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

August 28, 2017

Abstract

We present new identification results for a class of nonseparable nonparametric simultaneous equations models introduced by Matzkin (2008). These models combine traditional exclusion restrictions with a requirement that each structural error enter through a “residual index.” Our identification results are constructive and encompass a range of special cases with varying demands on the exogenous variation provided by instruments and the shape of the joint density of the structural errors. The most important results demonstrate identification when instruments have only limited variation. Even when instruments vary only over a small open ball, relatively mild conditions on the joint density suffice. We also show that the primary sufficient conditions for identification are verifiable and that the maintained hypotheses of the model are falsifiable.

*We received helpful comments from Xiaohong Chen, Elie Tamer, Alex Torgovitzky, the referees, seminar/conference audiences at the LSE, 2013 Econometrics of Demand Conference, 2014 SITE Summer Conference, 2014 NASM of the Econometric Society, 2014 IAAE Conference, 2014 CEMMAP Nonparametric Demand Conference, and especially Louise Laage. We thank Wayne Gao, Koo Hyun Kwon and Ming Li for capable research assistance, and 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 almost any economic setting in which multiple agents interact or a single agent makes multiple interrelated choices (see Appendix A for examples). The identifiability of simultaneous equations models is therefore an important question for a wide range of topics in empirical economics. Although early work on (parametric) identification treated systems of simultaneous equations as a primary focus, nonparametric identification 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 simultaneous system (more general than we will allow) is given by

\[ m_j(Y, Z, U) = 0 \quad j = 1, \ldots, J \]  

(1)

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

\[ U_j = \rho_j(Y, Z) \quad j = 1, \ldots, J. \]  

(2)

Identification of such models was considered by ?, ?, ?, and ?. However, a claim made in ? and relied upon by the others implied that traditional exclusion restrictions would identify the model when \( U \) is independent of \( Z \). ? showed that this claim is incorrect, leaving uncertain the nonparametric identifiability of fully simultaneous models.

Completeness conditions (??) offer one possible approach, and in ? we showed how identification arguments in ? or ? can be adapted to an example of the class of models considered below. However, independent of general concerns one might have with the interpretability of completeness conditions, this approach may be particularly unsatisfactory in a simultaneous equations setting. A simultaneous equations model already specifies the structure generating 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. At a minimum, constructive arguments can complement identification results relying on completeness conditions by

\(^1\)See, e.g., ?, ?, and ?.

\(^2\)See, e.g., ?, ?, and ? for conditions ensuring invertibility in different contexts.

\(^3\)? show identification in a related model by combining completeness conditions and additional arguments using the classic change of variables approach.

\(^4\)Recent work on this issue includes ? and ?. 
demonstrating how exogenous variation can pin down the structural features of interest, how much variation is sufficient, and the extent to which conditions beyond exogenous variation in instruments may be needed.

Much recent work has focused on triangular (recursive) systems of equations (e.g., 2, 4). A two-equation version of the triangular model takes the form

\begin{align*}
Y_1 &= m_1(Y_2, Z, U_1) \\
Y_2 &= m_2(Z, X, U_2)
\end{align*}

where \(U_2\) is a scalar error entering \(m_2\) monotonically and \(X\) is an exogenous observable excluded from the first equation. This structure often arises in a program evaluation setting. To contrast this model with a fully simultaneous system, suppose \(Y_1\) represents the quantity sold of a good and that \(Y_2\) is its price. If (3) is the structural demand equation, (4) should be the reduced form for price, with \(X\) denoting a supply shifter excluded from demand. However, typically both the demand error \(U_1\) and the supply error \(U_2\) would enter the reduced form for price.\(^5\) One obtains the triangular model only when the two structural errors enter the reduced form for price monotonically through a single index. This is a strong index assumption quite different from the residual index structure we consider.\(^?\) provide a necessary and sufficient condition for a simultaneous model to reduce to the triangular model, pointing out that this condition is quite restrictive.

An important breakthrough in the literature on fully simultaneous models came in \(\text{?}\). Matzkin considered a model of the form

\[ m_j(Y, Z, \delta) = 0 \quad j = 1, \ldots, J \]

where \(\delta = (\delta_1(Z, X_1, U_1), \ldots, \delta_j(Z, X_j, U_j))^T\) is a vector of indices

\[ \delta_j(Z, X_j, U_j) = g_j(Z, X_j) + U_j, \]

and each \(g_j(Z, X_j)\) is strictly increasing in \(X_j\). Here \(X = (X_1, \ldots, X_J)^T \in \mathbb{R}^J\) play a special role as exogenous observables (instruments) specific to each equation. This formulation thus respects traditional exclusion restrictions in that \(X_j\) is excluded from equations \(k \neq j\) (e.g., there is a “demand shifter” that enters only the demand equation and a “cost shifter” that enters only the supply equation). However, it restricts the more general model (1) by requiring \(X_j\) and \(U_j\) to enter the nonparametric function \(m_j\) through a “residual index” \(\delta_j(Z, X_j, U_j)\). Given invertibility of \(m\) (now in \(\delta\)), the analog of (2) is \(\delta_j(Z, X_j, U_j) = r_j(Y, Z), \quad j = 1, \ldots, J\), or equivalently,\(^6\)

\[ r_j(Y, Z) = g_j(Z, X_j) + U_j \quad j = 1, \ldots, J. \]

\(^5\)With \(J\) goods, all \(2J\) demand shocks and cost shocks would typically enter the reduced form for each price. Example 3 in Appendix A illustrates.

\(^6\)This model can be interpreted as a generalization of the transformation model to a simultaneous system. The semiparametric transformation model (e.g., \(\text{?}\)) takes the form \(t(Y) = Z\beta + U\), where \(Y \in \mathbb{R}, U \in \mathbb{R}\), and the unknown transformation function \(t\) is strictly increasing. Besides replacing
This is the model we study as well. Appendix A illustrates this structure in several important classes of applications. Some of these generalize classic systems of simultaneous equations that arise when multiple agents interact in equilibrium. The residual index structure can be directly imposed on the system of nonparametric simultaneous equations or derived from assumptions on primitives generating this system. In Appendix A we illustrate the latter in an equilibrium model of differentiated product markets. This appendix also shows how the simultaneous equations model arises from the interdependent decisions of a single agent, using an example of firm input demand. In that example, the residual index structure again emerges naturally from assumptions on model primitives.

Matzkin (2008) showed that the residual index model is identified when $U$ is independent of $X$, $(g_1 (X_1 Z), \ldots, g_J (X_J, Z))$ has large support conditional on $Z$, and the joint density (or log density) of $U$ satisfies certain global restrictions. This was, to our knowledge, the first result demonstrating identification in a fully simultaneous nonparametric model with nonseparable errors. This provided additional results and estimation strategies for a special case in which each residual index function $g_j (Z, X_j)$ is linear conditional on $Z$.

We provide new constructive identification results for the model (6). Along the way we point out 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 are falsifiable. After completing the model setup in section 2, in section 3 we develop a general sufficient condition for identification of the functions $g_j$. This “rectangle regularity” condition is implied by Matzkin’s (2008) combination of large support and global density conditions, but also holds when the instruments $X$ have limited support under a local density condition. Once each function $g_j$ is known, identification of the model follows as in the special case of a linear residual index function. To exploit this fact, in section 4 we review that special case and provide new sufficient conditions for identification. By combining these results with $Z\beta$ and $g (Z, X)$, (6) generalizes this model by dropping the requirement of a monotonic transformation function and, more fundamental, allowing a vector of outcomes $Y$ to enter each unknown transformation function. \textsuperscript{7} considers a nonparametric version of the single-equation transformation model. See also \textsuperscript{8}.

\textsuperscript{7} used a new characterization of observational equivalence to provide sufficient conditions for identification in a linear simultaneous equations model, a single equation model, a triangular (recursive) model, and a fully simultaneous nonparametric model (her “supply and demand” example) of the form (6).

\textsuperscript{8} also provides conditions for identification of certain features in models that partially relax the residual index structure. \textsuperscript{?} consider a nonparametric framework permitting simultaneity, providing a characterization of sharp identified sets under various restrictions. \textsuperscript{?} considers a linear (semiparametric) simultaneous equations model—either with two equations or having a “linear-in-means” structure—with random coefficients on the endogenous variables. Using combinations of support and density restrictions, he considers identification of the marginal and joint distributions of the random coefficients.

---

\textsuperscript{7} also provides conditions for identification of certain features in models that partially relax the residual index structure. \textsuperscript{8} consider a nonparametric framework permitting simultaneity, providing a characterization of sharp identified sets under various restrictions. \textsuperscript{?} considers a linear (semiparametric) simultaneous equations model—either with two equations or having a “linear-in-means” structure—with random coefficients on the endogenous variables. Using combinations of support and density restrictions, he considers identification of the marginal and joint distributions of the random coefficients.
(section 5) one obtains identification of the full (nonlinear index) model (6) under a variety of alternative support and density conditions. Our most important result may be Corollary 2, which allows instruments with limited support—for example, a small open ball. Given the maintained technical conditions of the model, this result requires that the log density of $U$ have a nondegenerate local maximum and a Hessian matrix that is invertible at a sufficiently rich set of (possibly isolated) points. These requirements are met by familiar parametric log densities on $\mathbb{R}^J$, and a genericity result in the Supplemental Material demonstrates one formal sense in which these density conditions may be viewed as mild.

2 The Model

2.1 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 $Z$ are important in applications but 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, and all assumptions should be interpreted to hold conditional on $Z$. Stacking the equations in (6), we then consider the model

$$r(Y) = g(X) + U,$$

where $r(Y) = (r_1(Y), \ldots, r_J(Y))^\top$ and $g(X) = (g_1(X_1), \ldots, g_J(X_J))^\top$. Let $\mathbb{X} = \text{int} (\text{supp}(X))$, and let $\mathbb{Y}$ denote the pre-image of $\text{supp}(g(X) + U)$ under $r$. The following describes the maintained assumptions on the model, following Matzkin (2008).

**Assumption 1.** (i) $X$ is nonempty and connected; (ii) $g$ is continuously differentiable, with $\partial g_j(x_j)/\partial x_j > 0 \forall j, x_j$; (iii) $U$ is independent of $X$ and has a twice continuously differentiable joint density $f$ that is positive on $\mathbb{R}^J$; (iv) $r$ is injective, twice differentiable, and has nonzero Jacobian determinant $J(y) = \det (\partial r(y)/\partial y) \forall y \in \mathbb{Y}$.

Part (i) rules out instruments with discrete or disconnected support. Part (ii) requires monotonicity and differentiability in the instruments. The primary role of parts (iii) and (iv) is to allow us to attack the identification problem using a standard change of variables approach (see, e.g., ?), relating the joint density of observables to that of the structural errors. In particular, letting $\phi(\cdot | X)$ denote the joint density of

---

9We strengthen Matzkin’s (2008) assumption that $f$ is continuously differentiable to twice continuous differentiability. Although we maintain this assumption from the beginning to simplify exposition, all results up to equation (32) hold with only continuous differentiability, letting Condition $M'$ in Appendix B replace Condition $M$ in the text. Our identification arguments using second derivatives also generalize to differenced first derivatives (see ?).

10? provide an initial exploration of identification when instruments are discrete.
conditional on $X$, we have
\[ \phi(y|x) = f(r(y) - g(x))|J(y)|. \] (8)

In addition, we have the following lemma.

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

**Proof.** See Appendix C.

With this result, below we treat $\phi(y|x)$ as known for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}$.

### 2.2 Normalizations

We impose three standard normalizations.\(^{11}\) First, observe that all relationships between $(Y, X, U)$ would be unchanged if for some constant $\kappa_j$, $g_j(X_j)$ 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 $\hat{y} \in \mathbb{Y}$ and arbitrary point $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_J) \in \mathbb{R}^J$ we set
\[ r_j(\hat{y}) = \hat{r}_j \quad \forall j. \] (9)

Similarly, since even with (9), (7) would be unchanged if, for every $j$, $g_j(X_j)$ were replaced by $g_j(X_j) + \kappa_j$ for some constant $\kappa_j$ while $U_j$ were replaced by $U_j - \kappa_j$, we take an arbitrary point $\hat{x} \in \mathbb{X}$ and set
\[ g_j(\hat{x}_j) = \hat{x}_j \quad \forall j. \] (10)

Given (9), this fixes the location of each $U_j$, but we must still choose its scale.\(^{12}\) In particular, since (7) would continue to hold if, for each $j$, we multiplied $r_j$, $g_j$ and $U_j$ by a nonzero constant $\kappa_j$, we normalize the scale of each $U_j$ by setting
\[ \frac{\partial g_j(\hat{x}_j)}{\partial x_j} = 1 \quad \forall j. \] (11)

Finally, given (9) and (10), a convenient choice of $\hat{r}$ sets each $\hat{r}_j = \hat{x}_j$, so that
\[ r_j(\hat{y}) - g_j(\hat{x}_j) = 0 \quad \forall j. \] (12)

\(^{11}\)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 (10). Alternatively we could follow ?, who makes no normalizations in her supply and demand example and shows only that derivatives of each $r_j$ and $g_j$ are identified up to scale.

\(^{12}\)Typically the location and scale of the structural errors 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 of referring to these restrictions as normalizations.
2.3 Identifiability, Verifiability, and Falsifiability

Before proceeding, we must define some key terminology. Following ?? and ?, a *structure* $S$ is a data generating process, i.e., a set of probabilistic or functional relationships between the observable and latent variables that implies (generates) a joint distribution of the observables. Let $\mathcal{S}$ denote the set of all structures. The true structure is denoted $S_0 \in \mathcal{S}$. A *hypothesis* is any nonempty subset of $\mathcal{S}$. A hypothesis $H$ is true (satisfied) if $S_0 \in H$.\(^{13}\)

A *structural feature* $\theta(S_0)$ is a functional of the true structure $S_0$. A structural feature $\theta(S_0)$ is *identified* (or *identifiable*) under the hypothesis $H$ if $\theta(S_0)$ is uniquely determined within the set \{ $\theta(S): S \in H$ \} by the joint distribution of observables. The primary structural features of interest in our setting are the functions $r$ and $g$.\(^{14}\) However, we will also be interested in binary features indicating whether key hypotheses hold. Given a maintained hypothesis $\mathcal{M}$, we will say that a hypothesis $H \subset \mathcal{M}$ is *verifiable* if the indicator $1 \{ S_0 \in H \}$ is identified under $\mathcal{M}$.\(^{15}\) Thus, when a hypothesis is verifiable, its satisfaction or failure is an identified feature.\(^{16}\)

A weaker and more familiar notion is that of falsifiability. Let $P_H$ denote the set of probability distributions (for the observables) generated by structures in a set $H$. Given a maintained hypothesis $\mathcal{M}$, $H \subset \mathcal{M}$ is *falsifiable* if $P_H \neq P_{\mathcal{M}}$. Thus, as usual, a hypothesis is falsifiable when it implies a restriction on the observables. A hypothesis 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 satisfactory statistical test. Although our arguments may suggest approaches for estimation or hypothesis testing, we leave all such issues for future work.

---

\(^{13}\) ?? and ? call any strict subset of $\mathcal{S}$ a *model*. Some authors make distinctions between the notions of “model,” “identifying assumptions,” or “overidentifying assumptions.” All of these notions are nested by the term hypothesis.

\(^{14}\) Note that the joint density $f$ is determined by these functions and the observables. In practice, quantities of interest will often include particular functionals of $(r, g, f)$. As pointed out by ??, identification of such functionals may sometimes be obtained under weaker conditions than those needed for identification of the model. Exploration of such possibilities in particular applications is a potentially important topic for further work. See ?? for some results in the case of differentiated products supply and demand.

\(^{15}\) We use the symbol $\subset$ for all (proper or not) subset relationships.

\(^{16}\) We are not aware of prior formal use of the notion of verifiability in the econometrics literature although informal use is common and, as our definition makes clear, this is merely a particular case of identifiability.
3 Identification of the Index Functions

We begin by considering identification of the index functions $g_j$. Taking logs in (8) and differentiating yields

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

(13)

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

(14)

Together (13) and (14) imply

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

(15)

Our approach in this section builds on an insight in Matzkin (2008), isolating unknowns in (15) with critical points of the log density $\ln f$ and “tangencies” to its level sets. We begin in section 3.1 by introducing a property of $(\mathbb{X}, f, g)$ that we call rectangle regularity. In section 3.2 we then show that rectangle regularity is sufficient for identification of the index functions $g_j$. Finally, we discuss two simpler sufficient conditions for rectangle regularity. The most important of these requires only that $\ln f$ have a local maximum somewhere on its support.

3.1 Rectangle Regularity

To state our general sufficient condition for identification of $g$, we require two new definitions. The first is standard and provided here only to avoid ambiguity.

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

Whenever we refer to a “rectangle” below, we mean a $J$-dimensional rectangle. Next, we introduce a notion of regularity, requiring that $\ln f$ have a critical point $u^*$ in a rectangular neighborhood $\mathcal{U}$ in which its level sets are “nice” in a sense defined by part (ii) of the following definition.

**Definition 2.** Given a $J$-dimensional rectangle $\mathcal{U} \equiv \times_{j=1}^J (u_j, \pi_j)$, $\ln f$ is regular on $\mathcal{U}$ if (i) there exists $u^* \in \mathcal{U}$ such that $\partial \ln f(u^*)/\partial u_j = 0 \ \forall j$; and (ii) for all $j$, almost all $u'_j \in (u_j, \pi_j)$, and some $\hat{u}_j (j, u'_j) = (\hat{u}_1 (j, u'_j), \ldots, \hat{u}_J (j, u'_j)) \in \mathcal{U}$,

$$\hat{u}_j (j, u'_j) = u'_j \text{ and}$$

$J(y)$ is a polynomial in the first partial derivatives of $r$ and is therefore differentiable. Then because $J(y)$ is everywhere nonzero, it can take only one sign on $Y$, ensuring that $|J(y)|$ (and therefore $\ln |J(y)|$) is differentiable.

7
\[
\frac{\partial \ln f(\hat{u}(j, u_j'))}{\partial u_k} \neq 0 \quad \text{iff} \ k = j.
\]

In Definition 2, \(\hat{u}(j, u_j')\) has a geometric interpretation as a point of tangency between a level set of \(\ln f\) and the hyperplane \(\{u \in \mathbb{R}^J : u_j = u_j'\}\). Part (ii) of the definition 2 requires such a tangency within the rectangle \(\mathcal{U}\) in each dimension \(j\).

Figure 1 illustrates an example in which \(J = 2\) and \(u^*\) is a local maximum. Within a neighborhood of \(u^*\) the level sets of \(\ln f\) are connected and smooth, representing strictly increasing values of \(\ln f\) as one moves towards \(u^*\). 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 log density is regular. One such rectangle, \(\mathcal{U} = (u_1, \bar{u}_1) \times (u_2, \bar{u}_2)\), is illustrated and is itself defined by tangencies to a level set.

We show below that the following condition ensures identification of the index functions \(g\).

**Assumption 2** (“Rectangle Regularity”). For every \(x \in \mathbb{X}\), \(\ln f\) is regular on a rectangle \(\mathcal{U}(x) = \times_j (u_j(x), \bar{u}_j(x))\), where

\[
\begin{align*}
  u_j(x) &= u_j^*(x) + g_j(x_j) - g_j(\bar{x}_j(x)) \quad \forall j, \\
  \bar{u}_j(x) &= u_j^*(x) + g_j(x_j) - g_j(x_j) \quad \forall j, \\
  u^*(x) &= (u_1^*(x), \ldots, u_J^*(x)) \quad \text{is a critical point of} \ \ln f, \ \text{and} \ \mathcal{X}(x) = \times_j (\underline{x}_j(x), \bar{x}_j(x))
  \end{align*}
\]

satisfies \(x \in \mathcal{X}(x) \subset \mathbb{X}\).

Rectangle regularity requires, for each \(x\), that \(\ln f\) be regular on a rectangular neighborhood around a critical point that maps through the model (7) to a rectangular neighborhood in \(\mathbb{X}\) around \(x\). Specifically, take an arbitrary \(x\). Let \(u^*(x)\) be a critical point of \(\ln f\) and let \(\times_j (u_j(x), \bar{u}_j(x)) \ni u^*(x)\) denote a rectangle on which \(\ln f\) is regular. Define \(y^*(x)\) by

\[
  r(y^*(x)) - g(x) = u^*(x),
\]

and define \(\underline{x}(x)\) and \(\bar{x}(x)\) by (substituting (17) into (16))

\[
\begin{align*}
  \underline{u}_j(x) &= r_j(y^*(x)) - g_j(\bar{x}_j(x)) \\
  \bar{u}_j(x) &= r_j(y^*(x)) - g_j(x_j)
\end{align*}
\]

for all \(j\). Figure 2 illustrates. Assumption 2 is satisfied if, for every \(x\), there exist \(u^*(x)\) and \(\times_j (\underline{u}_j(x), \bar{u}_j(x))\) such that the resulting rectangle \(\mathcal{X}(x) = \times_j (\underline{x}_j(x), \bar{x}_j(x))\) lies within \(\mathbb{X}\). It should be emphasized that although we write \(u^*(x)\), the same critical point may be used to construct \(\mathcal{X}(x)\) for many (even all) values of \(x\).

Because \(\mathbb{X}\) is open, a rectangle \(\mathcal{X}\) such that \(x \in \mathcal{X} \subset \mathbb{X}\) exists for every \(x \in \mathbb{X}\). Furthermore, when \(\mathbb{X}\) includes a rectangle \(\mathcal{M}\), it also includes all rectangles \(\mathcal{X} \subset \mathcal{M}\). Thus, because \(g(\mathcal{X})\) is a rectangle whenever \(\mathcal{X}\) is, the set \(\mathcal{X}(x)\) required by rectangle
Figure 1: The solid curves are level sets of a bivariate log density in a region of its support. The point $u^*$ is a local maximum. For each $u_1' \in (u_1, \overline{u}_1)$ the point $\hat{u}(1, u_1') = (u_1', \hat{u}_2(1, u_1'))$ is a point of tangency between the vertical line $U_1 = u_1'$ and a level set. The log density is regular on $\mathcal{U} = \times_j(u_j, \overline{u}_j)$.

Figure 2: For arbitrary $x \in \mathcal{X}$, the rectangle $\mathcal{U}$ in Figure 1 is mapped to a rectangle $\mathcal{X}(x)$ using (17) and (18), thereby satisfying (16).
regularity is guaranteed to exist as long as $\ln f$ is regular on some rectangle in $\mathbb{R}^J$ that is not too large relative to the support of $X$ around $x$. We use this insight to provide more transparent sufficient conditions for rectangle regularity in section 3.3 below (see also Appendix B).

The following result, proved in Appendix C, shows that rectangle regularity is equivalent to a condition on observables.

**Proposition 1.** Assumption 2 is verifiable.

### 3.2 Identification Under Rectangle Regularity

Under rectangle regularity, identification of the index functions $g_j$ follows in three steps. The first exploits a critical point of $\ln f$ to pin down derivatives of the Jacobian determinant at a point $y^*(x)$ for any $x$. The second uses tangencies to identify the ratios $\frac{\partial g_j(x')/\partial x_j}{\partial g_j(x)/\partial x_j}$ for all pairs of points $(x^0, x')$ in a sequence of overlapping rectangular subsets of $X$. The final step links these rectangular neighborhoods so that, using the normalization (11), we can integrate up to the functions $g_j$, using (10) as boundary conditions.

The first step is straightforward. For any $x \in X$, if $u^*(x)$ is a critical point of $\ln f$ and $y^*(x)$ is defined by (17), equation (14) yields

$$\frac{\partial \ln |J(y^*(x))|}{\partial y_k} = \frac{\partial \ln \phi(y^*(x) | x)}{\partial y_k} \forall k. \tag{19}$$

For arbitrary $x$ and $x'$ we can then rewrite (15) as

$$\sum_j \frac{\partial \ln \phi(y^*(x) | x')}{\partial x_j} \frac{\partial r_j((y^*(x)) / \partial y_k}{\partial x_j} = \frac{\partial \ln \phi(y^*(x) | x)}{\partial y_k} - \frac{\partial \ln \phi((y^*(x) | x')}{\partial y_k}. \tag{20}$$

By (13), $u^*(x)$ is a critical point of $\ln f$ if and only if, for $y^*(x)$ defined by (17), $\partial \ln \phi(y^*(x) | x) / \partial x_j = 0$ for all $j$. The values of all such $y^*(x)$ are observable. Given any such $y^*(x)$, this leaves the ratios $\frac{\partial r_j(y^*(x))}{\partial g_j(x'/x_j)}$ as the only unknowns in (20). This will allow us to demonstrate the second step in Lemma 2 below. Here we exploit the fact that, under Assumption 2, as $\hat{x}$ varies around the arbitrary point $x$, $r(y^*(x)) - g(\hat{x})$ takes on all values in a rectangular neighborhood of $u^*(x)$ on which $\ln f$ is regular.

**Lemma 2.** Let Assumptions 1 and 2 hold. Then for every $x \in X$ there exists a $J$-dimensional rectangle $\mathcal{X}(x) \ni x$ such that for all $x^0 \in \mathcal{X}(x) \setminus x$ and $x' \in \mathcal{X}(x) \setminus x$, the ratio $\frac{\partial g_j(x')}{\partial g_j(x)}$ is identified for all $j = 1, \ldots, J$.

**Proof.** Take arbitrary $x \in X$. Let $u^*$ and $U = \times_{j} (u_j, \pi_j)$ be as defined in Assumption 2, and let $y^*$ be as defined by (17).\(^{18}\) By Assumption 2 there exists $X = \times_{i} (x_i, \pi_i) \subset X$

\(^{18}\)To simplify notation, here we suppress dependence of $u^*, U, y^*, \bar{X}, \bar{u}, \bar{\pi}, \bar{u}, \bar{\pi}$ on the point $x$. 

10
(with \(x \in \mathcal{X}\)) such that (18) holds and (recalling (13)) such that for each \(j\) and almost every \(x'_j \in (\mathbf{x}_j, \mathbf{\bar{x}}_j)\) there is a \(J\)-vector \(\hat{x}\left(j, x'_j\right) \in \mathcal{X}\) satisfying

\[
\hat{x}_j\left(j, x'_j\right) = x'_j \quad \text{and} \quad \partial \ln \phi\left(y^*|\hat{x}(j, x'_j)\right) / \partial x_j \neq 0 \text{ iff } k = j. \tag{22}
\]

Since \(\phi (y|x)\) and its derivatives are observed for all \(y \in \mathcal{Y}\) and \(x \in \mathcal{X}\), every tuple \((y^*, \mathbf{x}_1, \mathbf{\bar{x}}_1, \ldots, \mathbf{x}_j, \mathbf{\bar{x}}_j)\) satisfying these conditions is identified, as are the (not necessarily unique) associated points \(\hat{x}\left(j, x'_j\right)\) satisfying (21) and (22). Take one such tuple and associated points. Taking arbitrary \(j, k, x'_j \in (\mathbf{x}_j, \mathbf{\bar{x}}_j)\), and the known point \(\hat{x}\left(j, x'_j\right)\), (20) becomes

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

By (22), we may rewrite this as

\[
\frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j} = \frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(j, x'_j))}{\partial y_k} / \partial x_j. \tag{23}
\]

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 (and, by continuity, all) \(x'_j \in (\mathbf{x}_j, \mathbf{\bar{x}}_j)\). By the same arguments leading up to (23), but with \(x'_j\) taking the role of \(x_j\), we obtain

\[
\frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j} = \frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(j, x'_j))}{\partial y_k} / \partial x_j. \tag{24}
\]

yielding identification of \(\frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j}\) for all \(x'_j \in (\mathbf{x}_j, \mathbf{\bar{x}}_j)\). Because the Jacobian determinant \(J(y)\) is nonzero, \(\frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j}\) cannot be zero for all \(k\). Thus for each \(j\) there is some \(k\) such that the ratio

\[
\frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j} / \frac{\partial r_j(y^*) / \partial y_k}{\partial g_j(x'_j) / \partial x_j}
\]

is well defined. This establishes the result.\(^{19}\)

\(^{19}\)Since the last step of 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.
The final step of the argument yields the main result of this section.

**Theorem 1.** Let Assumptions 1 and 2 hold. Then \( g \) is identified on \( \mathbb{X} \).

**Proof.** We first claim that Lemma 2 implies identification of the ratios \( \frac{\partial g_j(x_j^I)}{\partial x_j} / \frac{\partial g_j(x_j^0)}{\partial x_j} \) for all \( j \) and any two points \( x^0 \) and \( x^I \) in \( \mathbb{X} \). This follows immediately if there is some \( x \) such that \( \mathcal{X}(x) = \mathbb{X} \). Otherwise, observe that because each rectangle \( \mathcal{X}(x) \) guaranteed to exist by Lemma 2 is open, \( \{\mathcal{X}(x)\}_{x \in \mathbb{X}} \) is an open cover of \( \mathbb{X} \). Since \( \mathbb{X} \) is connected, for any \( x^0 \) and \( x^I \) in \( \mathbb{X} \) there exists a simple chain from \( x^0 \) to \( x^I \) consisting of elements (rectangles) from \( \{\mathcal{X}(x)\}_{x \in \mathbb{X}} \). Since the ratios \( \frac{\partial g_j(x_j^I)}{\partial x_j} / \frac{\partial g_j(x_j^0)}{\partial x_j} \) are known for all pairs \( (x_j^I, x_j^0) \) in each of these rectangles, it follows that the ratios \( \frac{\partial g_j(x_j^I)}{\partial x_j} / \frac{\partial g_j(x_j^0)}{\partial x_j} \) are identified for all \( j \). Taking \( x_j^0 = \dot{x}_j \) for all \( j \), the conclusion of the theorem then follows from the normalization (11) and boundary condition (10). \( \square \)

### 3.3 Sufficient Conditions for Rectangle Regularity

Here we offer two alternative sufficient conditions for Assumption 2 that are more easily interpreted. The first combines large support for the instruments \( X \) with regularity of \( \ln f \) on \( \mathbb{R}^J \); this is equivalent to the combination of conditions required by Matzkin (2008).\(^{21}\)

**Proposition 2.** Let Assumption 1 hold. Suppose that \( g(\mathbb{X}) = \mathbb{R}^J \) and that \( \ln f \) is regular on \( \mathbb{R}^J \). Then Assumption 2 holds.

**Proof.** Let \( \mathcal{X}(x) = \times_j (g_j^{-1}(\infty), g_j^{-1}(\infty)) \) for all \( x \). Then by (16), \( \mathcal{U}(x) = \mathbb{R}^J \), yielding the result. \( \square \)

Our second sufficient condition allows limited—even arbitrarily small—support for \( X \) while requiring only a local condition on the log density \( \ln f \).\(^{22}\)

**Condition M.** \( \ln f \) has a nondegenerate local maximum.

**Proposition 3.** Let Assumption 1 and Condition M hold. Then Assumption 2 holds.

**Proof.** See Appendix B. \( \square \)

\(^{20}\)See, e.g., ?, Lemma 1.5.21.

\(^{21}\)The density restriction stated in Matzkin (2008) is actually stronger, equivalent to assuming regularity of \( \ln f \) on \( \mathbb{R}^J \) but replacing “almost all \( u'_j \in (u_j, u_j) \)” in the definition of regularity with “all \( u'_j \in (u_j, u_j) \).” 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).

\(^{22}\)A nondegenerate local minimum would also suffice.
Our proof of Proposition 3 requires several steps, but intuition can be gained by returning to Figure 1. Recall that rectangle regularity holds when, for each point \( x \), \( \ln f \) is regular (has the requisite critical point and tangencies) on a rectangle that is not too big relative to the support of \( X \) around \( x \). In Figure 1, \( u^* \) is a non-degenerate local max. By the Morse lemma (e.g., \( \text{?} \), Corollary 2.18.), nondegenerate critical points are isolated. As the figures suggests, this ensures that there exist arbitrarily small rectangles around \( u^* \) on which \( \ln f \) is regular.

Neither of these two sufficient conditions for rectangle regularity implies the other.\(^{23}\) Nonetheless, Proposition 3 may be more important as it allows instruments with limited support (even a small open ball) while requiring only a mild condition on the log density.\(^{24}\)

4 Identification with a Linear Index Function

When each \( g_j \) is known, the model (7) reduces to the special case\(^{25}\)

\[
\begin{align*}
    r_j \left( Y \right) &= X_j + U_j & j = 1, \ldots, J. \\
\end{align*}
\]

where (8) becomes

\[
\phi(y|x) = f(r(y) - x) |J(y)|. 
\]

We consider this “linear index model” primarily to address identification of each \( r_j \) given knowledge of the functions \( g_j \) obtained through Theorem 1. However, the linear index model is also of independent interest and has been studied previously by \( ? \). Below we give two theorems demonstrating identification of this model.\(^{26}\) The first shows that when instruments have large support, there is no need for a density restriction. The second demonstrates identification with limited (even arbitrarily small) support. Although the latter requires a condition on the log density, this condition is covered by the genericity result given in the Supplemental Material.

\(^{23}\) Any setting in which \( g(X) \neq \mathbb{R}^J \) violates the requirements of Proposition 2. And if \( g(X) = \mathbb{R}^J \), a log density satisfying the requirements of Proposition 2 but violating Condition M is one whose critical points all lie in flat regions but are sufficiently separated that tangencies can be found somewhere in \( \mathbb{R}^J \) for every \( j = 1, \ldots, J \) and \( u'_j \in \mathbb{R} \).

\(^{24}\) The Supplemental Material provides a genericity result covering Condition M.

\(^{25}\) More formally, for each \( j \), redefine \( X_j = g_j \left( X_j \right) \), then redefine \( g_j \) to be the identity function. All properties required by Assumption 1 are preserved. Note that if one specifies \( g_j \left( X_j \right) = X_j \beta_j \), the normalization (11) implies \( \beta_j = 1 \ \forall j \).

\(^{26}\) Similarly, given identification of \( g \), falsifiable restrictions for the case of a linear index (see Appendix D) imply falsifiable restrictions in the more general model. We also show in Appendix D that, under the verifiable hypothesis of Assumption 2, the model defined by (7) and Assumption 1 is falsifiable.
4.1 Identification without Density Restrictions

When the instruments $X$ have large support (e.g., Matzkin (2008)), there is no need to restrict the log density $\ln f$.

**Theorem 2.** Let Assumption 1 hold and suppose $X = \mathbb{R}^J$. Then in the linear index model, $r$ is identified on $Y$.

**Proof.** Because $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y|\tilde{x}) \, dx = 1$, (26) implies

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

So by (26),

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

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

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

the value of $F_j(r_j(y) - \hat{x}_j)$ is identified for all $x_j \in \mathbb{R}$ and $y \in Y$. By (12), $F_j(r_j(y) - \hat{x}_j) = F_j(0)$. For every $y \in Y$ we can then find the value $\hat{o}(y)$ such that $F_j(r_j(y) - \hat{o}(y)) = F_j(0)$, which reveals $r_j(y) = \hat{o}(y)$. This identifies each function $r_j$ on $Y$.

Note that although $J(y)$ is a functional of $r$, this relationship was not imposed in our proof; rather, the Jacobian determinant was treated as a nuisance parameter to be identified separately. Thus, the definition $J(y) = \det(\partial r(y)/\partial y)$ provides a falsifiable restriction of the model and large support assumption.

**Proposition 4.** If $X = \mathbb{R}^J$, then the joint hypothesis of the linear index model (25) and Assumption 1 is falsifiable.

4.2 Identification with Limited Support

To demonstrate identification when $X$ has limited support, a different approach is needed. In the linear index model (13) and (15) become, respectively,

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

\[27\] The argument used to show Theorem 2 was first used by ? in combination with additional assumptions and arguments to demonstrate identification in a model of differentiated products demand and supply.
and

\[
\frac{\partial \ln \phi(y|x)}{\partial y_k} = \frac{\partial \ln |J(y)|}{\partial y_k} - \sum_j \frac{\partial \ln \phi(y|x)}{\partial x_j} \frac{\partial r_j(y)}{\partial y_k}.
\]  

(29)

We rewrite (29) as

\[ a_k(x,y) = d(x,y)^T b_k(y) \]  

(30)

where we define \( a_k(x,y) = \frac{\partial \ln \phi(y|x)}{\partial y_k}, \) \( d(x,y)^T = \left(1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \ldots, -\frac{\partial \ln \phi(y|x)}{\partial x_J}\right)\), and

\[ b_k(y) = \left(\frac{\partial \ln \phi(y|x)}{\partial y_k}, \frac{\partial r_1(y)}{\partial y_k}, \ldots, \frac{\partial r_J(y)}{\partial y_k}\right)^T. \]

Here \( a_k(x,y) \) and \( d(x,y) \) are observable whereas \( b_k(y) \) involves unknown derivatives of the functions \( r_j \). From (30) it is clear that \( b_k(y) \) is identified if there exist points \( \tilde{x} = (\tilde{x}^0, \ldots, \tilde{x}^J)^T \), with each \( \tilde{x}^j \in \mathbb{X} \), such that the \((J+1) \times (J+1)\) matrix

\[
D(\tilde{x}, y) \equiv \begin{pmatrix}
d(\tilde{x}^0, y)^T \\
\vdots \\
d(\tilde{x}^J, y)^T
\end{pmatrix}
\]  

(31)

has full rank.\(^{28}\) This yields the following observation, obtained previously by ?.

**Lemma 3.** Let Assumption 1 hold. For a given \( y \in \mathcal{Y} \), suppose there exists no nonzero vector \( c = (c_0, c_1, \ldots, c_J)^T \) such that \( d(x,y)^T c = 0 \ \forall x \in \mathcal{X} \). Then in the linear index model, \( \partial r(y)/\partial y_k \) is identified for all \( k \).

This result provides identification of \( \partial r(y)/\partial y_k \) at a point \( y \) when the support of \( X \) covers points \( \tilde{x} \) such that \( D(\tilde{x}, y) \) is invertible. Matzkin (2010) has provided a sufficient condition: that there exist \( \tilde{x} \) such that \( D(\tilde{x}, y) \) is diagonal with nonzero diagonal terms. Using our earlier geometric interpretation, that condition requires the log density to have an appropriate set of critical points and tangencies within the set \( \{r(y) - \mathbb{X}\} \). When \( \mathbb{X} = \mathbb{R}^J \), that is a mild requirement (and can be made slightly milder by requiring only triangular \( D(\tilde{x}, y) \) with nonzero diagonal). However, Theorem 2 shows that with large support we may dispense with all density restrictions. And when \( \mathbb{X} \neq \mathbb{R}^J \), densities having the requisite tangencies and critical points in \( \{r(y) - \mathbb{X}\} \) to obtain a diagonal or triangular \( D(\tilde{x}, y) \) for every \( y \) (or even all \( y \) in a substantial subset of its support) would be quite special.

Of course, most invertible matrices are not diagonal or triangular, suggesting that these sufficient conditions are much stronger than necessary. We can obtain a useful characterization of other sufficient conditions for identification by considering the

\(^{28}\)In particular, let \( A_k(\tilde{x}, y) = \left(a_k(\tilde{x}^0, y) \cdots a_k(\tilde{x}^J, y)\right)^T \) and stack the equations obtained from (30) at each of the points \( \tilde{x}^{(0)}, \ldots, \tilde{x}^{(J)} \), yielding \( A_k(\tilde{x}, y) = D(\tilde{x}, y) b_k(y) \). When \( D(\tilde{x}, y) \) is invertible we obtain \( b_k(y) = D(\tilde{x}, y)^{-1} A_k(\tilde{x}, y) \).
Define the second-derivative matrix
\[ H_\phi(x, y) = \frac{\partial^2 \ln \phi(y|x)}{\partial x \partial x^T} = \begin{pmatrix} \frac{\partial^2 \ln \phi(y|x)}{\partial x_1^2} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_1 \partial x_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln \phi(y|x)}{\partial x_J \partial x_1} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_J^2} \end{pmatrix}. \] (32)

Observe that, fixing a value of \( y \) and \( c = (c_0, c_1, \ldots, c_J)^T \), \( d(x, y)^T c \) is a function of \( x \), with gradient \(-H_\phi(x, y)(c_1, \ldots, c_J)^T \). This leads to the following lemma.

**Lemma 4.** For a nonzero vector \( c = (c_0, c_1, \ldots, c_J)^T \),
\[ d(x, y)^T c = 0 \quad \forall x \in \mathbb{X} \] (33)
if and only if for the nonzero vector \( \tilde{c} = (c_1, \ldots, c_J)^T \)
\[ H_\phi(x, y) \tilde{c} = 0 \quad \forall x \in \mathbb{X}. \] (34)

**Proof.** See Appendix C. \( \square \)

For a given value of \( y \), we can be certain that there is no \( \tilde{c} \) satisfying (34) if \( H_\phi(x, y) \) is nonsingular at some \( x \in \mathbb{X} \). This yields the following result. Note that although Theorem 3 allows for the possibility of identification on a strict subset of \( \mathbb{Y} \), the case \( \mathbb{Y}' = \mathbb{Y} \) is also covered.

**Theorem 3.** Let Assumption 1 hold and let \( \mathbb{Y}' \subset \mathbb{Y} \) be open and connected. Suppose that, for almost all \( y \in \mathbb{Y}' \), \( \partial^2 \ln f(u)/\partial u \partial u^T \) is nonsingular at some \( u \in \{r(y) - \mathbb{X}\} \). Then in the linear index model, \( r \) is identified on \( \mathbb{Y}' \).

**Proof.** From (28) and the definition (32), \( H_\phi(y|x) = \partial^2 \ln f(r(y) - x)/\partial u \partial u^T \). Identification of \( \partial r(y)/\partial y_k \) for all \( k \) and \( y \in \mathbb{Y}' \) then follows from Lemma 4, Lemma 3, and continuity of the derivatives of \( r \). 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 linear (and, thus piecewise continuously differentiable) path in \( \mathbb{Y}' \). With the boundary condition (9), identification of \( r_j(y) \) for all \( y \) and \( j \) then follows from the fundamental theorem of calculus for line integrals. \( \square \)

This result covers many different combinations of restrictions on \( (\mathbb{X}, f) \) sufficient for identification. Given Theorem 2, those of interest permit instruments with limited support. For example, if \( \partial^2 \ln f(u)/\partial u \partial u^T \) is nonsingular almost everywhere,

\footnote{Matzkin (2015, Theorem 4.2) considers a different type of second-derivative condition to show partial identification in a related model. ? combine a completeness condition with a different second-derivative condition to obtain identification in a discrete choice model with more than one instrument per equation.}

\footnote{Using related arguments, Propositions 6 and 7 in Appendix D demonstrate falsifiability of the linear index model.}

\footnote{See, e.g., ?, Theorem 6.63.}
identification of $r$ on $Y$ holds even with arbitrarily small $X$. Nonsingularity of $\partial^2 \ln f(u) / \partial u \partial u^\top$ almost everywhere holds for many standard joint probability distributions; examples of densities that violate this condition are those that are flat (uniform) or log-linear (exponential) on an open set. Of course, nonsingularity almost everywhere is much more than required: for Theorem 3 to apply, it is sufficient that $\partial^2 \ln f(u) / \partial u \partial u^\top$ be nonsingular once in $\{r(y) - X\}$ for each $y \in Y$.$^{32}$ Even when $X$ is a small open ball, this is a fairly modest requirement—one that is satisfied by typical parametric densities on $\mathbb{R}^J$ and covered by the genericity result provided in the Supplemental Material.

In addition, we observe that (28) immediately implies verifiability of the Hessian condition hypothesized in Theorem 3.

Proposition 5. For any $y \in Y$, the following condition is verifiable: $\partial^2 \ln f(u) / \partial u \partial u^\top$ is nonsingular at some $u \in \{r(y) - X\}$.

5 Identification of the Full Model

Together, the results in sections 3 and 4 yield many combinations of sufficient conditions for identification of the full (nonlinear index) model. We summarize many of these combinations in two corollaries. The first follows the prior literature in exploiting instruments with large support, but generalizes Matzkin's (2008) result by allowing Condition M to replace regularity on $\mathbb{R}^J$. The second Corollary offers a more significant step forward. It provides the first result showing identification of the full model without a large support condition.

Corollary 1. Suppose Assumption 1 holds and that $g(X) = \mathbb{R}^J$. If either (a) Condition M holds or (b) $f$ is regular on $\mathbb{R}^J$, then $g$ is identified on $X$, and $r$ is identified on $Y$.

Proof. By Theorem 1, identification of $g$ follows from Propositions 2 (in case (b)) and 3 (in case (a)). Identification of $r$ then follows from Theorem 2. \qed

We view Corollary 2, below, as our most important result. This result shows that identification of the full model can be attained even when instruments vary only over a small open ball.

Corollary 2. Let $Y' \subset Y$ be open and connected, and suppose that Assumption 1 and Condition M hold. If, for almost all $y \in Y'$, $\partial^2 \ln f(u) / \partial u \partial u^\top$ is nonsingular at some $u \in \{r(y) - g(X)\}$, then $g$ is identified on $X$ and $r$ is identified on $Y'$.

---

$^{32}$In some applications it may be reasonable to assume that $U_j$ and $U_k$ are independent for all $k \neq j$. Because $\partial^2 \ln f(u) / \partial u \partial u^\top$ is diagonal under independence, it is then sufficient that $\partial^2 \ln f_j(\hat{u}_j) / \partial u^2_j$ be nonzero for all $j$ at some $\hat{u} \in \{r(y) - X\}$. Other approaches to identification with independent errors have recently been explored by ?.
Proof. By Theorem 1 and Proposition 3, $g$ is identified on $X$. Identification of $r$ on $Y'$ then follows from Theorem 3.

Although Corollary 2 permits instruments with limited support, it requires two conditions on the log density: Condition M and invertibility of the Hessian matrix at some point in the set $\{r(y) - g(X)\}$ for almost all $y \in Y'$. In the Supplemental Material we use a standard notion of genericity to evaluate the strength of these conditions when we allow the instruments $X$ to have arbitrarily small support. We show that when $Y'$ is the pre-image of any bounded open connected subset of $\mathbb{R}^J$, densities simultaneously satisfying these conditions form a dense open subset of all $C^2$ log densities on $\mathbb{R}^J$ possessing a local maximum. This demonstrates one formal sense in which the conditions required by Corollary 2 may be viewed as mild.

6 Conclusion

We have developed new sufficient conditions for identification in a class of nonparametric simultaneous equations models introduced by Matzkin (2008, 2015). These models combine traditional exclusion restrictions with an index restriction linking the roles of unobservables and observables. Our results establish identification of these models under more general and more interpretable conditions than previously recognized. Our most important results are those demonstrating identification without a large support condition. Even when instruments have arbitrarily small support, the density conditions we require for identification hold for familiar parametric families and satisfy a standard (topological) notion of genericity. Instruments with large support allow identification under even weaker density conditions or, in the case of the linear index model, no density restriction at all. Key assumptions required by our results are verifiable, while the maintained assumptions of the model itself are falsifiable.

Together these results demonstrate the robust identifiability of fully simultaneous models satisfying Matzkin’s (2008, 2015) residual index structure. These models cover a range of important applications in economics. Although we have focused exclusively on nonparametric identification, our results provide a more robust foundation for existing (parametric and nonparametric) estimators and may suggest strategies for new estimation and testing approaches. Among other important topics left for future work are (a) a full treatment of identification when instruments have discrete support, and (b) in particular applications, the potential identifiability of specific counterfactual quantities of interest under conditions that relax the assumptions we impose to ensure point identification of the model itself.
Appendices

A  Examples of Simultaneous Economic Models

Here we provide a few examples of simultaneous systems arising in important economic applications, relate the structural forms of these models to their associated residual forms, and describe nonparametric functional form restrictions generating the residual index structure.

Example 1. Consider first a nonparametric version of the classical simultaneous equations model, with the structural equations given by

\[ Y_j = \Gamma_j (Y_{-j}, Z, X_j, U_j) \quad j = 1, \ldots, J, \quad (A.1) \]

where \( Y_{-j} \) denotes \( \{Y_k\}_{k \neq j} \). This system demonstrates full simultaneity in the most transparent form: all \( J \) endogenous variables \( Y \) appear in each of the \( J \) equations. The system \((A.1)\) also reflects the exclusion restrictions that each \( X_j \) appears only in equation \( j \). Thus, for each equation \( j \), \( X_{-j} \) are excluded exogenous variables that may serve as instruments for the included right-hand-side endogenous variables \( Y_{-j} \). Extensive discussion and examples can be found in the theoretical and applied literatures on parametric (typically, linear) simultaneous equations models.

The residual index structure arises by requiring

\[ \Gamma_j (Y_{-j}, Z, X_j, U_j) = \gamma_j (Y_{-j}, Z, \delta_j (Z, X_j, U_j)) \quad \forall j \]

where \( \delta_j (Z, X_j, U_j) = g_j (Z, X_j) + U_j \). The resulting model features nonseparable structural errors but requires them to enter the nonseparable nonparametric function \( \Gamma_j \) through the index \( \delta_j (Z, X_j, U_j) \). If each function \( \gamma_j \) is invertible (e.g., strictly increasing) in \( \delta_j (Z, X_j, U_j) \), then one obtains \((6)\) from the inverted structural equations by letting \( r_j = \gamma_j^{-1} \). Identification of the functions \( r_j \) and \( g_j \) for each \( j \) then implies identification of each \( \Gamma_j \).

Example 2. Although simultaneity often arises when multiple agents interact, single-agent settings involving interrelated choices also give rise to fully simultaneous systems. In addition, the structural equations obtained from the economic model need not take the form \((A.1)\). As an example illustrating both points, consider identification of a production function when firms are subject to shocks to the marginal product of each input. Suppose that a firm’s output is given by \( Q = \Psi (Y, \xi) \), where \( \Psi \) is a concave production function, \( Y \in \mathbb{R}^J_+ \) is a vector of input quantities, and \( \xi \in \mathbb{R}^J \) is a vector of factor-specific productivity shocks affecting the firm.\(^{33}\) The shocks are known by the firm when input levels are chosen, but unobserved to the econometrician. Let \( P \) and

\(^{33}\)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.
W denote exogenous prices of the output and inputs, respectively. The observables (from a population of firms) are \((Q, Y, P, W)\).

Profit-maximizing behavior is characterized by a system of first-order conditions

\[
p \frac{\partial \Psi (y, \epsilon)}{\partial y_j} = w_j \quad j = 1, \ldots, J.
\] (A.2)

The solution(s) to this system of equations define input demand correspondences \(y_j(p, w, \epsilon)\). Here, full simultaneity is reflected by the fact that each \(y_j(p, w, \epsilon)\) depends on the entire vector of shocks \(\epsilon\).

The index structure can be obtained by assuming that, for some unknown function \(\psi_j\) and unknown strictly increasing function \(h_j\),

\[
\frac{\partial \Psi (y, \epsilon)}{\partial y_j} = h_j(\psi_j(y) \epsilon_j).
\]

This restriction combines a form of multiplicative separability with a formalization of the notion that the shocks are factor-specific: \(\epsilon_j\) is the only shock directly affecting the marginal product of input \(j\).\(^{34}\)

The first-order conditions (A.2) then take the form

\[
h_j(\psi_j(y) \epsilon_j) = \frac{w_j}{p}
\]

or, equivalently,

\[
\psi_j(y) \epsilon_j = h_j^{-1}\left(\frac{w_j}{p}\right).
\]

Taking logs, we have

\[
\ln (\psi_j(y)) = g_j\left(\frac{w_j}{p}\right) - \ln (\epsilon_j) \quad j = 1, \ldots, J
\]

where \(g_j = \ln h_j\) is an unknown strictly increasing function. Defining \(X_j = \frac{W_j}{P}\), \(U_j = \ln (E_j)\), and \(r_j = \ln \psi_j\), we then obtain a model of the form (6). Our identification results above then imply identification of the functions \(\psi_j\) and \(g_j\) and, therefore, the realizations of each productivity shock \(E_j\). Since \(Q\) and \(Y\) are observed, this implies identification of the production function \(\Psi\).

Example 3. Example 1 covers the elementary supply and demand model for a single good in a competitive market. Allowing multiple goods (including substitutes or complements) and firms with market power typically leads to demand and “supply” equations with a different form. Let \(P = (P_1, \ldots, P_K)\) denote the prices of goods 1, \ldots, \(K\), with \(Q = (Q_1, \ldots, Q_K)\) denoting their quantities (expressed, e.g., in levels

\(^{34}\)A similar form of additive separability would also lead to the residual index structure.
or shares). Let $V_j \in \mathbb{R}$ and $\xi_j \in \mathbb{R}$ denote, respectively, observed and unobserved demand shifters for good $j$. All other observed demand shifters have been conditioned out, treating them fully flexibly. Let $V = (V_1, \ldots, V_K)$ and $\xi = (\xi_1, \ldots, \xi_K)$. Demand for each good $j$ then takes the form

$$Q_j = D_j (P, V, \xi). \quad (A.3)$$

Observe that each demand function $D_j$ depends on $K$ (endogenous) prices as well as $K$ structural errors $\xi$. To impose the index structure, first define

$$\delta_j = \alpha_j (V_j) + \xi_j$$

where $\alpha_j$ is strictly increasing. Then, letting $\delta = (\delta_1, \ldots, \delta_J)$, suppose that (A.3) can be written

$$Q_j = \sigma_j (P, \delta). \quad (A.4)$$

? derive this structure from an index restriction on a nonparametric random utility model.

On the supply side, let $W_j \in \mathbb{R}$ and $\omega_j \in \mathbb{R}$ denote observed and unobserved cost shifters, respectively (all other observed shifters of costs or markups have been conditioned out, treating these fully flexibly). Assuming single-product firms for simplicity, let each firm $j$ have a strictly increasing marginal cost function

$$c_j (\kappa_j),$$

where for some strictly increasing function $\beta_j$,

$$\kappa_j = \beta_j (W_j) + \omega_j.$$  

Let $\kappa = (\kappa_1, \ldots, \kappa_K)$. Suppose that prices are determined through oligopoly competition, yielding a reduced form$^{35}$

$$P_j = \pi_j (\delta, \kappa) \quad j = 1, \ldots, K. \quad (A.5)$$

Note that here each price $P_j$ depends on all $2K$ structural errors.

? and ? provide conditions ensuring that the system of equations (A.4) and (A.5) can be inverted, yielding a $2K \times 2K$ system

$$\alpha_j (V_j) + \xi_j = \sigma_j^{-1} (Q, P)$$

$$\beta_j (W_j) + \omega_j = \pi_j^{-1} (Q, P).$$

Letting $J = 2K$, $r = (\sigma_1^{-1}, \ldots, \sigma_K^{-1}, \pi_1^{-1}, \ldots, \pi_K^{-1})^T$, $X = (V_1, \ldots, V_K, W_1, \ldots, W_K)^T$, and $U = (\xi_1, \ldots, \xi_K, \omega_1, \ldots, \omega_K)^T$, we obtain the system (6). The primary objects of interest in applications include demand derivatives (elasticities) and firms’ marginal

---

$^{35}$? show that such a reduced form arises under standard models of oligopoly supply.
costs, as these allow construction of a wide range of counterfactual predictions. Identification of all $\alpha_j$, $\beta_j$, $\sigma_j^{-1}$ and $\pi_j^{-1}$ immediately implies identification of all $\pi_j$ and $\sigma_j$, and thus all demand derivatives. Specifying the extensive form of oligopoly competition then typically yields a mapping from prices, quantities, and the demand functions $\sigma_j$ to marginal costs (see ?), yielding identification of marginal costs as well.

B Proof of Proposition 3

In this appendix we show that, given Assumption 1, Condition M implies rectangle regularity (Assumption 2). Along the way we establish additional (weaker) sufficient conditions for rectangle regularity. We let $B(u, \epsilon)$ denote an $\epsilon$-ball around a point $u \in \mathbb{R}^J$. For $c \in \mathbb{R}$ and $S \subset \mathbb{R}^J$, we let $A(c; S)$ denote the upper contour set $\{u \in S : \ln f(u) \geq c\}$. We begin with a new condition and new definition.

Condition M'. There exists $c \in \mathbb{R}$ and a compact set $S \subset \mathbb{R}^J$ with nonempty interior such that (i) $A(c; S) \subset \text{int}(S)$, and (ii) the restriction of $\ln f$ to $A(c; S)$ attains a maximum $c > c$ at its unique critical point.

Definition 3. $\ln f$ satisfies local rectangle regularity if it possesses a critical point $u^*$ such that for all $\epsilon > 0$, $\ln f$ is regular on a rectangle $R(u^*, \epsilon)$ such that $u^* \in R(u^*, \epsilon) \subset B(u^*, \epsilon)$.

Condition M' requires a local maximum point $u^*$ such that if we “zoom in” (first to a compact set $S$ and then to some upper contour set of $\ln f$ on the restricted domain $S$) $u^*$ is the only critical point “in sight.” Note that Condition M' requires no second derivatives of $\ln f$. Local rectangle regularity requires that, around some critical point $u^*$, there exist arbitrarily small rectangles on which $\ln f$ is regular. Below we show that Condition M $\implies$ Condition M' $\implies$ local rectangle regularity $\implies$ rectangle regularity. The first and last steps are relatively straightforward.

Lemma 5. Condition M implies Condition M'.

Proof. Let $u^*$ be a point at which $\ln f$ has a nondegenerate local max, and let $\bar{c} = \ln f(u^*)$. By the Morse lemma (e.g., ?, Corollary 2.18.), a nondegenerate critical point is an isolated critical point. So there exists $\epsilon > 0$ such that on the open ball $B(u^*, \epsilon)$, $u^*$ is both the only critical point and a strict maximum. Let $S$ be a compact subset of $B(u^*, \epsilon)$ with $u^*$ in its interior. Because $u^*$ is the only critical point of $\ln f$ on $S$ and maximizes $\ln f$ on $S$, we need only show that there exists $c < \bar{c}$ such that the upper contour set $A(c; S)$ lies in the interior of $S$. Continuity of $\ln f$ implies that $A(c; S)$ is upper hemicontinuous in $c$. So because $u^*$ is the only point in $A(\bar{c}; S)$

36 Take any compact $\Omega \subset \mathbb{R}^J$ and continuous $h : \Omega \to \mathbb{R}$ with upper contour sets $A(c) = \{u \in \Omega : h(u) \geq c\}$. Since $\Omega$ is compact and $h$ is continuous, $A$ is compact-valued. Take any $\hat{c} \in \mathbb{R}$ and a sequence $c^n$ such that $c^n \to \hat{c}$. Let $u^n$ be a sequence such that $u^n \in A(c^n) \forall n$ and $u^n \to \hat{u}$.
and lies in the interior of $S$ by construction, we obtain $A(\zeta; S) \subset \text{int}(S)$ by setting $\zeta = \bar{c} - \delta$ for some sufficiently small $\delta > 0$. \hfill \Box$

**Lemma 6.** Local rectangle regularity implies Assumption 2 (rectangle regularity).

*Proof.* Let $u^*$ denote the critical point referenced in Definition 3. Given any $x \in X$, let $u^* (x) = u^*$ and let $\overline{X}(x) = x_j \left( \underline{x}_j (x), \overline{x}_j (x) \right)$ be a rectangle such that $x \in \overline{X}(x) \subset X$. Define $y^*(x)$ by (17) and let $\overline{U}(x) = x_j \left( \underline{u}_j (x), \overline{u}_j (x) \right)$ where, for $j = 1, \ldots, J$,

\[
\overline{u}_j (x) = u^*_j (x) + g_j (x_j) - g_j \left( \underline{x}_j (x) \right) \\
\underline{u}_j (x) = u^*_j (x) + g_j (x_j) - g_j \left( \overline{x}_j (x) \right).
\]

Local rectangle regularity guarantees that $\ln f$ is regular on some rectangle

\[
\mathcal{U} (x) = x_j \left( \underline{u}_j (x), \overline{u}_j (x) \right) \subset \overline{U} (x)
\]

such that $u^*(x) \in \mathcal{U}(x)$. If we now let $X(x) = x_j \left( \underline{x}_j (x), \overline{x}_j (x) \right)$, where each $\underline{x}_j (x)$ and $\overline{x}_j (x)$ is defined by (18), then by construction $X(x) \subset \overline{X}(x) \subset X$ and $\mathcal{U}(x)$ satisfies (16).

We now show that Condition $M'$ implies local rectangle regularity. We will make use of the following two results.

**Lemma 7.** Under Condition $M'$, the restriction of $A(\cdot; S)$ to $(\zeta, \bar{c}]$ is a nonempty-valued, compact-valued, continuous (upper and lower hemicontinuous) correspondence.

*Proof.* Letting $u^*$ denote the critical point referenced in Condition $M'$, we have $u^* \in A(c; S)$ for all $c \leq \bar{c}$. Because $S$ is compact and $\ln f$ is continuous, $A(c; S)$ is compact for all $c$. Upper hemicontinuity follows from continuity of $\ln f$ (see footnote 36). To show lower hemicontinuity,37 take any $\hat{c} \in (\zeta, \bar{c}]$, any $\hat{u} \in A(\hat{c}; S)$, and any sequence $c^n$ in $(\zeta, \bar{c}]$ such that $c^n \to \hat{c}$. If $\hat{u} = u^*$ then $\hat{u} \in A(c; S)$ for all $c \leq \bar{c}$, so with the constant sequence $u^n = \hat{u}$ we have $u^n \in A(c^n; S)$ for all $n$ and $u^n \to \hat{u}$. So now suppose that $\hat{u} \neq u^*$. Letting $\| \cdot \|$ denote the Euclidean norm, define a sequence $u^n$ by

\[
u^n = \arg \min_{u \in A(c^n; S)} \| u - \hat{u} \| \quad (B.1)
\]

so that $u^n \in A(c^n; S)$ by construction. We now show $u^n \to \hat{u}$. Take arbitrary $\epsilon > 0$. Because (i) $\ln f$ is continuous, (ii) $\hat{u} \in \text{int}(S)$, and (iii) $\ln f (\hat{u}) > \zeta$, for all sufficiently small $\delta > 0$ we have $B(\hat{u}, \delta) \subset A(\zeta; S)$. Thus, $\{ B(\hat{u}, \epsilon) \cap A(\zeta; S) \}$

If $\hat{u} \notin A(\hat{c})$ then, because $\hat{u}$ must lie in $\Omega$, we must have $h(\hat{u}) < \hat{c}$. But then continuity of $h$ would require $h(u^n) < c^n$ for $n$ sufficiently large, contradicting the fact that $u^n \in A(c^n)$ \forall n. So $\hat{u} \in A(\hat{c})$.

37This argument is similar to that used to prove Proposition 2 in \(?\).
contains an open set \( O \ni \hat{u} \). If \( \ln f (u) \leq \ln f (\hat{u}) \) for all \( u \in O \), \( \hat{u} \) would be a critical point of the restriction of \( \ln f \) to \( A (c; S) \); since \( \hat{u} \) is not a critical point, there must exist \( u^* \in \{ \mathcal{B} (\hat{u}, \epsilon) \cap A (c; S) \} \) such that \( \ln f (u^*) > \ln f (\hat{u}) \). Because \( \ln f (\hat{u}) \geq \hat{c} \), this implies \( \ln f (u^*) > \hat{c} \). Recalling that \( c^n \to \hat{c} \), for \( n \) sufficiently large we then have \( \ln f (u^*) > c^n \) and, therefore, \( u^* \in A (c^n; S) \). So, recalling (B.1), for \( n \) sufficiently large we have \( \| u^n - \hat{u} \| \leq \| u^* - \hat{u} \| < \epsilon \).

Lemma 8. Under Condition M', for all \( c \in (c, \bar{c}) \), \( A (c; S) \) is connected and has nonempty interior.

Proof. Let \( u^* \) denote the critical point referenced in Condition M' and take any \( c \in (c, \bar{c}) \). To show that \( A (c; S) \) has nonempty interior, observe that \( u^* \in A (c; S) \subset A (\underline{c}; S) \subset \text{int}(S) \). Because \( \ln f \) is continuous and \( A (c; S) \subset \text{int}(S) \), \( \ln f (u) = c \) for all \( u \) on the boundary of \( A (c; S) \). Since \( \ln f (u^*) = \bar{c} \), \( u^* \) must be an interior point.

To show that \( A (c; S) \) is connected, suppose (to the contrary) that the upper contour set \( A (c; S) \) is the union of disjoint nonempty open (relative to \( A (c; S) \)) sets \( A^1 \) and \( A^2 \). Without loss let \( u^* \) lie in \( A^1 \). By continuity of \( \ln f \), \( A (c; S) \) is a compact subset of \( \mathbb{R}^J \). This requires that \( A^2 \) be a compact subset of \( \mathbb{R}^J \) as well. \(^{38}\) The restriction of \( \ln f \) to \( A^2 \) must therefore attain a maximum at some point(s) \( u^{**} \), which must be on the interior of \( A (c; S) \). Any such \( u^{**} \) would be another critical point of \( \ln f \) on \( A (c; S) \), contradicting Condition M'.

The following result, whose construction is illustrated by Figure 3, then completes the proof of Proposition 3.

Lemma 9. Condition M' implies local rectangle regularity.

Proof. Let \( S, \underline{c} \) and \( \bar{c} \) be as defined in Condition M', and let \( u^* \) denote the critical point referenced in Condition M'. We first show that, for any \( \epsilon > 0 \), there exists \( c^0 \in (\underline{c}, \bar{c}) \) such that \( A (c^0; S) \subset U \subset \mathcal{B} (u^*, \epsilon) \) for some rectangle \( U \). We saw in the proof of Lemma 8 that \( u^* \in \text{int} (A (c; S)) \) for any \( c \in (\underline{c}, \bar{c}) \). By Lemma 7, \( \max_{u \in A (c; S)} u_j \) and \( \min_{u \in A (c; S)} u_j \) are continuous in \( c \in (\underline{c}, \bar{c}) \), implying continuity of the function

\[
H (c) = \max_j \frac{\max_{u^+, u^- \in A (c; S)} u_j^+ - u_j^-}{\epsilon_j} \quad c \in (\underline{c}, \bar{c})
\]

So, because \( H (\bar{c}) = 0 \), given any \( \epsilon > 0 \) there must exist \( c^0 \in (\underline{c}, \bar{c}) \) such that the rectangle (Lemma 8 ensures that each interval in the Cartesian product is nonempty)

\[
U = \times_j \left( \min_{u^- \in A (c^0; S)} u_j^-, \max_{u^+ \in A (c^0; S)} u_j^+ \right)
\]

\(^{38}\)Bounded is immediate. Suppose \( A^2 \) is not closed: let \( u \notin A^2 \) be a limit point of a sequence in \( A^2 \). Since \( A (c; S) \) is closed, it must then be that \( u \in A^1 \). But since \( u \) was a limit point of a sequence in \( A^2 \), for all \( \epsilon > 0 \) there exists \( \hat{u} \in \{ \mathcal{B} (u, \epsilon) \cap A (c; S) \} \) such that \( \hat{u} \in A^2 \). Because \( A^1 \) and \( A^2 \) are disjoint, this requires \( \hat{u} \notin A^1 \), contradicting openness of \( A^1 \) relative to \( A (c; S) \).
Figure 3: Curves show level sets of a bivariate log density in a region of its support. The shaded area is a compact set $S$. The darker subset of $S$ is an upper contour set $A(c; S)$ of the restriction of $\ln f$ to $S$. The point $u^*$ is a local max and the only critical point of $\ln f$ on $A(c; S)$. The rectangle $U$ is defined by tangencies to the upper contour set $A(c_0; S)$ for some $c_0 \in (c, \ln f(u^*))$. Given any $\epsilon > 0$, we obtain $U \in B(u^*, \epsilon)$ by setting $c^0$ sufficiently close to $\ln f(u^*)$. Thus, $A(c_0; S) \subset U \subset B(u^*, \epsilon)$. To complete the proof, we show that $\ln f$ is regular on $U$. By construction $u^* \in A(c^0; S) \subset U$. Now take arbitrary $j$ and any $u_j \neq u_j^*$ such that $(u_j, u_{-j}) \in U$ for some $u_{-j}$. By Lemma 8 and the definition of $U$, there must also exist $\tilde{u}_{-j}$ such that $(u_j, \tilde{u}_{-j}) \in A(c^0; S)$. Let $\hat{u}(j, u_j)$ solve

$$\max_{\hat{u} \in A(c^0; S): u_j = u_j} \ln f(\hat{u}).$$

This solution must lie in $A(c^0; S) \subset U$ and satisfy $\partial \ln f(\hat{u}(j, u_j)) / \partial u_k = 0$ for all $k \neq j$. Since $u_j \neq u_j^*$, we have $\partial \ln f(\hat{u}(j, u_j)) / \partial u_j \neq 0$.

\[\square\]

C Other Proofs Omitted from the Text

Proof of Lemma 1. By (7), part (iii) of Assumption 1 immediately implies (a) and (b). Parts (iii) and (iv) then imply that $r$ has a continuous inverse $r^{-1} : \mathbb{R}^J \to \mathbb{R}^J$. Connectedness of $Y$ then 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
Proof of Proposition 1. Fix an arbitrary $x \in \mathbb{X}$. By (13),

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

(C.1)

if and only if, for $u^*(x) = r(y^*(x)) - g(x)$, $\partial \ln f(u^*(x))/\partial u_j = 0$. Thus, existence of the critical point $u^*(x)$ in Assumption 2 is equivalent to existence of $y^*(x) \in \mathbb{Y}$ such that (C.1) holds. This is verifiable. Now observe that for $x \in \mathcal{X}(x)$ and $\mathcal{U}(x)$ as defined in Assumption 2,

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

Thus, Assumption 2 holds if and only if there is a rectangle $\mathcal{X}(x) = \times_j \left( \bar{x}_j(x), \underline{x}_j(x) \right)$, with $x \in \mathcal{X}(x) \subset \mathbb{X}$, such that for all $j$ and almost all $x'_j \in \left( \bar{x}_j(x), \underline{x}_j(x) \right)$ there exists $\hat{x} \left( j, x'_j \right) \in \mathcal{X}(x)$ satisfying

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

Satisfaction of these conditions is observable. \qed

Proof of Lemma 4. Recall that $d(x, y)^\top = \left( 1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \ldots, -\frac{\partial \ln \phi(y|x)}{\partial x_J} \right)$. Suppose first that (33) holds for nonzero $c = (c_0, c_1, \ldots, c_J)^\top$. Differentiating (33) with respect to $x$ yields (34), with $\tilde{c} = (c_1, \ldots, c_J)^\top$. 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 $d(x, y)^\top c = 0$, we must have $c_j \neq 0$ for some $j > 0$. Thus (34) must hold for some nonzero $\tilde{c}$. Now suppose (34) holds for nonzero $\tilde{c} = (c_1, \ldots, c_J)^\top$. Take an arbitrary point $x^0$ and let $c_0 = \sum_{j=1}^J \frac{\partial \ln \phi(y|x^0)}{\partial x_j} c_j$ so that, for $c = (c_0, c_1, \ldots, c_J)^\top$, $d(x^0, y)^\top c = 0$ by construction. Since the first component of $d(x, y)$ equals 1 for all $(x, y)$ and $\mathbb{X}$ is an open connected subset of $\mathbb{R}^J$, (34) implies that $\frac{\partial}{\partial x_j} [d(x, y)^\top c] = 0$ for all $j$ and every $x \in \mathbb{X}$. Thus (33) holds for some nonzero $c$. \qed

D Further Falsifiability Results

Here we provide additional falsifiability results—two for the linear index model and one for the full model. First, recalling (30) and the definition

$$b_k(y) = \left( \frac{\partial \ln \phi_j(y)|y_k}{\partial y_k}, \frac{\partial r_1(y)|y_k}{\partial y_k}, \ldots, \frac{\partial r_J(y)|y_k}{\partial y_k} \right)^\top,$$
observe that Theorem 3 shows, for all $y \in Y$ and $k = 1, \ldots, J$, separate identification of the derivatives $\partial r(y)/\partial y_k$ and the derivatives $\partial \ln |J(y)|/\partial y_k$. However, knowledge of the former also implies knowledge of the latter. So under the hypotheses of Theorem 3 we have the falsifiable restrictions

$$\frac{\partial}{\partial y_k} \ln \left| \det \begin{pmatrix} \frac{\partial r_1(y)}{\partial y_1} & \cdots & \frac{\partial r_1(y)}{\partial y_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_J(y)}{\partial y_1} & \cdots & \frac{\partial r_J(y)}{\partial y_J} \end{pmatrix} \right| = \frac{\partial}{\partial y_k} \ln |J(y)| \quad \forall k. \quad (D.1)$$

**Proposition 6.** Under the hypotheses of Theorem 3, the model defined by (25) and Assumption 1 is falsifiable.

Suppose now that at some value of $y$ there exist two sets of points satisfying the rank condition of Lemma 3—a verifiable condition. Then the maintained assumptions of the linear index model are falsifiable.

**Proposition 7.** Suppose that, for some $y \in Y$, $X$ contains two sets of points, $\tilde{x} = (\tilde{x}^0, \ldots, \tilde{x}^J)^T$ and $\tilde{\tilde{x}} = (\tilde{\tilde{x}}^0, \ldots, \tilde{\tilde{x}}^J)^T$, such that (i) $\tilde{x} \neq \tilde{\tilde{x}}$ and (ii) $D(\tilde{x}, y)$ and $D(\tilde{\tilde{x}}, y)$ have full rank. Then the model defined by (25) and Assumption 1 is falsifiable.

**Proof.** By Lemma 3, $\partial r(y)/\partial y_k$ is identified for all $k$ using the derivatives of $\phi(y|x)$ at points $x$ in $\tilde{x}$ (only) or in $\tilde{\tilde{x}}$ (only). Letting $\partial r(y)/\partial y_k [\tilde{x}]$ and $\partial r(y)/\partial y_k [\tilde{\tilde{x}}]$ denote the implied values of $\partial r(y)/\partial y_k$, we obtain the verifiable restrictions $\partial r(y)/\partial y_k [\tilde{x}] = \partial r(y)/\partial y_k [\tilde{\tilde{x}}]$ for all $k$. $\square$

As noted in the text, all falsifiability results for the linear index model extend to the full model when sufficient conditions for identification of $g$ hold. The following provides an additional falsifiable restriction of the full model.

**Proposition 8.** The joint hypothesis of (7), Assumption 1, and Assumption 2, is falsifiable.

**Proof.** The proof of Lemma 2 began with an arbitrary $x \in X$ and the associated $y^* (x)$ defined by (17). It was then demonstrated that for some open rectangle $\mathcal{X}(x) \ni x$ the ratios

$$\frac{\partial g_j(x'_j)}{\partial x_j} \quad \text{and} \quad \frac{\partial g_j(x^0_j)}{\partial x_j}$$

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

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

denote the identified value of $\frac{\partial g_j(x'_j)}{\partial x_j}$. Now take any point $\tilde{x} \in \mathcal{X}(x) \setminus x$ and repeat the argument, replacing $y^* (x)$ with the point $y^{**} (\tilde{x})$ such that (assuming the model...
is correctly specified) \( r(y^* (\tilde{x})) = g(\tilde{x}) + u^* \) where \( \partial \ln f (u^*) / \partial u_j = 0 \forall j \) and \( \ln f \) 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

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

for all \( j = 1, \ldots, J \), all \( x^0 \in \mathcal{X}(\tilde{x}) \setminus \tilde{x} \) and all \( x' \in \mathcal{X}(\tilde{x}) \setminus \tilde{x} \). 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 falsifiable restriction

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

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