tilde{\mathcal{X}}(x) \ni x$. Let $\mathcal{M} = g\left(\tilde{\mathcal{X}}(x)\right)$ and let $\mathcal{N}$ be a smaller rectangle $\times_j \left(\underline{u}_j, \overline{u}_j\right)$ on which f is regular (guaranteed to exist by local rectangle regularity). Because f is regular on $\mathcal{N}$, it has a critical point $u^* \in \mathcal{N}$. Taking such a point u^*, define $y^*(x)$ by (32). Now define $\underline{x}_j(x)$ and $\overline{x}_j(x)$ for all j by

$$r_j(y^*(x)) = g_j\left(\underline{x}_j(x)\right) + \overline{u}_j \tag{33}$$

$$r_j(y^*(x)) = g_j\left(\overline{x}_j(x)\right) + \underline{u}_j. \tag{34}$$

Let $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x)\right)$. Then by (31), (32), (33), and (34), $\mathcal{U}(x) = \mathcal{N}$. Thus f is regular on $\mathcal{U}(x)$. □

[18] The regularity assumption stated in Matzkin (2008) is actually stronger, equivalent to assuming regularity of f on $\mathbb{R}^J$ but replacing "almost all $u'_j \in \left(\underline{u}_j, \overline{u}_j\right)$" in the definition of regularity with "all $u'_j \in \left(\underline{u}_j, \overline{u}_j\right)$." The latter is unnecessarily strong and rules out many standard densities, including the multivariate normal. Throughout we interpret the weaker condition as that intended by Matzkin (2008).

Neither of these sufficient conditions implies the other. If X has large support, regularity on $\mathbb{R}^J$ holds when f is one of many standard densities, including the multivariate normal. Local rectangle regularity places no requirement on the support of X and (given our maintained assumption $f \in \mathcal{C}^1$) and holds for essentially any joint density f with an isolated critical value that is a local minimum or maximum. To make this precise (taking the case of a local max), for $c \in \mathbb{R}$ and $\Sigma \subset \mathbb{R}^J$ define the upper contour sets of the restriction of f to Σ:

$$\mathcal{A}(c;\Sigma) = \{u \in \Sigma : f(u) \geq c\}.$$

Remark 7. *Let f be continuously differentiable and suppose there is a compact connected set $S \subset \mathbb{R}^J$ with nonempty interior such that for some $\underline{c} \in \mathbb{R}$, (i) $\mathcal{A}(\underline{c};S) \subset int(S)$, and (ii) the restriction of f to $\mathcal{A}(\underline{c};S)$ attains a maximum $\bar{c} > \underline{c}$ at its unique critical value u^*. Then f satisfies local rectangle regularity.*

Proof. See Appendix A. □

Figure 4 illustrates the construction used in the proof of Remark 7.

5.2 Identification of the Index Functions

Under Assumption 4, identification of the index functions g_j follows in three steps. The first exploits a critical value u^* to pin down derivatives of the Jacobian determinant at a point $y^*(x)$ for any x. The second uses tangencies to identify the ratios $\frac{\partial g_j(x_j')/\partial x_j}{\partial g_j(x_j^0)/\partial x_j}$ for all pairs of points (x^0, x') in a sequence of overlapping rectangular subsets of $\mathbb{X}$. The final step links these rectangular neighborhoods so that, using the normalization (12), we can integrate up to the functions g_j, using (11) as boundary conditions.

The first step is straightforward. If u^* is a critical value of f and, for arbitrary $x \in \mathbb{X}$, $y^*(x)$ is defined by (32), equation (29) yields

$$\frac{\partial \ln |J(y^*(x))|}{\partial y_k} = \frac{\partial \ln \phi(y^*(x)\,|x)}{\partial y_k} \quad \forall k. \tag{35}$$

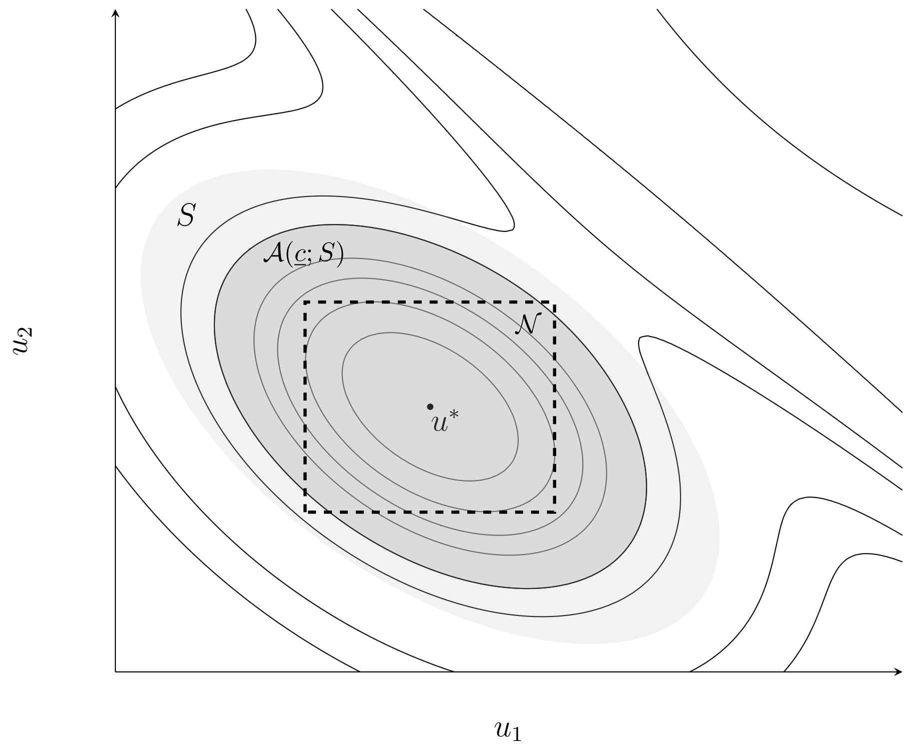


Figure 4: The shaded area is a connected compact set S. The darker subset of S is an upper contour set $\mathcal{A}(\underline{c}; S)$ of the restriction of f to S. The point u^* is a local max and the only critical value of f in $\mathcal{A}(\underline{c}; S)$. For any J-dimensional rectangle $\mathcal{M}$ there will exist $c^0 \geq \underline{c}$ such that (i) the rectangle $\mathcal{N} = \times_j \left(\min_{u^- \in \mathcal{A}(c^0;S)} u_j^-, \max_{u^+ \in \mathcal{A}(c^0;S)} u_j^+ \right)$ (see (A.2) in Appendix A) is smaller than $\mathcal{M}$ and (ii) f is regular on $\mathcal{N}$.

For arbitrary x and x', this allows us to rewrite (30) as

$$\sum_j \frac{\partial \ln \phi(y^*(x)\,|x')}{\partial x_j}\frac{\partial r_j((y^*(x))/\partial y_k}{\partial g_j(x'_j)/\partial x_j} = \frac{\partial \ln \phi(y^*(x)\,|x)}{\partial y_k} - \frac{\partial \ln \phi((y^*(x)\,|x')}{\partial y_k} \tag{36}$$

where the only unknowns are the ratios $\frac{\partial r_j(y^*(x))/\partial y_k}{\partial g_j(x'_j)/\partial x_j}$.

The second step is demonstrated in Lemma 4 below. This step exploits the fact that under Assumption 4, as $\hat{x}$ varies around the arbitrary point x, $r(y^*(x)) - \hat{x}$ takes on all values in a rectangular neighborhood of u^* on which f is regular.

Lemma 4. *Let Assumptions 1, 2, and 4 hold. Then for every $x \in \mathbb{X}$ there exists a J-dimensional rectangle $\mathcal{X}(x) \ni x$ such that for all $x^0 \in \mathcal{X}(x)\backslash x$ and $x' \in \mathcal{X}(x)\backslash x$, the ratio*

$$\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$$

is identified for all $j = 1, \ldots, J$.

Proof. Take arbitrary $x \in \mathbb{X}$ and let u^* and $\mathcal{U}(x) = \times_j \left(\underline{u}_j, \overline{u}_j\right)$ be as defined in Assumption 4.[19] Define y^* by (32). By Assumption 4 there exists $\mathcal{X} = \times_i (\underline{x}_i, \overline{x}_i) \subseteq \mathbb{X}$ (with $x \in \mathcal{X}$) such that

$$r_j(y^*) = g_j(\overline{x}_j) + \underline{u}_j, \quad j = 1, \ldots, J \tag{37}$$

$$r_j(y^*) = g_j(\underline{x}_j) + \overline{u}_j, \quad j = 1, \ldots, J \tag{38}$$

and (recalling (28)) such that for each j and almost every $x'_j \in \left(\underline{x}_j, \overline{x}_j\right)$ there is a J-vector $\hat{x}\left(x'_j\right) \in \mathcal{X}$ satisfying

$$\hat{x}_j\left(x'_j\right) = x'_j$$

$$\frac{\partial \ln \phi\left(y^*|\hat{x}(x'_j)\right)}{\partial x_j} \neq 0 \quad \text{and} \tag{39}$$

$$\frac{\partial \ln \phi\left(y^*|\hat{x}(x'_j)\right)}{\partial x_k} = 0 \quad \forall k \neq j. \tag{40}$$

[19] To simplify notation, we will suppress dependence of $y^*, \underline{x}_j, \overline{x}_j, \underline{u}_j$, and $\overline{u}_j$ on the arbitrary point x.

Since $\phi(y|x)$ and its derivatives are observed for all $y \in \mathbb{Y}, x \in \mathbb{X}$, the point y^* is identified, as are the pairs $(\underline{x}_j, \overline{x}_j)$ and the point $\hat{x}(x_j')$ for any j and $x_j' \in (\underline{x}_j, \overline{x}_j)$.[20] Taking arbitrary j, arbitrary $x_j' \in (\underline{x}_j, \overline{x}_j)$ and the known vector $\hat{x}(x_j')$ defined above, (36) becomes

$$\frac{\partial \ln \phi(y^*|\hat{x}(x_j'))}{\partial x_j} \frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j')/\partial x_j} = \frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(x_j'))}{\partial y_k}.$$

By (39), we may rewrite this as

$$\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j')/\partial x_j} = \frac{\frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(x_j'))}{\partial y_k}}{\frac{\partial \ln \phi(y^*|\hat{x}(x_j'))}{\partial x_j}}. \tag{41}$$

Since the right-hand side is known, $\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j')/\partial x_j}$ is identified for almost all $x_j' \in (\underline{x}_j, \overline{x}_j)$. By continuity, this yields identification of $\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j')/\partial x_j}$ for all $x_j' \in (\underline{x}_j, \overline{x}_j)$. By the same arguments leading up to (41), but with x_j^0 taking the role of x_j', we obtain

$$\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j^0)/\partial x_j} = \frac{\frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(x_j^0))}{\partial y_k}}{\frac{\partial \ln \phi(y^*|\hat{x}(x_j^0))}{\partial x_j}} \tag{42}$$

yielding identification of $\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j^0)/\partial x_j}$ for all $x_j^0 \in (\underline{x}_j, \overline{x}_j)$. Because the Jacobian determinant $|J(y^*)|$ is nonzero, $\partial r_j(y^*)/\partial y_k$ cannot be zero for all k. Thus for each $j = 1, \ldots, J$ there must exist some k such that the ratio $\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j^0)/\partial x_j} \bigg/ \frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x_j')/\partial x_j}$ is well defined. This establishes the result.[21] $\square$

The final step of the argument will yield the following result.

Theorem 3. *Let Assumptions 1, 2, and 4 hold and suppose that* $\mathbb{X}$ *is connected. Then* g *is identified on* $\mathbb{X}$.

[20] We do not require uniqueness of u^* or the set $\mathcal{U}$. Rather, we use only the fact that for a given x there exist both a value y^* mapping through (14) to a critical point u^* and a rectangle around x mapping through (37) and (38) to a rectangle around u^* on which f is regular. Through (28), such a y^* and a rectangle around x are both identified.

[21] Since the argument can be repeated for any k such that $\partial r_j(y^*)/\partial y_k \neq 0$, the ratios of interest in the lemma may typically be overidentified.

Proof. We first claim that Lemma 4 implies identification of $\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$ for all $x'_j \in \mathbb{X}$ and an arbitrary $x^0_j \in \mathbb{X}$. If $\mathcal{X}(x) = \mathbb{X}$ in Lemma 4, this is immediate. Otherwise, consider any two points x^0 and x' in $\mathbb{X}$. Because $\mathbb{X}$ is a connected open subset of $\mathbb{R}^J$, $\mathbb{X}$ is path-connected. Thus, there exists a continuous function $\rho : [0,1] \to \mathbb{X}$ such that $\rho(0) = x^0$ and $\rho(1) = x'$. Because each rectangle $\mathcal{X}(\rho(\lambda))$ guaranteed to exist by Lemma 4 is open, we know that for sufficiently small ϵ

$$\mathcal{X}\left(x^0\right) \ni \rho(\epsilon), \quad \mathcal{X}(x') \ni \rho(1-\epsilon)$$

while for every $\lambda \in (0,1)$ there exists $\delta(\lambda) > 0$ such that

$$\mathcal{X}(\rho(\lambda)) \ni \rho(\lambda - \delta(\lambda)), \quad \mathcal{X}(\rho(\lambda)) \ni \rho(\lambda + \delta(\lambda)).$$

Thus, there exists an ordered set of sequentially overlapping rectangles

$$\{\mathcal{X}(\rho(\lambda))\}_{\lambda \in [0,1]}$$

linking x^0 to x'. Since the ratios $\frac{\partial g_j(x^1_j)/\partial x_j}{\partial g_j(x^2_j)/\partial x_j}$ known for all points $\left(x^1_j, x^2_j\right)$ in each of these rectangles, it follows that the ratios $\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x_j)/\partial x_j}$ are known for all x^0 and x' in $\mathbb{X}$. The claim then follows. Finally, taking $x^0_j = \dot{x}_j$ for all j, the conclusion of the Theorem follows from the normalization (12) and boundary condition (11). □

5.3 Identification of the Model

Theorem 3 is important because once g is known, identification results for the linear index model extend immediately to the setting with a nonlinear index. This follows by letting

$$\tilde{X}_j = g_j(X) \quad \forall j$$

so that (9) becomes

$$r(Y) = \tilde{X} + U.$$

This allows many possible combinations of sufficient conditions for identification of (r, f, g). We will give three examples. We begin with the identification result for nonparametric fully simultaneous models given (with different proof) in Matzkin (2008).

Corollary 6. *Let Assumptions 1 and 2 hold. Suppose that $g(\mathbb{X}) = \mathbb{R}^J$ and that f is regular on $\mathbb{R}^J$. Then g is identified on $\mathbb{X}$, and r and f are identified.*

Proof. Since $g(\mathbb{X}) = \mathbb{R}^J$ and each g_j has everywhere strictly positive derivative, g has a continuous inverse g^{-1}on $\mathbb{R}^J$. Since the image of a path-connected set under a continuous mapping is path-connected, $\mathbb{X} = g^{-1}\left(\mathbb{R}^J\right)$ is path-connected. Thus by Lemma 5 and Theorem 3, g is identified on $\mathbb{X}$. Since regularity of f on $\mathbb{R}^J$ implies the conditions of Corollary 2, identification of r and f then follows. □

The next result shows that if the large support assumption of Corollary 6 is retained, it also suffices that f satisfy local rectangle regularity.

Corollary 7. *Let Assumptions 1 and 2 hold. Suppose that $g(\mathbb{X}) = \mathbb{R}^J$ and that f satisfies local rectangle regularity. Then g is identified on $\mathbb{X}$, and r and f are identified.*

Proof. Identification of g on $\mathbb{X}$ follows as in the proof of Corollary 6. The result then follows from Theorem 1. □

Finally, we provide one of the many possible results demonstrating identification with only limited support for X.

Corollary 8. *Let Assumptions 1 and 2 hold. Suppose that $\mathbb{X}$ is connected, f satisfies local rectangle regularity, and that $\partial^2 \ln f(u) / \partial u \partial u'$ is nonsingular almost everywhere. Then g is identified on $\mathbb{X}$, and r and f are identified.*

Proof. By Lemma 6 and Theorem 3, g is identified on $\mathbb{X}$. The result then follows from Corollary 4. □

5.4 Verifiability and Falsifiability

5.4.1 Verifiability

Although Assumption 4 involves a condition on the joint distribution of latent variables, it is equivalent to a verifiable condition on observables.

Remark 8. *Assumption 4 is verifiable.*

Proof. Fix an arbitrary $x \in \mathbb{X}$. By (28),

$$\frac{\partial \phi(y^*(x)\,|x)}{\partial x_j} = 0 \tag{43}$$

if and only if, for $u^* = r(y^*(x)) - g(x)$, $\frac{\partial f(u^*)}{\partial u_j} = 0\ \forall j$. Thus part (a) of Assumption 4 holds iff there exists $y^*(x) \in \mathbb{Y}$ such that (43) holds. Thus, part (a) is verifiable. Now observe that for $\mathcal{X}(x)$ and $\mathcal{U}(x)$ as defined in part (b) of Assumption 4,

$$x \in \mathcal{X}(x) \iff (r(y^*(x)) - g(x)) \in \mathcal{U}(x).$$

Thus, part (b) holds if and only if there exists a rectangle $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x)\right) \subset \mathbb{X}$, with $x \in \mathcal{X}(x)$ such that for all j and almost all $x_j \in \left(\underline{x}_j(x), \overline{x}_j(x)\right)$ there exists $\hat{x}(x_j) \in \mathcal{X}(x)$ satisfying

$$\begin{aligned}
\hat{x}_j(x_j) &= x_j \\
\frac{\partial \phi(y^*(x)\,|\hat{x}(x_j))}{\partial x_j} &\neq 0 \\
\frac{\partial \phi(y^*(x)\,|\hat{x}(x_j))}{\partial x_k} &\neq 0 \quad \forall k \neq j.
\end{aligned}$$

Satisfaction of this condition is determined by the observables, implying that part (b) is verifiable. □

5.4.2 Falsifiability

We proved identification of the model by showing that, once g is known, all is as if we are back in the case of a linear residual index. By the same logic, given identification of g, the falsifiable restrictions derived in Remark 3 above imply falsifiable restrictions of the residual index model in the case of the nonlinear index as well. We omit a formal statement of this falsifiability result. However, we can also derive a new set of overidentifying restrictions.

Remark 9. *Under Assumptions 1, 2, and 4, the model defined by (9) is falsifiable.*

Proof. The proof of Lemma 4 began with an arbitrary $x \in \mathbb{X}$ and the associated $y^*(x)$ defined by (32). It was then demonstrated that for some open rectangle $\mathcal{X}(x) \ni x$ the ratios

$$\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$$

are identified for all $j = 1, \ldots, J$, all $x^0 \in \mathcal{X}(x) \backslash x$ and all $x' \in \mathcal{X}(x) \backslash x$. Let

$$\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}[x]$$

denote the identified value of $\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$. Now take any point $\tilde{x} \in \mathcal{X}(x) \backslash x$ and repeat the argument, replacing $y^*(x)$ with the point $y^{**}(\tilde{x})$ such that $r(y^{**}(\tilde{x})) = g(\tilde{x}) + u^{**}$ where $\partial f(u^{**})/\partial u_j = 0\ \forall j$ and f is regular on a rectangle around u^{**} (u^{**} may equal u^*, but this is not required). For some open rectangle $\mathcal{X}(\tilde{x})$, this again leads to identification of the ratios

$$\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$$

for all $j = 1, \ldots, J$, all $x^0 \in \mathcal{X}(\tilde{x}) \backslash \tilde{x}$ and all $x' \in \mathcal{X}(\tilde{x}) \backslash \tilde{x}$. Let

$$\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}[\tilde{x}]$$

denote the identified value of $\frac{\partial g_j(x'_j)/\partial x_j}{\partial g_j(x^0_j)/\partial x_j}$. Because both x^* and $\tilde{x}$ are in the open set $\mathcal{X}(x)$,

$\{\mathcal{X}(x) \cap \mathcal{X}(\tilde{x})\} \neq \emptyset$. Thus we obtain the verifiable restriction

$$\frac{\partial g_j(x_j')/\partial x_j}{\partial g_j(x_j^0)/\partial x_j}\left[x\right] = \frac{\partial g_j(x_j')/\partial x_j}{\partial g_j(x_j^0)/\partial x_j}\left[\tilde{x}\right]$$

for all j and all pairs $(x^0, x') \in \{\mathcal{X}(x) \cap \mathcal{X}(\tilde{x})\}$. □

6 Conclusion

We have considered identification in a class of nonparametric simultaneous equations models with a residual index structure—models first introduced in important work by Matzkin (2008, 2010). Building on Matzkin'z work, we have shown that this class of models permits identification under considerably more general conditions than previously recognized. Our results admit a broad range of special cases that offer tradeoffs between the size of the support of the instruments and the restrictiveness of shape assumptions on the density of the unobservables. Further, our conditions on the support of instruments or the shape of the density are verifiable, while the other maintained assumptions of the model are falsifiable.

One set of results focused on models with an index that is linear in the instruments. Building on our own earlier work, we showed that a full support condition on the instruments can identify the linear index model even with no restriction on the density of unobservables. At an opposite extreme, we provided conditions on the invertibility of the second derivative of the log-density such that a model with a linear index can be identified using arbitrarily small support on the instruments. Inbetween, our results allow a wide range of possible intermediate conditions on support and density. We have also shown how these results can be extended to models with a nonlinear index. For that case we have shown that mild restrictions on the behavior of the density near a critical point can allow identification even with very limited support for the instruments.

Together these results substantially expand the set of conditions under which nonparametric fully simultaneous equations models are known to be identified. They also demonstrate the robust identifiability that holds in models with Matzkin's residual index structure. These results are relevant to a wide range of applications of simultaneous equations models in economics. Although we have focused exclusively on identification, our results provide a more robust foundation for existing estimators and, through our constructive identification arguments, may suggest strategies for new estimation and testing approaches.

Appendices

A Proof of Remark 7

Below we let $\mathcal{B}(u, \epsilon)$ denote an ϵ-ball around a point $u \in \mathbb{R}^J$. To prove Remark 7 we rely on three lemmas.

Lemma 5. *Let S be a connected compact subset of $\mathbb{R}^J$ with nonempty interior, and let $h : S \to \mathbb{R}$ be a continuous function with upper contour sets $\mathcal{A}(c) = \{u \in S : h(u) \geq c\}$. Suppose that for some $\underline{c} < c_{\max} \equiv \arg\max_{u \in S} h(u)$, $\mathcal{A}(\underline{c}) \subset int(S)$. Then $\mathcal{A}(c)$ has nonempty interior for all $c < c_{\max}$.*

Proof. If $h(u) = c_{\max}$ for all $u \in S$, the result is immediate. Otherwise, because the continuous image of a connected set is connected, $h(S)$ is a nonempty interval. Thus, for any $c < c_{\max}$ there exists $u \in S$ such that $h(u) \in (\max\{\underline{c}, c\}, c_{\max})$. Since $\mathcal{A}(\underline{c}) \subset \text{int}(S)$, such u lies in $\{\mathcal{A}(c) \cap \mathcal{A}(\underline{c})\} \subset \text{int}(S)$. Thus, for sufficiently small $\epsilon > 0$, $\mathcal{B}(u, \epsilon) \subset S$ and, by continuity of h, $h(\hat{u}) > c \; \forall \hat{u} \in \mathcal{B}(u, \epsilon)$. Thus $\mathcal{A}(c)$ contains an open subset of $\mathbb{R}^J$. □

Lemma 6. *Let S be a connected compact subset of $\mathbb{R}^J$ with nonempty interior, and let $h : S \to \mathbb{R}$ be a continuous function with upper contour sets $\mathcal{A}(c) = \{u \in S : h(u) \geq c\}$. Suppose that for some $\underline{c} \in \mathbb{R}$, (i) $\mathcal{A}(\underline{c}) \subset int(S)$ and (ii) the restriction of h to $\mathcal{A}(\underline{c})$ attains a maximum $\overline{c} > \underline{c}$ at its unique critical value u^*. Then $\mathcal{A}$ is a continuous correspondence on $(\underline{c}, \overline{c}]$.*

Proof. For all $c \in (\underline{c}, \overline{c}]$, $\mathcal{A}(c)$ contains u^* and is therefore nonempty. Since S is compact and h is continuous, $\mathcal{A}$ is compact-valued. Suppose upper hemi-continuity of $\mathcal{A}$ fails at some point $\hat{c} \in (\underline{c}, \overline{c}]$. Then there must exist sequences $c^n \to \hat{c}$ and $u^n \to u$ such that $u^n \in \mathcal{A}(c^n)$ for all n but $u \notin \mathcal{A}(\hat{c})$. The latter requires $h(u) < \hat{c}$, since $\lim_{n\to\infty} u^n$ must lie in S. But by continuity of h this would imply $h(u^n) < c^n$ for sufficiently large n—a contradiction. To show lower hemi-continuity,[22] take arbitrary $\hat{c} \in (\underline{c}, \overline{c}]$, $\hat{u} \in \mathcal{A}(\hat{c})$, and $c^n \to \hat{c}$. If $\hat{u} = u^*$

[22] This argument is similar to that used to prove Proposition 2 in Honkapohja (1987).

then $\hat{u} \in \mathcal{A}(c)$ for all $c \leq \overline{c}$, so with the constant sequence $u^n = \hat{u}$ we have $u^n \in \mathcal{A}(c^n)$ for all n and $u^n \to \hat{u}$. So now suppose that $\hat{u} \neq u^*$. Define a sequence u^n by

$$u^n = \arg\min_{u \in \mathcal{A}(c^n)} \|u - \hat{u}\| \tag{A.1}$$

so that $u^n \in \mathcal{A}(c^n)$ by construction. We now show $u^n \to \hat{u}$. Take arbitrary $\epsilon > 0$. Because $\hat{u}$ lies in the pre-image of the open set $(\underline{c}, \overline{c})$ under the continuous map h, $\hat{u}$ must lie in the interior of $\mathcal{A}(\underline{c})$. So for sufficiently small $\delta > 0$ we have $\mathcal{B}(\hat{u}, \delta) \subset \mathcal{A}(\underline{c})$. Thus, $\{\mathcal{B}(\hat{u}, \epsilon) \cap \mathcal{A}(\underline{c})\}$ contains an open set $\mathcal{B}(\hat{u}, \delta)$ for some $\delta > 0$. If $h(u) \leq h(\hat{u})$ for all u in that set, $\hat{u}$ would be a critical value of h. Since $\hat{u}$ is not a critical value, there must exist $u^\epsilon \in \{\mathcal{B}(\hat{u}, \epsilon) \cap \mathcal{A}(\underline{c})\}$ such that $h(u^\epsilon) > h(\hat{u})$. Since we also have $h(\hat{u}) \geq \hat{c}$ and $c^n \to \hat{c}$, for n sufficiently large we have (by continuity of h) $h(u^\epsilon) > c^n$, implying that $u^\epsilon \in \mathcal{A}(c^n)$. So by (A.1), for n sufficiently large we have $\|u^n - \hat{u}\| \leq \|u^\epsilon - \hat{u}\| \leq \epsilon$. □

Lemma 7. *Let S be a connected compact subset of $\mathbb{R}^J$ with nonempty interior, and let $h : S \to \mathbb{R}$ be a continuous function with upper contour sets $\mathcal{A}(c) = \{u \in S : h(u) \geq c\}$. Suppose that for some $\underline{c}$, (a) $\mathcal{A}(\underline{c})$ $\in$int(S) and (b) the restriction of h to $\mathcal{A}(\underline{c})$ attains a maximum $\overline{c} > \underline{c}$ at its unique critical value u^*. Then $\mathcal{A}(c)$ is a connected set for all $c \in (\underline{c}, \overline{c})$.*

Proof. Proceeding by contradiction, suppose that for some $c \in (\underline{c}, \overline{c})$ the upper contour set $\mathcal{A}(c)$ is the union of disjoint nonempty open (relative to $\mathcal{A}(c)$) sets $\mathcal{A}^1$ and $\mathcal{A}^2$. Let $u^* \in \mathcal{A}^1$ without loss. Because $\mathcal{A}(\underline{c})$ $\in$int(S), $\mathcal{A}^2$ must be bounded and, by continuity of h, closed relative to $\mathbb{R}^J$. The restriction of h to $\mathcal{A}^2$ must therefore attain a maximum u^{**}. But since u^{**} must lie in the interior of $\mathcal{A}(\underline{c})$, u^{**} must be a critical value of h on $\mathcal{A}(\underline{c})$.

□

With these preliminary results, we can now prove Remark 7, restated below for convenience. Recall that for $c \in \mathbb{R}$ and $\Sigma \subset \mathbb{R}^J$ we let $\mathcal{A}(c; \Sigma)$ denote the upper contour set of the restriction of f to Σ.

Remark 7. Let f be continuously differentiable and suppose there is a compact connected set $S \subset \mathbb{R}^J$ with nonempty interior such that for some $\underline{c} \in \mathbb{R}$, (i) $\mathcal{A}(\underline{c}; S) \subset \text{int}(S)$, and (ii) the restriction of f to $\mathcal{A}(\underline{c}; S)$ attains a maximum $\overline{c} > \underline{c}$ at its unique critical value u^*. Then f satisfies local rectangle regularity.

Proof. We first show that, for any J-dimensional rectangle $\mathcal{M}$, there exists $c^0 \in (\underline{c}, \bar{c})$ such that $\mathcal{A}(c^0; S)$ is contained in a J-dimensional rectangle $\mathcal{N}$ that is smaller than $\mathcal{M}$. Because S is compact and f is continuous, $\mathcal{A}(c; S)$ is compact for all c.Further, by Lemma 6, $\mathcal{A}(c; S)$ is a continuous correspondence on $(\underline{c}, \overline{c}]$.[23] Thus $\max_{u \in \mathcal{A}(c;S)} u_j$ and $\min_{u \in \mathcal{A}(c;S)} u_j$ are continuous in $c \in (\underline{c}, \overline{c}]$, implying that the function $H : (\underline{c}, \overline{c}] \to \mathbb{R}$ defined by

$$H(c) = \max_j \max_{\substack{u^+ \in \mathcal{A}(c;S) \\ u^- \in \mathcal{A}(c;S)}} u_j^+ - u_j^-$$

is continuous. Thus, since $H(\bar{c}) = 0$, for some $c^0 \in (\underline{c}, \bar{c})$ the J-dimensional rectangle (Lemma 5 ensures that each interval is nonempty)

$$\mathcal{N} = \times_j \left(\min_{u^- \in \mathcal{A}(c^0;S)} u_j^-, \max_{u^+ \in \mathcal{A}(c^0;S)} u_j^+ \right) \tag{A.2}$$

is smaller than $\mathcal{M}$. To complete the proof, we show that f is regular on $\mathcal{N}$. By construction $u^* \in \mathcal{A}(c^0; S) \subset \mathcal{N}$. Now take arbitrary j and any $u_j \neq u_j^*$ such that $(u_j, u_{-j}) \in \mathcal{N}$ for some u_{-j}. By Lemma 7 and the definition of $\mathcal{N}$, there must also exist $\tilde{u}_{-j}$ such that $(u_j, \tilde{u}_{-j}) \in \mathcal{A}(c^0; S)$. Let $\hat{u}(u_j)$ solve

$$\max_{\hat{u} \in \mathcal{A}(\underline{c};S):\ \hat{u}_j = u_j} f(\hat{u}).$$

This solution must lie in $\mathcal{A}(c^0; S) \subset \mathcal{N}$ and satisfy the first-order conditions

$$\frac{\partial f(\hat{u}(u_j))}{u_k} = 0 \quad \forall k \neq j.$$

[23]In Lemmas 5–7, let $\mathcal{A}(c) = \mathcal{A}(c; S)$ and let h be the restriction of f to S.

Since $u_j \neq u_j^*$, we have $\frac{\partial f(\hat{u}(u_j))}{u_j} \neq 0$. □

B Additional Discussion of Matzkin 2010

In the text we showed that the identification results in Matzkin (2008, 2010) for the residual index model are special cases of our results. Matzkin (2010) also included two alternative sets of conditions sufficient for partial identification in a closely related model with two equations and a single instrument. In this appendix we show that natural extensions her conditions for partial identification lead to full identification of the residual index model. Further, both of these extensions turn out to be special cases of our Theorem 2.

Matzkin (2010) (section 4.1) considers a two-equation model of the form

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

She considers the case in which one is interested in identification only of the first equation. Maintaining the assumptions[24] that U has a joint density with support $\mathbb{R}^2$, that there exists a one-to-one, onto, $\mathcal{C}^2$ function $r = (r_1, r_2)$ such that

$$\begin{aligned} r_1(Y) &= U_1 \\ r_2(Y) &= U_2 + X_2 \end{aligned}$$

she offers two alternative sufficient conditions for identification of the derivatives $\partial m_1(y_2, u_1)/\partial y_2$.[25]

[24]She also assumes that X_2 has a $\mathcal{C}^1$ density.

[25]Matzkin stated these conditions as two alternative pairs of conditions. Her Assumptions 4.5 and 4.6 are together equivalent to Assumption M. Her Assumptions 4.5′ and 4.6′ together are equivalent to Assumption M′.

Assumption M. For all y, there exists $x_2^*(y) \in \mathbb{X}$ such that

$$\frac{\partial^2 \ln f\left(r_1(y), r_2(y) - x_2^*(y)\right)}{\partial u_2 \partial u_2} = 0.$$

At any such value

$$\frac{\partial^2 \ln f_U\left(r_1(y), r_2(y) - x_2^*(y)\right)}{\partial u_1 \partial u_2} \neq 0.$$

Assumption M′. For all y, there exist distinct values $x_2^*(y)$ and $x_2^{**}(y)$ in $\mathbb{X}$ such that

$$\frac{\partial \ln f\left(r_1(y), r_2(y) - x_2^*(y)\right)}{\partial u_2} = \frac{\partial \ln f\left(r_1(y), r_2(y) - x_2^{**}(y)\right)}{\partial u_2} = 0.$$

At any such values,

$$\frac{\partial \ln f\left(r_1(y), r_2(y) - x_2^*(y)\right)}{\partial u_1} \neq \frac{\partial \ln f\left(r_1(y), r_2(y) - x_2^{**}(y)\right)}{\partial u_1}.$$

Matzkin (2010) does not discuss extension of her identification results to identification of r and f to the full model[26]

$$\begin{aligned} r_1(Y) &= U_1 + X_1 \\ r_2(Y) &= U_2 + X_2. \end{aligned}$$

However, such extension is straightforward under a natural extension of either Assumption M or Assumption M′. Further, although Matzkin's identification arguments for the two cases are quite different, we show that both can be obtained as special cases of our Theorem 2.

[26]Because the distinction between $J = 2$ and $J > 2$ is trivial here, we keep $J = 2$ for simplicity.

We begin with the proposed extensions of Matzkin's assumptions:

Assumption EM. For all y and $j \in \{1,2\}$, there exists $x^*(y,j) \in \mathbb{X}$ such that

$$\frac{\partial^2 \ln f(r(y) - x^*(y,j))}{\partial u_j \partial u_j} = 0.$$

At such values

$$\frac{\partial^2 \ln f(r(y) - x^*(y,j))}{\partial u_j \partial u_k} \neq 0 \quad \text{for } k \neq j.$$

Assumption EM′. For all y and $j \in \{1,2\}$, there exist distinct values $x^*(y,j)$ and $x^{**}(y,j)$ in $\mathbb{X}$ such that

$$\frac{\partial \ln f(r(y) - x^*(y,j))}{\partial u_j} = \frac{\partial \ln f(r(y) - x^{**}(y,j))}{\partial u_j} = 0.$$

At any such values,

$$\frac{\partial \ln f(r(y) - x^*(y,j))}{\partial u_{-j}} \neq \frac{\partial \ln f(r(y) - x^{**}(y,j))}{\partial u_{-j}}.$$

Remark 10. *Assumption EM implies that for all y,*[27] *there exists points $x^*(y,1)$ and $x^*(y,2)$ in $\mathbb{X}$ such that the matrix*

$$\begin{pmatrix} \frac{\partial^2 \ln f(r(y)-x^*(y,1))}{\partial u_1 \partial u_1} & \frac{\partial^2 \ln f(r(y)-x^*(y,1))}{\partial u_1 \partial u_2} \\ \frac{\partial^2 \ln f(r(y)-x^*(y,2))}{\partial u_2 \partial u_1} & \frac{\partial^2 \ln f(r(y)-x^*(y,2))}{\partial u_2 \partial u_2} \end{pmatrix}$$

has zero diagonal terms, nonzero anti-diagonal terms, and is therefore is nonsingular. Thus, given Assumptions 1 and 2, identification of (r, f) under Assumption EM follows from Corollary 3.

Proof. Immediate. □

[27] Note that the argument here would hold if this property were true only for almost all y. The same is true for Remark 11 below.

Remark 11. *Assumption EM′ implies that for all* y*, the matrix*

$$\begin{pmatrix} d\left(y, x^{*}\left(y,1\right)\right) \\ d\left(y, x^{**}\left(y,1\right)\right) \\ d\left(y, x^{*}\left(y,2\right)\right) \\ d\left(y, x^{**}\left(y,2\right)\right) \end{pmatrix}$$

has full column Rank. Thus, given Assumption 1, identification of (r, f) *under Assumption EM′ follows from Theorem 2.*

Proof. By Assumption EM′, the matrix can be written

$$\begin{pmatrix} 1 & 0 & a \\ 1 & 0 & b \\ 1 & c & 0 \\ 1 & d & 0 \end{pmatrix}$$

where we know $a \neq b$ and $c \neq d$. One step of Gaussian elimination transforms the matrix to the form

$$\begin{pmatrix} 1 & 0 & a-b \\ 0 & 0 & 0 \\ 1 & c & 0 \\ 1 & d & 0 \end{pmatrix}.$$

Since the matrix

$$\begin{pmatrix} 1 & 0 & a-b \\ 1 & c & 0 \\ 1 & d & 0 \end{pmatrix}$$

has determinant $(d-c)(a-b)$, the result follows. □

References

Andrews, D. W. K. (2011): "Examples of L^2-Complete and Boundedly Complete Distributions," Discussion Paper 1801, Cowles Foundation, Yale University.

Ayer, A. J. (1936): *Language, Truch and Logic*. Penguin, London.

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. 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 1732, Cowles Foundation, Yale University.

——— (2013): "Identification in a Class of Nonparametric Simultaneous Equations Models," Discussion Paper 1787RR, Cowles Foundation, Yale University.

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82(5), 1749–1797.

Blundell, R., and R. L. Matzkin (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, 5(2), 271–295.

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).

D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471.

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.

Giaquinta, M., and G. Modica (2007): *Mathematical Analysis: Linear and Metric Structures and Continuity*. Birkhaeuser.

Honkapohja, S. (1987): "On Continuity of Compensated Demand," *International Economic Review*, 28, 545–557.

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.

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

Lehmann, E. L., and H. Scheffe (1950): "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya*, 10, 305–340.

——— (1955): "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya*, 15, 219–236.

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.