7-1-2015

Identification of Nonparametric Simultaneous Equations Models with a Residual Index Structure

Steven T. Berry

Philip A. Haile

Recommended Citation

IDENTIFICATION OF NONPARAMETRIC SIMULTANEOUS EQUATIONS MODELS WITH A RESIDUAL INDEX STRUCTURE

By

Steven T. Berry and Philip A. Haile

July 2015
Revised December 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2008R

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

December 14, 2016

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 of these results demonstrate identification even when instruments have limited variation. A genericity result demonstrates a formal sense in which the associated density conditions may be viewed as mild, even when instruments vary only over a small open ball.

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,[1] 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

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

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

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

Identification of such models was considered by Brown (1983), Roehrig (1988), Brown and Matzkin (1998), and Brown and Wegkamp (2002). However, 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.

Completeness conditions (Lehmann and Scheffe (1950, 1955)) offer one possible approach, and in Berry and Haile (2014) we showed how identification arguments in Newey and Powell (2003) or Chernozhukov and Hansen (2005) can be adapted to an example of the class of models considered below.[3] 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.

---

[1]See, e.g., Koopmans (1950), Hood and Koopmans (1953), and Fisher (1966).

[2]See, e.g., Palais (1959), Gale and Nikaido (1965), and Berry, Gandhi, and Haile (2013) for conditions ensuring invertibility in different contexts.

[3]Chiappori and Komunjer (2009b) show identification in a related model by combining completeness conditions with arguments using the classic change of variables approach.

A high-level assumption like completeness implicitly places further restrictions on the model, although the nature of these restrictions is typically unclear.[4]

Much recent work has focused on triangular (recursive) systems of equations (e.g., Chesher (2003), Imbens and Newey (2009), Torgovitzky (2015)). A two-equation version of the triangular model takes the form

$$Y_1 = m_1(Y_2, Z, U_1) \tag{3}$$
$$Y_2 = m_2(Z, X, U_2) \tag{4}$$

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. Blundell and Matzkin (2014) 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 Matzkin (2008). Matzkin considered a model of the form

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

where $\delta = (\delta_1 (Z, X_1, U_1), \ldots, \delta_J (Z, X_J, U_J))^\mathsf{T}$ is a vector of indices

$$\delta_j (Z, X_j, U_j) = g_j (Z, X_j) + U_j, \tag{5}$$

and each $g_j (Z, X_j)$ is strictly increasing in X_j. Here $X = (X_1, \ldots, X_J)^\mathsf{T} \in \mathbb{R}^J$ 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 δ), the analog of (2) is $\delta_j (Z, X_j, U_j) = r_j (Y, Z)$, $j = 1, \ldots, J$, or equivalently,[6]

$$r_j (Y, Z) = g_j (Z, X_j) + U_j \qquad j = 1, \ldots, J. \tag{6}$$

---

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

[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

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.[7] This was, to our knowledge, the first result demonstrating identification in a fully simultaneous nonparametric model with nonseparable errors. Matzkin (2015) 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.[8]

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 mild local density restriction. 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 spe-

---

system. The semiparametric transformation model (e.g., Horowitz (1996)) 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 $Z\beta$ with $g(Z, X)$, (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. Chiappori and Komunjer (2009a) considers a nonparametric version of the single-equation transformation model. See also Berry and Haile (2009).

[7]Matzkin (2008) used a new characterization of observational equivalence to show 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).

[8]Matzkin (2015) also provides conditions for identification of certain features in models that partially relax the residual index structure. Chesher and Rosen (2015) consider a nonparametric framework permitting simultaneity, providing a characterization of sharp identified sets under various restrictions. Masten (2015) 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.

cial case and provide new sufficient conditions for identification. By combining these results (section 5), one obtains identification of the full (nonlinear index) model (6) under a variety of support and density conditions.

At one extreme, we show that the model is identified under large support conditions without any restriction on the joint density of unobservables. Perhaps the most important result, however, is given by Corollary 3 in section 5, which allows instruments that vary only over a small open ball. Given the maintained technical conditions of the model (Assumption 1), this result shows that identification holds under two relatively mild conditions on the log density of unobservables. The first, "Condition M," requires the log density of U to have a nondegenerate local maximum. The second requires the Hessian of that log density to be invertible at a sufficiently rich set of (possibly isolated) points; a sufficient condition for this restriction is given by "Condition H" in section 6. We argue that these are relatively mild density restrictions, satisfied (for example) by typical parametric densities on $\mathbb{R}^J$. Formalizing this notion, section 6 establishes that densities satisfying Conditions M and H are generic among all C^2 log densities that have a local maximum somewhere on $\mathbb{R}^J$.[9] Thus, under mild (generic) density conditions, the model is identified even when instruments have arbitrarily small support.

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

$$r(Y) = g(X) + U, \tag{7}$$

where $r(Y) = (r_1(Y), \ldots, r_J(Y))^\mathsf{T}$ and $g(X) = (g_1(X_1), \ldots, g_J(X_J))^\mathsf{T}$. Let $\mathbb{X} = \mathrm{int}(\mathrm{supp}(X))$, and let $\mathbb{Y}$ denote the pre-image of $\mathrm{supp}(g(X) + U)$ under r. The following describes the maintained assumptions on the model, following Matzkin (2008).[10]

---

[9]These generic log densities are sufficient to identify the functions $(g_1, \ldots, g_J)$ on the support of X and to identify the functions $(r_1, \ldots, r_J)$ on the pre-image of any rectangle $\times_j \left(\underline{u}_j, \overline{u}_j\right)$ in $\mathbb{R}^J$.

[10]We 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 (31) 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 Appendix F).

Assumption 1. (i) $\mathbb{X}$ 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 $\mathtt{J}(y) = \det(\partial r(y)/\partial y) \ \forall y \in \mathbb{Y}$.

Part (i) rules out instruments with discrete or disconnected support.[11] 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., Koopmans, Rubin, and Leipnik (1950)), relating the joint density of observables to that of the structural errors. In particular, letting $\phi(\cdot|X)$ denote the joint density of Y conditional on X, we have

$$\phi(y|x) = f(r(y) - g(x))\,|\mathtt{J}(y)|. \tag{8}$$

In addition, we have the following lemma.

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. 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.[12] 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 arbitrary point $\dot{r} = (\dot{r}_1, \dots, \dot{r}_J) \in \mathbb{R}^J$ we set

$$r_j(\dot{y}) = \dot{r}_j \quad \forall j. \tag{9}$$

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 κ_j while U_j were replaced by $U_j - \kappa_j$, we take an an arbitrary point $\dot{x} \in \mathbb{X}$ and set

$$g_j(\dot{x}_j) = \dot{x}_j \quad \forall j. \tag{10}$$

---

[11]Appendix G provides an initial exploration of identification when instruments are discrete.

[12]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 Matzkin (2008), 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.

Given (9), this fixes the location of each U_j, but we must still choose its scale.[13] 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 κ_j, we normalize the scale of each U_j by setting

$$\frac{\partial g_j\left(\dot{x}_j\right)}{\partial x_j} = 1 \quad \forall j. \tag{11}$$

Finally, given (9) and (10), a convenient choice of $\dot{r}$ sets each $\dot{r}_j = \dot{x}_j$, so that

$$r_j\left(\dot{y}\right) - g_j\left(\dot{x}_j\right) = 0 \quad \forall j. \tag{12}$$

2.3 Identifiability, Verifiability, and Falsifiability

Before proceeding, we must define some key terminology. Following Hurwicz (1950) and Koopmans and Reiersol (1950), 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 $\mathfrak{S}$ denote the set of all structures. The true structure is denoted $S_0 \in \mathfrak{S}$. A *hypothesis* is any nonempty subset of $\mathfrak{S}$. A hypothesis $\mathcal{H}$ is true (satisfied) if $S_0 \in \mathcal{H}$.[14]

A *structural feature* $\theta\left(S_0\right)$ is a functional of the true structure S_0. A structural feature $\theta\left(S_0\right)$ is *identified* (or *identifiable*) under the hypothesis $\mathcal{H}$ if $\theta\left(S_0\right)$ is uniquely determined within the set $\{\theta\left(S\right) : S \in \mathcal{H}\}$ by the joint distribution of observables. The primary structural features of interest in our setting are the functions r and g.[15] 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 $\mathcal{H} \subset \mathcal{M}$ is *verifiable* if the indicator $1\{S_0 \in \mathcal{H}\}$ is identified under $\mathcal{M}$.[16] Thus, when a hypothesis is verifiable, its satisfaction or failure is an identified feature.[17]

---

[13]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.

[14]Hurwicz (1950) and Koopmans and Reiersol (1950) call any strict subset of $\mathfrak{S}$ 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.

[15]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 Hurwicz (1950), 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 Berry and Haile (2014) for some results in the case of differentiated products supply and demand.

[16]We use the symbol $\subset$ for all (proper or not) subset relationships.

[17]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.

A weaker and more familiar notion is that of falsifiability. Let $\mathcal{P}_{\mathcal{H}}$ denote the set of probability distributions (for the observables) generated by structures in $\mathcal{H}$. Given a maintained hypothesis $\mathcal{M}$, $\mathcal{H} \subset \mathcal{M}$ is *falsifiable* if $\mathcal{P}_{\mathcal{H}} \neq \mathcal{P}_{\mathcal{M}}$. 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.

3 Identification of the Index Functions

We begin by considering identification of the index functions g_j. Taking logs in (8) and differentiating yields[18]

$$\frac{\partial \ln \phi(y|x)}{\partial x_j} = -\frac{\partial \ln f(r(y) - g(x))}{\partial u_j} \frac{\partial g_j(x_j)}{\partial x_j} \tag{13}$$

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = \sum_j \frac{\partial \ln f(r(y) - g(x))}{\partial u_j} \frac{\partial r_j(y)}{\partial y_k} + \frac{\partial \ln |\mathtt{J}(y)|}{\partial y_k}. \tag{14}$$

Together (13) and (14) imply

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = -\sum_j \frac{\partial \ln \phi(y|x)}{\partial x_j} \frac{\partial r_j(y)/\partial y_k}{\partial g_j(x_j)/\partial x_j} + \frac{\partial \ln |\mathtt{J}(y)|}{\partial y_k}. \tag{15}$$

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 first introduce a property of $(\mathbb{X}, f, g)$ that we call *rectangle regularity*. 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.

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.

---

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

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 $\mathcal{U} \equiv \times_{j=1}^{J} \left(\underline{u}_j, \overline{u}_j \right)$, $\ln f$ 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 $u'_j \in \left(\underline{u}_j, \overline{u}_j \right)$, and some $\hat{u} \left(j, u'_j \right) = \left(\hat{u}_1 \left(j, u'_j \right), \dots, \hat{u}_J \left(j, u'_j \right) \right) \in \mathcal{U}$,

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

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

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 $\left\{ u \in \mathbb{R}^J : u_j = u'_j \right\}$. 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 $\ln f$ are connected and smooth, and represent 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, $\mathcal{U} = (\underline{u}_1, \overline{u}_1) \times (\underline{u}_2, \overline{u}_2)$, is defined by tangencies to a single level set. For each $u'_1 \in (\underline{u}_1, \overline{u}_1)$, the point $\hat{u}_2(1, u'_1)$ is the value of U_2 at a tangency between the vertical line $U_1 = u'_1$ and a level set of $\ln f$ closer to u^* than that defining $\mathcal{U}$.

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

Assumption 2 ("Rectangle Regularity"). For all $x \in \mathbb{X}$ there exists a rectangle $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x) \right) \subset \mathbb{X}$ containing x such that for (i) some $u^*(x)$ such that $\partial \ln f \left(u^*(x) \right) / \partial u_j = 0$ for all j and (ii) $\underline{u}_j(x)$ and $\overline{u}_j(x)$ defined by

$$\begin{aligned}
\underline{u}_j(x) &= u_j^*(x) + g_j(x_j) - g_j(\overline{x}_j(x)) \\
\overline{u}_j(x) &= u_j^*(x) + g_j(x_j) - g_j(\underline{x}_j(x)),
\end{aligned} \tag{16}$$

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

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 $\times_j \left(\underline{u}_j(x), \overline{u}_j(x) \right) \ni u^*(x)$ denote a rectangle on which $\ln f$ is regular. Define $y^*(x)$ by

$$r(y^*(x)) = g(x) + u^*(x), \tag{17}$$

and define $\underline{x}(x)$ and $\overline{x}(x)$ by

$$r_j(y^*(x)) = g_j(\underline{x}_j) + \overline{u}_j(x) = g_j(\overline{x}_j) + \underline{u}_j(x) \ \forall j. \tag{18}$$

8

Figure 1: The solid curves are the level sets of a bivariate log density in a region of its support. The point u^* is a local maximum. For each $u_1' \in (\underline{u}_1, \overline{u}_1)$ 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 $\mathcal{U} = \times_j (\underline{u}_j, \overline{u}_j)$.
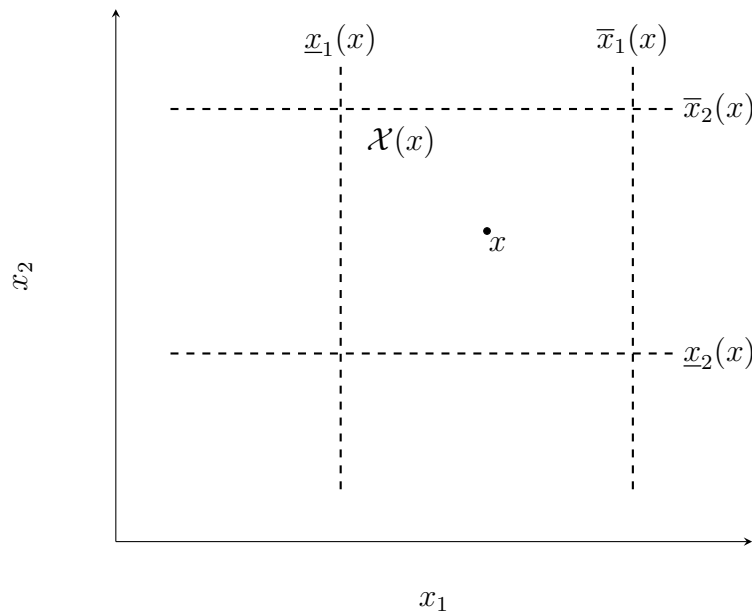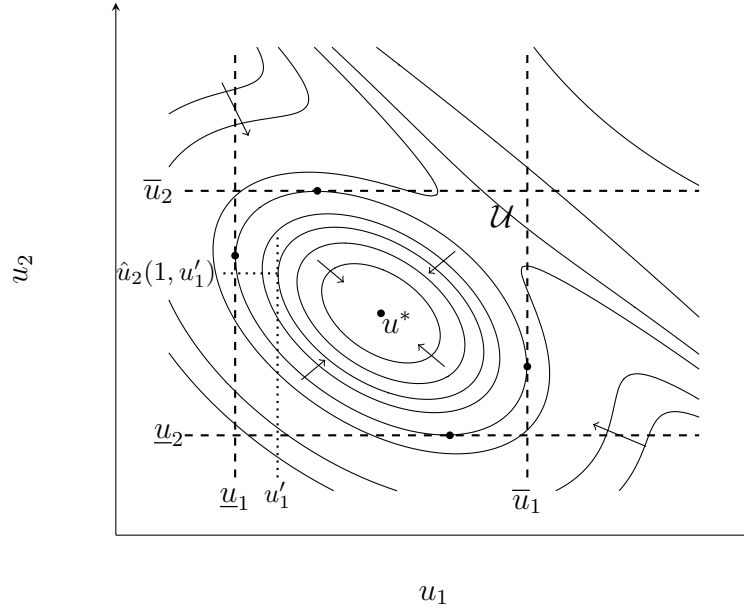
Figure 2: For arbitrary $x \in \mathbb{X}$, the rectangle $\mathcal{U}$ in Figure 1 is mapped to a rectangle $\mathcal{X}(x)$ using (17) and (18), thereby satisfying (16).

9

Figure 2 illustrates. Assumption 2 is satisfied if, for every x, there exist $u^*(x)$ and $\times_j \left(\underline{u}_j(x), \overline{u}_j(x)\right)$ such that the resulting rectangle $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x)\right)$ 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 regularity is guaranteed to exist as long as $\ln f$ is regular on some rectangle in $\mathbb{R}^J$ whose dimensions are 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.

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 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 (11), we can integrate up to the functions g_j, using (10) as boundary conditions.

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

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

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

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

By (13), the point $y^*(x)$ is identified, so the only unknowns in (20) are the ratios $\frac{\partial r_j(y^*(x))/\partial y_k}{\partial g_j(x_j')/\partial x_j}$. 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^* on which $\ln f$ is regular.

Lemma 2. *Let Assumptions 1 and 2 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 $x \in \mathbb{X}$. Let u^* and $\mathcal{U} = \times_j \left(\underline{u}_j, \overline{u}_j\right)$ be as defined in Assumption 2, and let y^* be as defined by (17).[19] By Assumption 2 there exists $\mathcal{X} = \times_i \left(\underline{x}_i, \overline{x}_i\right) \subset \mathbb{X}$ (with $x \in \mathcal{X}$) such that (18) holds and (recalling (13)) such that for each j and almost every $x'_j \in \left(\underline{x}_j, \overline{x}_j\right)$ there is a J-vector $\hat{x}\left(j, x'_j\right) \in \mathcal{X}$ satisfying

$$\hat{x}_j\left(j, x'_j\right) = x'_j \quad \text{and}$$

$$\frac{\partial \ln \phi\left(y^*|\hat{x}(j, x'_j)\right)}{\partial x_k} \neq 0 \quad \text{iff } k = j. \tag{21}$$

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

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

By (21), we may rewrite this as

$$\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x'_j)/\partial x_j} = \frac{\frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(j,x'_j))}{\partial y_k}}{\frac{\partial \ln \phi(y^*|\hat{x}(j,x'_j))}{\partial x_j}}. \tag{22}$$

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) $x'_j \in \left(\underline{x}_j, \overline{x}_j\right)$. By the same arguments leading up to (22), but with x^0_j taking the role of x'_j, we obtain

$$\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x^0_j)/\partial x_j} = \frac{\frac{\partial \ln \phi(y^*|x)}{\partial y_k} - \frac{\partial \ln \phi(y^*|\hat{x}(x^0_j))}{\partial y_k}}{\frac{\partial \ln \phi(y^*|\hat{x}(x^0_j))}{\partial x_j}} \tag{23}$$

yielding identification of $\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x^0_j)/\partial x_j}$ for all $x^0_j \in \left(\underline{x}_j, \overline{x}_j\right)$. Because the Jacobian determinant $\mathsf{J}(y)$ is nonzero, $\partial r_j(y^*)/\partial y_k$ cannot be zero for all k. Thus for each j there is some k such that the ratio

$$\frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x^0_j)/\partial x_j} \bigg/ \frac{\partial r_j(y^*)/\partial y_k}{\partial g_j(x'_j)/\partial x_j}$$

---

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

[20]Note that we do not require uniqueness of u^* or $\mathcal{U}$, nor therefore uniqueness of the associated y^* or $\mathcal{X}$. We use only the fact (under Assumption 2) that for a given x there exist both (i) a value y^* mapping through (17) to a critical point u^* and (ii) a rectangle $\mathcal{X}$ around x mapping through (18) to a rectangle $\mathcal{U}$ around u^* on which $\ln f$ is regular.

is well defined. This establishes the result.[21] □

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 $\frac{\partial g_j(x_j')/\partial x_j}{\partial g_j(x_j^0)/\partial x_j}$ for all j and any two points x^0 and x' 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' in $\mathbb{X}$ there exists a simple chain from x^0 to x^1 consisting of elements (rectangles) from $\{\mathcal{X}(x)\}_{x\in\mathbb{X}}$.[22] Since the ratios $\frac{\partial g_j(x_j^1)/\partial x_j}{\partial g_j(x_j^2)/\partial x_j}$ are known for all points (x_j^1, x_j^2) in each of these rectangles, it follows that the ratios $\frac{\partial g_j(x_j')/\partial x_j}{\partial g_j(x_j^0)/\partial x_j}$ are identified for all j. Now observe that because $\mathbb{X}$ is a connected open subset of $\mathbb{R}^J$, $\mathbb{X}$ is path-connected. 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). □

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 with regularity of $\ln f$ on $\mathbb{R}^J$; this is equivalent to the combination of conditions required by Matzkin (2008).[23]

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 $\mathcal{X}(x) = \times_j \left(g_j^{-1}(-\infty), g_j^{-1}(\infty) \right)$ for all x. Then by (16), $\mathcal{U}(x) = \mathbb{R}^J$, yielding the result. □

Our second sufficient condition allows limited—even arbitrarily small—support for X while requiring only a local condition on the log density $\ln f$.[24]

Condition M. $\ln f$ has a nondegenerate local maximum.

---

[21] 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.

[22] See, e.g., van Mill (2002), Lemma 1.5.21.

[23] 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 $u_j' \in (\underline{u}_j, \overline{u}_j)$" in the definition of regularity with "all $u_j' \in (\underline{u}_j, \overline{u}_j)$." 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).

[24] A nondegenerate local minimum would also suffice.

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

Proof. See Appendix B. □

The 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., Matsumoto (2002), 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.[25] Nonetheless, because Proposition 3 avoids any extreme requirement, it may be more important. In fact it does not merely avoid the large support assumption: it permits instruments that vary only over a small open ball. And it does so while requiring only a relatively mild condition on the log density (see section 6).

4 Identification with a Linear Index Function

When each g_j is known, the model (7) reduces to the special case[26]

$$r_j(Y) = X_j + U_j \quad j = 1, \ldots, J. \tag{24}$$

where (8) becomes

$$\phi(y|x) = f(r(y) - x) \, |\mathsf{J}(y)| . \tag{25}$$

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 Matzkin (2015).[27] Below we give two theorems demonstrating identification of this model.[28] 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 restriction on the log density, this condition is covered by our genericity result in section 6.

---

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

[26] More formally, for each j, redefine $X_j = g_j(X_j)$, then redefine g_j to be the identity function. All properties required by Assumption 1 are preserved.

[27] Note also that if one specifies $g_j(X_j) = X_j\beta_j$, the normalization (11) implies $\beta_j = 1 \; \forall j$.

[28] Similarly, given identification of g, the falsifiable restrictions derived in Propositions 6 and 7 below imply falsifiable restrictions in the more general model. Under the verifiable hypothesis of Assumption 2, the model defined by (7) and Assumption 1 is also falsifiable (see Proposition 8 in Appendix C).

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 $\ln f$.[29]

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

Proof. Because $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(r\left(y\right) - x\right) \, dx = 1$, (25) implies

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

So by (25),

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

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 $y \in \mathbb{Y}$. Let F_j denote the marginal CDF of U_j. Since

$$\int_{\hat{x}_j \geq x_j, \hat{x}_{-j}} f\left(r\left(y\right) - \hat{x}\right) \, d\hat{x} = F_j\left(r_j\left(y\right) - x_j\right) \tag{26}$$

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

Note that although $\mathsf{J}(y)$ 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 $\mathsf{J}\left(y\right) = \det\left(\partial r(y)/\partial y\right)$ provides a falsifiable restriction of the model and large support assumption.

Proposition 4. *If $\mathbb{X} = \mathbb{R}^J$, then the joint hypothesis of the linear index model (24) 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,

$$\frac{\partial \ln \phi(y|x)}{\partial x_j} = -\frac{\partial \ln f(r(y) - x)}{\partial u_j} \tag{27}$$

---

[29]The argument used to show Theorem 2 was first used by Berry and Haile (2014) in combination with additional assumptions and arguments to demonstrate identification in a model of differentiated products demand and supply.

and

$$\frac{\partial \ln \phi(y|x)}{\partial y_k} = \frac{\partial \ln |J(y)|}{\partial y_k} - \sum_j \frac{\partial \ln \phi(y|x)}{\partial x_j} \frac{\partial r_j(y)}{\partial y_k}. \tag{28}$$

We rewrite (28) as

$$a_k(x,y) = d(x,y)^\mathsf{T} b_k(y) \tag{29}$$

where we define $a_k(x,y) = \frac{\partial \ln \phi(y|x)}{\partial y_k}$, $d(x,y)^\mathsf{T} = \left(1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \ldots, -\frac{\partial \ln \phi(y|x)}{\partial x_J}\right)$, and $b_k(y) = \left(\frac{\partial \ln |J(y)|}{\partial y_k}, \frac{\partial r_1(y)}{\partial y_k}, \ldots, \frac{\partial r_J(y)}{\partial y_k}\right)^\mathsf{T}$. Here $a_k(x,y)$ and $d(x,y)$ are observable whereas $b_k(y)$ involves unknown derivatives of the functions r_j. From (29) it is clear that $b_k(y)$ is identified if there exist points $\tilde{\mathbf{x}} = \left(\tilde{x}^0, \ldots, \tilde{x}^J\right)^\mathsf{T}$, with each $\tilde{x}^j \in \mathbb{X}$, such that the $(J+1) \times (J+1)$ matrix

$$D(\tilde{\mathbf{x}}, y) \equiv \begin{pmatrix} d(\tilde{x}^0, y)^\mathsf{T} \\ \vdots \\ d(\tilde{x}^J, y)^\mathsf{T} \end{pmatrix} \tag{30}$$

has full rank.[30] This yields the following observation, obtained previously by Matzkin (2015).

Lemma 3. *Let Assumption 1 hold. For a given $y \in \mathbb{Y}$, suppose there exists no nonzero vector $c = (c_0, c_1, \ldots, c_J)^\mathsf{T}$ such that $d(x,y)^\mathsf{T} c = 0 \; \forall x \in \mathbb{X}$. 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 $\tilde{\mathbf{x}}$ such that $D(\tilde{\mathbf{x}}, y)$ is invertible. Matzkin (2010) has provided a sufficient condition: that there exist $\tilde{\mathbf{x}}$ such that $D(\tilde{\mathbf{x}}, y)$ is diagonal with nonzero diagonal terms. Using our earlier geometric interpretation, that condition requires the log density to have an appropriate set of critical values 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* $D(\tilde{\mathbf{x}}, y)$ 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 values in $\{r(y) - \mathbb{X}\}$ to obtain a diagonal or triangular $D(\tilde{\mathbf{x}}, y)$ 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. By instead considering conditions on the second derivatives of $\ln f$, we can show that a mild restriction

---

[30]In particular, let $A_k(\tilde{\mathbf{x}}, y) = \left(a_k(\tilde{x}^0, y) \; \cdots \; a_k(\tilde{x}^J, y)\right)^\mathsf{T}$ and stack the equations obtained from (29) at each of the points $\tilde{x}^{(0)}, \ldots, \tilde{x}^{(J)}$, yielding $A_k(\tilde{\mathbf{x}}, y) = D(\tilde{\mathbf{x}}, y) b_k(y)$. When $D(\tilde{\mathbf{x}}, y)$ is invertible we obtain $b_k(y) = D(\tilde{\mathbf{x}}, y)^{-1} A_k(\tilde{\mathbf{x}}, y)$.

ensures identification, even when X has arbitrarily small support.[31]

Define the second-derivative matrix

$$H_\phi(x, y) = \frac{\partial^2 \ln \phi(y|x)}{\partial x \partial x^\mathsf{T}} = \begin{pmatrix} \frac{\partial^2 \ln \phi(y|x)}{\partial x_1^2} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_J \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln \phi(y|x)}{\partial x_1 \partial x_J} & \cdots & \frac{\partial^2 \ln \phi(y|x)}{\partial x_J^2} \end{pmatrix}. \tag{31}$$

Observe that, fixing a value of y and $c = (c_0, c_1, \ldots, c_J)^\mathsf{T}$, $d(x, y)^\mathsf{T} c$ is a function of x, with gradient $H_\phi(x, y)(c_1, \ldots, c_J)^\mathsf{T}$. This leads to the following lemma, and the theorem that follows.[32] Note that although the theorem allows for the possibility of identification on a strict subset of $\mathbb{Y}$, the case $\mathbb{Y}' = \mathbb{Y}$ is also covered.

Lemma 4. *For a nonzero vector $c = (c_0, c_1, \ldots, c_J)^\mathsf{T}$,*

$$d(x, y)^\mathsf{T} c = 0 \quad \forall x \in \mathbb{X} \tag{32}$$

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

$$H_\phi(x, y) \tilde{c} = 0 \quad \forall x \in \mathbb{X}. \tag{33}$$

Proof. See Appendix C. $\qquad\qquad\square$

Theorem 3. *Let Assumption 1 hold and let $\mathbb{Y}' \subset \mathbb{Y}$ be open and connected. Suppose that, for almost all $y \in \mathbb{Y}'$, there is no nonzero J-vector $\tilde{c}$ such that*

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

Then in the linear index model, r is identified on $\mathbb{Y}'$.

Proof. From (27), $\partial^2 \ln \phi(y|x) / \partial x_j \partial x_k = \partial^2 \ln f(r(y) - x) / \partial u_j \partial u_k$. Identification of $\partial r(y) / \partial y_k$ for all k and $y \in \mathbb{Y}'$ then follows from the definition (31), 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 $\mathbb{Y}'$.[33] 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. $\qquad\square$

Corollary 1 follows immediately from Theorem 3.

---

[31]Matzkin (2015, Theorem 4.2) considers a different type of second-derivative condition to show partial identification in a related model. Chiappori and Komunjer (2009b) combine a completeness condition with a different second-derivative condition to obtain identification in a discrete choice model requiring more than one instrument per equation.

[32]Using related arguments, Propositions 6 and 7 in Appendix E demonstrate falsifiability of the linear index model.

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

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

This corollary 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 $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ is nonsingular almost everywhere, identification of r on $\mathbb{Y}$ holds even with arbitrarily small $\mathbb{X}$. Nonsingularity of $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ 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 Corollary 1 to apply, it is sufficient that $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ be nonsingular once in $\{r(y) - \mathbb{X}\}$ for each $y \in \mathbb{Y}'$.[34] We will see below that this is requirement holds generically, even when $\mathbb{X}$ is a small open ball.

In addition, we observe that (27) immediately implies verifiability of the rank condition hypothesized in Corollary 1.

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

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

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. □

We view Corollary 3, 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.

---

[34] In some applications it may be reasonable to assume that U_j and U_k are independent for all $k \neq j$. Because $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ is diagonal under independence, it is then sufficient that $\partial^2 \ln f_j(\hat{u}_j) / \partial u_j^2$ be nonzero for all j at some $\hat{u} \in \{r(y) - \mathbb{X}\}$.

Corollary 3. *Let* $\mathbb{Y}' \subset \mathbb{Y}$ *be open and connected, and suppose that Assumption 1 and Condition M hold. If, for almost all* $y \in \mathbb{Y}'$, $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ *is nonsingular at some* $u \in \{r(y) - g(\mathbb{X})\}$, *then* g *is identified on* $\mathbb{X}$ *and* r *is identified on* $\mathbb{Y}'$.

Proof. By Theorem 1 and Proposition 3, g is identified on $\mathbb{X}$. Identification of r on $\mathbb{Y}'$ then follows from Corollary 1. $\square$

Although Corollary 3 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 $\{r(y) - g(\mathbb{X})\}$ for almost all $y \in \mathbb{Y}'$. The importance of the result depends on the restrictiveness of these requirements. This is a topic we take up in the following section. There we use a standard notion of genericity to show that, even when X has arbitrarily small support, simultaneous satisfaction of these conditions can be viewed as "typical" among log densities on $\mathbb{R}^J$ that are twice continuously differentiable an possess a local maximum.[35] Thus, there is at least one formal sense in which the requirements of Corollary 3 may be viewed as mild.

6   Generic Identification

In this section we demonstrate that, if $\mathbb{Y}'$ is the pre-image of any (open) rectangle in $\mathbb{R}^J$ (or more generally, any bounded open connected set), the requirements of Corollary 3 hold generically among log densities on $\mathbb{R}^J$ that are twice continuously differentiable an possess a local maximum. This is true even when instruments vary only over an arbitrarily small open ball. This implies a form of generic identification of g on $\mathbb{X}$ and of r on $\mathbb{Y}'$.[36]

To demonstrate this result, first let $\mathcal{G} = \times_j \left[g_j \left(\underline{x}_j \right), g_j \left(\overline{x}_j \right) \right] \subset g(\mathbb{X})$ be a compact "square" in $\mathbb{R}^J$ with width $w_x = g_j \left(\overline{x}_j \right) - g_j \left(\underline{x}_j \right) > 0$ for all j.[37] We form a tessellation of $\mathbb{R}^J$ using squares of width $w_x / 2$.[38] Let $\sigma = (\sigma_1, \ldots, \sigma_J)$ denote a J-vector of

---

[35] Existence of a local maximum is a mild requirement for a continuously differentiable log density on $\mathbb{R}^J$. A local max always exists when $J = 1$. More generally, one sufficient condition is existence of a global maximum, which is guaranteed under the following vanishing tails condition: for any $\epsilon > 0$ there exists a compact set $S(\epsilon)$ such that $f(u) < \epsilon$ for all $u \notin S(\epsilon)$.

[36] Because $r(\mathbb{Y}')$ may be arbitrarily large, the gap between generic identification on $\mathbb{Y}'$ and generic identification on the pre-image of $\mathbb{R}^J$ (i.e., on $\mathbb{Y}$) may be of little importance.

[37] Such a square must exist. Because $\mathbb{X}$ is open and each component g_j of g is continuous and strictly increasing, $g(\mathbb{X})$ is open. Given any open ball in $g(\mathbb{X})$, there exists $w_x > 0$ sufficiently small that some square of width w_x lies inside the ball.

[38] In three or more dimensions, a tessellation is also known as a honeycomb, and a what we call a square is a cube or hypercube. For simplicity, as with our use of the term "rectangle," we use the language of the two-dimensional case. We point out, however, that our definition of rectangle involves an open set while what we call a square is closed.

integers. For each $\sigma \in \mathbb{Z}^J$ define the square

$$sq_\sigma = \times_j \left[\frac{2\sigma_j - 1}{4} w_x, \frac{2\sigma_j + 1}{4} w_x \right].$$

Then $\{sq_\sigma\}_{\sigma \in \mathbb{Z}^J}$ forms a regular tessellation of $\mathbb{R}^J$ such that, for every $y \in \mathbb{Y}$, the set $\{r(y) - \mathcal{G}\}$ (and therefore $\{r(y) - g(\mathbb{X})\}$) covers some square sq_σ.

Given any open set $S \subset \mathbb{R}^J$ that is bounded and connected, let $\mathbb{Y}_S$ denote the pre-image of S under r. Because r is continuous, $\mathbb{Y}_S$ is open; and because r has continuous inverse (see proof of Lemma 1), $\mathbb{Y}_S$ is connected. Boundedness of S implies that there is a finite set $\mathbb{Z}_S \subset \mathbb{Z}^J$ such that

$$\cup_{\sigma \in \mathbb{Z}_S} sq_q \supset \{r(\mathbb{Y}_S) - \mathcal{G}\}, \tag{34}$$

where here the minus sign denotes the Minkowski difference. By construction, for every $y \in \mathbb{Y}_S$ there exists $\sigma \in \mathbb{Z}_S$ such that $\{r(y) - g(\mathbb{X})\}$ covers sq_σ. So if Assumption 1 and Condition M hold, the following is sufficient for Corollary 3 to apply, yielding identification of g on $\mathbb{X}$ and of r on $\mathbb{Y}_S$.

Condition H. For every $\sigma \in \mathbb{Z}_S$, $\partial^2 \ln f(u) / \partial u \partial u^\mathsf{T}$ is nonsingular at some $u \in sq_\sigma$.

We show below that simultaneous satisfaction of Conditions M and H is generic in the space of log densities on $\mathbb{R}^J$ that are twice continuously differentiable and possess a local maximum. To define this space, first let $C^2(\mathbb{R}^J)$ denote the space of twice continuously differentiable real valued functions on $\mathbb{R}^J$. We define a topology on $C^2(\mathbb{R}^J)$ using the C^2 extended norm $\|\cdot\|_{C^2}$, where

$$\|h\|_{C^2} = \sup_{u \in \mathbb{R}^J} |h(u)| + \max_{j \in \{1,\dots,J\}} \sup_{u \in \mathbb{R}^J} \left| \frac{\partial h(u)}{\partial u_j} \right| + \max_{j,k \in \{1,\dots,J\}} \sup_{u \in \mathbb{R}^J} \left| \frac{\partial^2 h(u)}{\partial u_j \partial u_k} \right|$$

for any $h \in C^2(\mathbb{R}^J)$. Under the induced extended metric,[39] two functions h and $\hat{h}$ in $C^2(\mathbb{R}^J)$ are deemed to be "close" if these functions and their partial derivatives up to order 2 are uniformly close.[40] Let $\mathcal{F} \subset C^2(\mathbb{R}^J)$ denote the subspace (with subspace topology) of twice continuously differentiable log densities on $\mathbb{R}^J$ that possess a local maximum. We say that functions in a set $\mathcal{H} \subset \mathcal{F}$ are *generic* in $\mathcal{F}$ if $\mathcal{H}$ is a dense open subset of $\mathcal{F}$ (see, e.g., Mas-Colell (1985)).[41]

---

[39] A metric inducing the same topology is $\tilde{d}(h', h) = \frac{\|h' - h\|_{C^2}}{1 + \|h' - h\|_{C^2}}$. We work with the C^2 extended metric to simplify exposition.

[40] Our genericity result also holds under the alternative (coarser) topology of compact convergence (of sequences of functions in $C^2(\mathbb{R}^J)$ and their partial derivatives up to order 2). Results regarding denseness in $\mathcal{F}$ trivially extend to any coarser topology. And our arguments demonstrating openness rely only on convergence within explicitly defined compact subsets of $\mathbb{R}^J$.

[41] A weaker notion of genericity is that of a *residual* set (countable intersection of dense open

Theorem 4. *Let Assumption 1 hold, let $S \subset \mathbb{R}^J$ be bounded, open, and connected, and let $\mathbb{Z}_S$ be as defined in (34). The set $\mathcal{F}_S^* = \{\ln f \in \mathcal{F} : \text{Conditions M and H hold}\}$ is dense and open in $\mathcal{F}$.*

To prove this result, define the following subsets of $\mathcal{F}$:

$$
\begin{aligned}
\mathcal{F}_\sigma^H &= \left\{\ln f \in \mathcal{F} : \partial^2 \ln f(u)/\partial u \partial u^\mathsf{T} \text{ is nonsingular at some } u \in sq_\sigma\right\} \\
\mathcal{F}^H &= \{\ln f \in \mathcal{F} : \text{Condition H holds}\} \\
\mathcal{F}^M &= \{\ln f \in \mathcal{F} : \text{Condition M holds}\} .
\end{aligned}
$$

Then we have

$$
\mathcal{F}^H = \cap_{\sigma \in \mathbb{Z}_S} \mathcal{F}_\sigma^H \tag{35}
$$

and

$$
\mathcal{F}_S^* = \mathcal{F}^M \cap \mathcal{F}^H. \tag{36}
$$

Thus $\mathcal{F}_S^*$ is a finite intersection of subsets of $\mathcal{F}$. In Corollary 4 below (section 6.1), we show that $\mathcal{F}_S^*$ is dense in $\mathcal{F}$. Lemmas 7 and 8 below (section 6.2) show that $\mathcal{F}^M$ and each $\mathcal{F}_\sigma^H$ (for $\sigma \in \mathbb{Z}^J$) is open in $\mathcal{F}$. Theorem 4 then follows from (35) and (36).[42]

6.1 Dense

Let

$$
\mathcal{F}^* = \mathcal{F}^M \cap_{\sigma \in \mathbb{Z}^J} \mathcal{F}_\sigma^H. \tag{37}
$$

In this subsection we prove the following result, whose Corollary is immediate from the fact that $\mathcal{F}^* \subset \mathcal{F}_S^*$.

Lemma 5. *$\mathcal{F}^*$ is dense in $\mathcal{F}$.*

Corollary 4. *$\mathcal{F}_S^*$ is dense in $\mathcal{F}$.*

Fix any $\ln f \in \mathcal{F}$ and let u^* denote a local max. To prove Lemma 5, it is sufficient to show that for any $\epsilon > 0$, there exists $\ln f^* \in \mathcal{F}^*$ such that $||\ln f^* - \ln f||_{C^2} < \epsilon$. Let

---

subsets). With minor modification, our arguments below demonstrate that even the set $\mathcal{F}^* \subset \mathcal{F}_S^*$ (defined in (37)) is residual in $\mathcal{F}$. By Corollary 3, log densities in $\mathcal{F}^*$ ensure identification even for the case $\mathbb{Y}' = \mathbb{Y}$. However, because $\mathcal{F}$ is not a complete metric space, we use the more demanding genericity notion, requiring a dense open subset. As noted by Mas-Colell (1985), these topological notions of genericity are standard in infinite dimensional spaces but can sometimes provide too weak a notion of "typical" (see also Hunt, Sauer, and Yorke (1992), Anderson and Zame (2001), or Stinchcombe (2002)). Exploration of other notions of genericity applicable to our setting is beyond the scope of this paper.

[42] Although we focus on genericity of the conditions required by Corollary 3, Lemmas 5 and 7 below also imply that $\mathcal{F}^M$ is a dense open subset of $\mathcal{F}$ (see Corollary 2).

$w_f > 0$ be such that $\ln f(u^*) \geq \ln f(u)$ for all u in the square $\times_j \left[u_j^* - \frac{w_f}{2}, u_j^* + \frac{w_f}{2} \right]$.[43] Let s^* be a closed square with center u^* and width

$$w = \min \left\{ \frac{w_x}{4}, w_f \right\}.$$

For all j, let $\underline{u}_j$ and $\overline{u}_j$ be defined such that $s^* = \times_j \left[\underline{u}_j, \overline{u}_j \right]$.

Starting from s^*, form another tessellation of $\mathbb{R}^J$ using squares of width w. Let $\tau = (\tau_1, \ldots, \tau_J)$ denote a J-vector of integers. For each $\tau \in \mathbb{Z}^J$ define the square $s_\tau = \times_j \left[\underline{u}_j + \tau_j w, \overline{u}_j + \tau_j w \right]$. Then $\{s_\tau\}_{\tau \in \mathbb{Z}^J}$ forms a regular tessellation of $\mathbb{R}^J$.[44] Let $\dot{u}_\tau = (\dot{u}_{\tau 1}, \ldots, \dot{u}_{\tau J})$ denote the center of square τ. For $\tau = (0, \ldots, 0)$, we then have $s_\tau = s^*$ and $\dot{u}_\tau = u^*$. For all $u \in \mathbb{R}^J$, let $\tau(u) \in \mathbb{Z}^J$ denote the index of a square such that $u \in s_{\tau(u)}$.[45] Observe that every cell of the tessellation $\{sq_\sigma\}_{\sigma \in \mathbb{Z}^J}$ covers at least one cell of the tessellation $\{s_\tau\}_{\tau \in \mathbb{Z}^J}$. We prove Lemma 5 by constructing an arbitrarily small perturbation of $\ln f$ that (i) lies in $\mathcal{F}$, (ii) has a nondegenerate local max at u^*, and (iii) has a nonsingular Hessian matrix at the center of every square s_τ.

Let f denote the probability density function associated with the fixed log density $\ln f$ (i.e., $f = \exp(\ln f)$). For $v \in \mathbb{R}$ define[46]

$$p(v) = 1\{|v| \leq 1\} \left(1 - v^2 \right)^3.$$

Given any $\lambda > 0$ and $\{\lambda_\tau \in (0, \lambda]\}_{\tau \in \mathbb{Z}^J}$, for all $u \in \mathbb{R}^J$ let

$$f_\lambda(u) = \kappa f(u) \exp \left(\lambda_{\tau(u)} \prod_{j=1}^{J} p \left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2} \right) \right), \tag{38}$$

with particular values of each λ_τ to be determined below. The scalar κ is chosen to ensure that $f_\lambda(u)$ integrates to one on $\mathbb{R}^J$, i.e.,

$$\kappa = \left[\int_{\mathbb{R}^J} f(u) \exp \left(\lambda_{\tau(u)} \prod_{j=1}^{J} p \left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2} \right) \right) du \right]^{-1}. \tag{39}$$

---

[43]Such w_f must exist since around any local max is an open ball on which $\ln f(u^*)$ is (at least weakly) maximal.

[44]Unlike the tessellation $\{sq_\sigma\}_{\sigma \in \mathbb{Z}^J}$, this tessellation may vary with the choice of $\ln f$ through the choice of the point u^* and width w.

[45]There will be only one such square for almost all u. However, any u on the boundary of one square will also be on the boundary of at least one other square. How the function $\tau(u)$ resolves this ambiguity does not matter for what follows.

[46]The function p is proportional to a triweight kernel (see, e.g., Silverman (1986)). In Appendix D we discuss some relevant properties of p and of a log density perturbed on a square using this function as in (38).

Because the term $\lambda_{\tau(u)} \prod_{j=1}^{J} p\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right)$ takes only values between 0 and $\lambda_{\tau(u)}$, κ must lie in the interval $[\exp(-\lambda), 1]$. Thus, by construction the perturbed function f_λ is positive on $\mathbb{R}^J$, integrates to one, and (see Appendix D) is twice continuously differentiable on $\mathbb{R}^J$.

We first show that $\ln f_\lambda$ is made arbitrarily close to $\ln f$ by setting λ sufficiently small. The proof follows easily from (38) and is relegated to Appendix C.

Claim 5. *For any $\epsilon > 0$, there exists $\lambda^* > 0$ such that for any $\lambda \in (0, \lambda^*)$ and any $\{\lambda_\tau \in (0, \lambda]\}_{\tau \in \mathbb{Z}^J}$, $\|\ln f_\lambda - \ln f\|_{C^2} < \epsilon$.*

To complete the proof of Lemma 5 we show that, given any $\lambda > 0$, the scaling parameters $\{\lambda_\tau\}_{\tau \in \mathbb{Z}^J}$ can be chosen to ensure that $\ln f_\lambda \in \mathcal{F}^*$.

Claim 6. *For any $\lambda > 0$, there exist $\{\lambda_\tau \in (0, \lambda]\}_{\tau \in \mathbb{Z}^J}$ such that $\ln f_\lambda \in \mathcal{F}^*$.*

Proof. Fix any $\lambda > 0$. We first show that there exist $\{\lambda_\tau \in (0, \lambda]\}_{\tau \in \mathbb{Z}^J}$ such that, for all $\sigma \in \mathbb{Z}^J$, $\partial^2 \ln f_\lambda(u)/\partial u \partial u^\intercal$ is nonsingular at some $u \in sq_\sigma$. Because every square sq_σ covers the square s_τ for some $\tau \in \mathbb{Z}^J$, it is sufficient that, for each τ and some $\lambda_\tau \in (0, \lambda]$, $\partial^2 \ln f_\lambda(u)/\partial u \partial u^\intercal$ is nonsingular at $u = \dot{u}_\tau$. Take any $\tau \in \mathbb{Z}^J$. Equations (C.4)–(C.6) imply[47]

$$\frac{\partial^2 \ln f_\lambda(\dot{u}_\tau)}{\partial u \partial u^\intercal} = \frac{\partial^2 \ln f(\dot{u}_\tau)}{\partial u \partial u^\intercal} - \left(\frac{24\lambda_\tau}{w^2}\right) I_J,$$

where I_J is the identity matrix. The eigenvalues of $\partial^2 \ln f_\lambda(\dot{u}_\tau)/\partial u \partial u^\intercal$ are therefore equal to those of $\partial^2 \ln f(\dot{u}_\tau)/\partial u \partial u^\intercal$ minus $\frac{24\lambda_\tau}{w^2}$. So for almost all values of $\lambda_\tau \in (0, \lambda]$, all eigenvalues of $\partial^2 \ln f_\lambda(\dot{u}_\tau)/\partial u \partial u^\intercal$ are nonzero, ensuring that $\ln f_\lambda \in \mathcal{F}^H$.[48] Fixing any such $\{\lambda_\tau \in (0, \lambda]\}_{\tau \in \mathbb{Z}^J}$, we then complete the proof by showing that $\ln f_\lambda \in \mathcal{F}^M$. By our choice of the point u^* and square s^*, $u^* \in \arg\max_{u \in s^*} \ln f(u)$. And because $u^* = \dot{u}_\tau$ for $\tau = (0, \ldots, 0)$, $u^* = \arg\max_{u \in s^*} [\ln f_\lambda(u) - \ln f(u)]$. Thus, u^* is a local maximum of $\ln f_\lambda$ and, by the choice of λ_τ above for $\tau = (0, \ldots, 0)$, $\partial^2 \ln f_\lambda(u^*)/\partial u \partial u^\intercal$ is nonsingular. $\square$

6.2 Open

To prove the required openness results, we begin with a result from the literature on Morse functions.[49] Given any compact $K \subset \mathbb{R}^J$, let $C^2(K)$ denote the space of twice

---

[47]See the values of $p(0)$, $p'(0)$ and $p''(0)$ given in section D.

[48]Since we needed only *one* such λ_τ we have shown more than necessary. Thus the "abundance" of perturbations lying in $\mathcal{F}^*$ is even greater than required by the notion of C^2-denseness.

[49]See, e.g., Lemma 5.32 in Banyaga and Hurtubise (2004). We provide a proof in Appendix C.

continuously differentiable real valued functions on K. For $h \in C^2(K)$, let

$$\|h\|_{C_K^2} = \sup_{u \in K} |h(u)| + \max_{j \in \{1,\dots,J\}} \sup_{u \in K} \left| \frac{\partial h(u)}{\partial u_j} \right| + \max_{j,k \in \{1,\dots,J\}} \sup_{u \in K} \left| \frac{\partial^2 h(u)}{\partial u_j \partial u_k} \right|.$$

Lemma 6. *Suppose that $\mathfrak{f} \in C^2(K)$ has no degenerate critical point. Then there exists $\epsilon > 0$ such that any $\mathfrak{g} \in C^2(K)$ satisfying $\|\mathfrak{f} - \mathfrak{g}\|_{C_K^2} < \epsilon$ has no degenerate critical point.*

Next we show that $\mathcal{F}^M$ is an open subset of $\mathcal{F}$.

Lemma 7. *For every $\ln f \in \mathcal{F}^M$ there exists $\epsilon > 0$ such that if $\ln \hat{f} \in \mathcal{F}$ and $\left\| \ln \hat{f} - \ln f \right\|_{C^2} < \epsilon$, then $\ln \hat{f} \in \mathcal{F}^M$.*

Proof. Take any $\ln f \in \mathcal{F}^M$ and let u^* denote a nondegenerate local maximum. In the proof of Lemma 9 (see Appendix B) we show that for some compact connected set Σ with nonempty interior there exists $\underline{c}$ such that (i) the upper contour set $A(\underline{c}, \Sigma) = \{u \in \Sigma : \ln f(u) \geq \underline{c}\}$ lies in the interior of Σ and (ii) the restriction of $\ln f$ to $A(\underline{c}, \Sigma)$ attains a maximum $\bar{c} = \ln f(u^*) > \underline{c}$ at its unique critical point. Let $K = A(\underline{c}, \Sigma)$. Because $\ln f$ is continuous, $A(\underline{c}, \Sigma)$ is closed in $\mathbb{R}^J$. And because $A(\underline{c}, \Sigma)$ lies on the interior of the compact set Σ, $A(\underline{c}, \Sigma)$ is bounded. Thus K is compact, and $\ln f$ has no degenerate critical point on K. So by Lemma 6, for all sufficiently small $\epsilon > 0$, $\left\| \ln \hat{f} - \ln f \right\|_{C^2} < \epsilon$ ensures that also has no degenerate critical point on K. To complete the proof it suffices to show that, for all sufficiently small $\epsilon > 0$, $\left\| \ln \hat{f} - \ln f \right\|_{C^2} < \epsilon$ ensures that the restriction of $\ln \hat{f}$ to K has a maximum on the interior of K. By continuity of $\ln f$, $\ln f(u) = \underline{c}$ for all u on the boundary of K. So when $\sup_{u \in K} \left| \ln \hat{f}(u) - \ln f(u) \right| < \epsilon$ (implied by $\left\| \ln \hat{f} - \ln f \right\|_{C^2} < \epsilon$), $\ln \hat{f}$ must take values no more than $\underline{c} + \epsilon$ on the boundary of K and no less than $\bar{c} - \epsilon$ at u^*. For sufficiently small $\epsilon > 0$ we have $\underline{c} + \epsilon < \bar{c} - \epsilon$, requiring that the restriction of $\ln \hat{f}$ to K have an interior maximum. $\qquad\square$

Finally, we show that for every $\sigma \in \mathbb{Z}^J$, F_σ^H is an open subset of $\mathcal{F}$.

Lemma 8. *For any $\sigma \in \mathbb{Z}^J$ and any $\ln f \in \mathcal{F}_\sigma^H$, there exists $\epsilon > 0$ such that if $\ln \hat{f} \in \mathcal{F}$ and $\left\| \ln \hat{f} - \ln f \right\|_{C^2} < \epsilon$, $\ln \hat{f} \in \mathcal{F}_\sigma^H$.*

Proof. Fix $\sigma \in \mathbb{Z}^J$ and $\ln f \in \mathcal{F}_\sigma^H$, the latter implying that for some $\delta > 0$ and some $\hat{u} \in sq_\sigma$, $|\det(\partial^2 \ln f(\hat{u})/\partial u \partial u^\mathsf{T})| > \delta$. Recall that $\|h\|_{C^2} < \epsilon$ requires $\max_{j,k \in \{1,\dots,J\}} \sup_{u \in S} \left| \frac{\partial^2 h(u)}{\partial u_j \partial u_k} \right| < \epsilon$. So for sufficiently small $\epsilon > 0$, $\left\| \ln f - \ln \hat{f} \right\|_{C^2} < \epsilon$ implies $\left| \det\left(\frac{\partial^2 \ln f(\hat{u})}{\partial u \partial u^\mathsf{T}} \right) - \det\left(\frac{\partial^2 \ln \hat{f}(\hat{u})}{\partial u \partial u^\mathsf{T}} \right) \right| < \delta$, ensuring that $\left| \det\left(\frac{\partial^2 \ln \hat{f}(\hat{u})}{\partial u \partial u^\mathsf{T}} \right) \right| > 0$. $\quad\square$

7   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. Indeed, our genericity result demonstrates a standard formal sense in which, even when instruments have arbitrarily small support, our sufficient conditions for identification may be viewed as mild. We have also shown that 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. We have also shown that key identifying assumptions required by our results are verifiable, and that the maintained assumptions of the model 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*

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

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

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

where $\delta_j\left(Z, X_j, U_j\right) = g_j\left(Z, X_j\right) + U_j$. The resulting model features nonseparable structural errors but requires them to enter the nonseparable nonparametric function Γ_j through the index $\delta_j\left(Z, X_j, U_j\right)$. If each function γ_j is invertible (e.g., strictly increasing) in $\delta_j\left(Z, X_j, U_j\right)$, then one obtains (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 Γ_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 $Q = \Psi\left(Y, \mathcal{E}\right)$, where Ψ is a concave production function, $Y \in \mathbb{R}_+^J$ is a vector of input quantities, and $\mathcal{E} \in \mathbb{R}^J$ is a vector of factor-specific productivity shocks affecting the firm.[50] The shocks are known by the firm when input levels are chosen, but unobserved to the econometrician. Let P and*

---

[50]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

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

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 demand for each input j (or, equivalently, each first-order condition in (A.2)) depends on the entire vector of shocks.

The index structure can be obtained by assuming that, for some unknown function ψ_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 weak form of multiplicative separability with a formalization of the notion that the shocks are factor-specific: ϵ_j is the only shock directly affecting the marginal product of input j.[51]

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 $X_j = \frac{W_j}{P}$, $U_j = \ln(\mathcal{E}_j)$, 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 ψ_j and, therefore, the realizations of each productivity shock $\mathcal{E}_j$. Since Q and Y are observed, this implies identification of the production function Ψ.

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*

---

[51] A similar form of additive separability would also lead to the residual index structure.

$1, \ldots, K$, *with* $Q = (Q_1, \ldots, Q_K)$ *denoting their quantities (expressed, e.g., in levels 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*

$$Q_j = D_j\left(P, V, \xi\right). \tag{A.3}$$

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

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

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

$$Q_j = \sigma_j\left(P, \delta\right). \tag{A.4}$$

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

$$c_j\left(\kappa_j\right),$$

where, for some strictly increasing function β_j,

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

Let $\kappa = (\kappa_1, \ldots, \kappa_K)$. *Suppose that prices are determined through oligopoly competition, yielding a reduced form*[52]

$$P_j = \pi_j\left(\delta, \kappa\right) \quad j = 1, \ldots, K. \tag{A.5}$$

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

Berry, Gandhi, and Haile (2013) and Berry and Haile (2014) provide conditions ensuring that the system of equations (A.4) and (A.5) can be inverted, yielding a $2K \times 2K$ system

$$\alpha_j\left(V_j\right) + \xi_j = \sigma_j^{-1}\left(Q, P\right)$$
$$\beta_j\left(W_j\right) + \omega_j = \pi_j^{-1}\left(Q, P\right).$$

Letting $J = 2K$, $r = \left(\sigma_1^{-1}, \ldots, \sigma_K^{-1}, \pi_1^{-1}, \ldots, \pi_K^{-1}\right)^\mathsf{T}$, $X = (V_1, \ldots, V_K, W_1, \ldots, W_K)^\mathsf{T}$, *and* $U = (\xi_1, \ldots, \xi_K, \omega_1, \ldots, \omega_K)^\mathsf{T}$, *we obtain the system (6). The primary objects of*

---

[52]Berry and Haile (2014) show that such a reduced form arises under standard models of oligopoly supply.

interest in applications include demand derivatives (elasticities) and firms' marginal costs, as these allow construction of a wide range of counterfactual predictions. Identification of all α_j, β_j, σ_j^{-1} and π_j^{-1} immediately implies identification of all π_j and σ_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 σ_j to marginal costs (see Berry and Haile (2014)), 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 $\mathcal{B}(u, \epsilon)$ denote an ϵ-ball around a point $u \in \mathbb{R}^J$. 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 $\ln f$ to Σ. We begin with a new condition and new definition.

Condition M′.   There exists $\underline{c} \in \mathbb{R}$ and a compact connected set $S \subset \mathbb{R}^J$ with nonempty interior such that (i) $\mathcal{A}(\underline{c}; S) \subset \text{int}(S)$, and (ii) the restriction of $\ln f$ to $\mathcal{A}(\underline{c}; S)$ attains a maximum $\bar{c} > \underline{c}$ 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 $\mathcal{R}(u^*, \epsilon)$ such that $u^* \in \mathcal{R}(u^*, \epsilon) \subset \mathcal{B}(u^*, \epsilon)$.

Condition M′ requires that if we "zoom in" to a sufficiently small neighborhood of a local max (first to S, then further to an upper contour set of $\ln f$ on the restricted domain S), the local max 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.

Lemma 9. *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, a nondegenerate critical point is an isolated critical point.[53] So there exists $\epsilon > 0$ such that on the open ball $\mathcal{B}(u^*, \epsilon)$, u^* is both the only critical point and a strict maximum. Let S be a compact connected subset of $\mathcal{B}(u^*, \epsilon)$ 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 $\underline{c} < \bar{c}$ such that the upper contour set $\mathcal{A}(\underline{c}; S)$ lies in the interior of S. Continuity of $\ln f$ implies that $\mathcal{A}(c; S)$

---

[53]See, e.g., Matsumoto (2002), Corollary 2.18.

28

is upper hemicontinuous in c.[54] So because u^* is the only point in $\mathcal{A}(\bar{c}; S)$ and S has nonempty interior, we obtain $\mathcal{A}(\underline{c}; S) \subset \text{int} S$ by setting $\underline{c} = \bar{c} - \delta$ for any sufficiently small $\delta > 0$. $\qquad\square$

Lemma 10. *Local rectangle regularity implies Assumption 2 (rectangle regularity).*

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

$$g_j\left(\underline{\underline{x}}_j(x) \right) + \overline{\overline{u}}_j(x) = g_j\left(\overline{\overline{x}}_j(x) \right) + \underline{\underline{u}}_j(x) = g_j(x) + u^*(x) \quad j = 1, \ldots, J.$$

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

$$\mathcal{U}(x) = \times_j \left(\underline{u}_j(x), \overline{u}_j(x) \right) \subset \overline{\overline{\mathcal{U}}}(x)$$

such that $u^*(x) \in \mathcal{U}(x)$. Let $\mathcal{X}(x) = \times_j \left(\underline{x}_j(x), \overline{x}_j(x) \right)$, where each $\underline{x}_j(x)$ and $\overline{x}_j(x)$ is defined by (18). By construction, $\mathcal{X}(x) \subset \overline{\overline{\mathcal{X}}}(x) \subset \mathbb{X}$ and $\mathcal{U}(x)$ satisfies (16). $\qquad\square$

The final step is to show that Condition M′ implies local rectangle regularity. We will require two lemmas before showing this. For these two results, 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 $\mathcal{A}(c) = \{u \in S : h(u) \geq c\}$

Lemma 11. *Suppose that for some $\underline{c} < c_{\max} \equiv \max_{u \in S} h(u)$, $\mathcal{A}(\underline{c}) \subset \text{int}(S)$. Then $\mathcal{A}(c)$ has nonempty interior for all $c < c_{\max}$.*

Proof. Since $\mathcal{A}(\underline{c}) \subset \text{int}(S)$, we must have $h(\tilde{u}) < c_{\max}$ for some $\tilde{u} \in S$. Therefore, since the continuous image of a connected set is connected, $h(S)$ is a nondegenerate interval. For any $c < c_{\max}$ there must then exist $u \in S$ such that $\max\{\underline{c}, c\} < h(u) < c_{\max}$. 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 $\epsilon > 0$, we have both $\mathcal{B}(u, \epsilon) \subset S$ and (by continuity of h) $h(\hat{u}) > c$ $\forall \hat{u} \in \mathcal{B}(u, \epsilon)$. $\qquad\square$

Lemma 12. *Suppose that for some $\underline{c} \in \mathbb{R}$, $\mathcal{A}(\underline{c}) \subset \text{int}(S)$ and the restriction of h to $\mathcal{A}(\underline{c})$ attains a maximum $\bar{c} > \underline{c}$ at its unique critical point u^*. Then (a) $\mathcal{A}$ is a continuous (upper and lower hemicontinuous) correspondence on $(\underline{c}, \bar{c}]$; and (b) $\mathcal{A}(c)$ is a connected set for all $c \in (\underline{c}, \bar{c})$*

---

[54] Take any compact $\Omega \subset \mathbb{R}^J$ and continuous $h : \Omega \to \mathbb{R}$ with upper contour sets $\mathcal{A}(c) = \{u \in \Omega : h(u) \geq c\}$. Since Ω is compact and h is continuous, $\mathcal{A}$ 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 $u^n \in \mathcal{A}(c^n) \, \forall n$ and $u^n \to \hat{u}$. If $\hat{u} \notin \mathcal{A}(\hat{c})$ then, because $\hat{u}$ must lie in Ω, 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 $u^n \in \mathcal{A}(c^n) \, \forall n$. So $\hat{u} \in \mathcal{A}(\hat{c})$.

Proof. (a) Upper hemicontinuity follows from continuity of h (see footnote 54). To show lower hemicontinuity,[55] take any $\hat{c} \in (\underline{c}, \overline{c}]$, any $\hat{u} \in \mathcal{A}(\hat{c})$, and any sequence $c^n \to \hat{c}$. If $\hat{u} = u^*$ 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 $\hat{u} \neq u^*$. Letting $\| \cdot \|$ denote the Euclidean norm, define a sequence u^n by

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

so that $u^n \in \mathcal{A}(c^n)$ by construction. We now show $u^n \to \hat{u}$. Take arbitrary $\epsilon > 0$. Since h is continuous and $h(\hat{u}) > \underline{c}$, for all 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. If $h(u) \leq h(\hat{u})$ for all u in that set, $\hat{u}$ would be a critical point of the restriction of h to $\mathcal{A}(\underline{c})$. Since $\hat{u}$ is not a critical point, 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 $h(\hat{u}) \geq \hat{c}$, this implies $h(u^\epsilon) > \hat{c}$. Recalling that $c^n \to \hat{c}$, for n sufficiently large we then have $h(u^\epsilon) > c^n$ and, therefore, $u^\epsilon \in \mathcal{A}(c^n)$. So, recalling (B.1), for n sufficiently large we have $\|u^n - \hat{u}\| \leq \|u^\epsilon - \hat{u}\| < \epsilon$.

(b) 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 $\mathcal{A}^2$. Without loss let u^* lie in $\mathcal{A}^1$. By continuity of h, $\mathcal{A}(c)$ is a compact subset of $\mathbb{R}^J$. This requires that $\mathcal{A}^2$ be a compact subset of $\mathbb{R}^J$ as well.[56] The restriction of h to $\mathcal{A}^2$ must therefore attain a maximum at some point(s) u^{**}, which must be on the interior of $\mathcal{A}(c)$. Any such u^{**} would be another critical point of h on $\mathcal{A}(c)$. $\qquad\square$

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

Lemma 13. *Condition M′ implies local rectangle regularity.*

Proof. Let S, $\underline{c}$ and $\overline{c}$ 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 $c^0 \in (\underline{c}, \overline{c})$ such that $\mathcal{A}(c^0; S) \subset \mathcal{U} \subset \mathcal{B}(u^*, \epsilon)$ for some rectangle $\mathcal{U}$. Observe that by continuity of $\ln f$, $u^* \in \text{int}\,\mathcal{A}(c; S)$ for all $c^0 \in (\underline{c}, \overline{c})$. Because S is compact and $\ln f$ is continuous, $\mathcal{A}(c; S)$ is compact for all c. And by Lemma 12 (part (a)), $\mathcal{A}(c; S)$ is a continuous correspondence on $(\underline{c}, \overline{c}]$.[57] Thus $\max_{u \in \mathcal{A}(c;S)} u_j$ and $\min_{u \in \mathcal{A}(c;S)} u_j$ are

---

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

[56] Bounded is immediate. Suppose $\mathcal{A}^2$ is not closed: let $u \notin \mathcal{A}^2$ be a limit point of a sequence in $\mathcal{A}^2$. Since $\mathcal{A}(c)$ is closed, it must then be that $u \in \mathcal{A}^1$. But since u was a limit point of a sequence in $\mathcal{A}^2$, for all $\epsilon > 0$ there exists $\hat{u} \in \{\mathcal{B}(u, \epsilon) \cap \mathcal{A}(c)\}$ such that $\hat{u} \in \mathcal{A}^2$. Because $\mathcal{A}^1$ and $\mathcal{A}^2$ are disjoint, this requires $\hat{u} \notin \mathcal{A}^1$, contradicting openness of $\mathcal{A}^1$ relative to $\mathcal{A}(c)$.

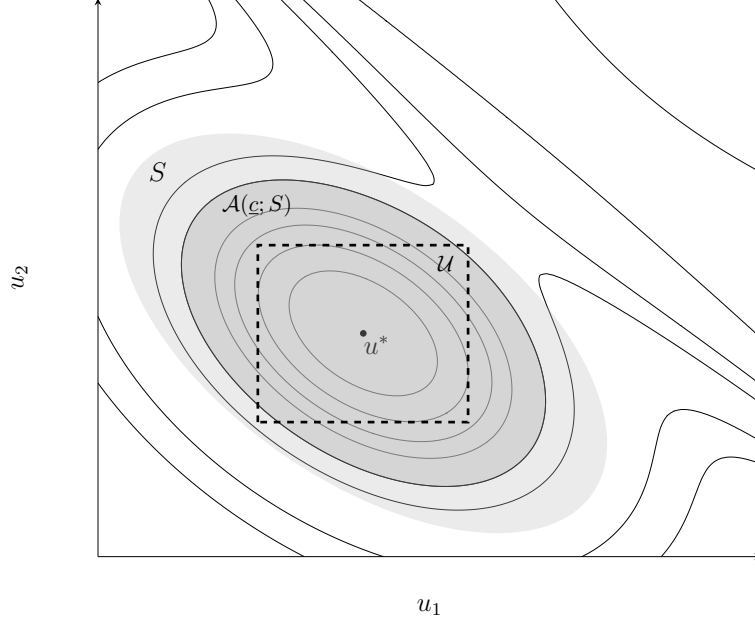[57] In Lemmas 11 and 12, let $\mathcal{A}(c) = \mathcal{A}(c; S)$ and let h be the restriction of $\ln f$ to S.

Figure 3: Curves show level sets of a bivariate log density in a region of its support. 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 $\ln f$ to S. The point u^* is a local max and the only critical point of $\ln f$ on $\mathcal{A}(\underline{c}; S)$. The rectangle $\mathcal{U}$ is defined by tangencies to the upper contour set $\mathcal{A}(c^0; S)$ for some $c^0 \in (\underline{c}, \ln f(u^*))$. Given any $\epsilon > 0$, we obtain $\mathcal{U} \in \mathcal{B}(u^*, \epsilon)$ by setting c^0 sufficiently close to $\ln f(u^*)$.

continuous in $c \in (\underline{c}, \overline{c}]$, implying continuity of the function

$$H(c) = \max_j \max_{\substack{u^+ \in \mathcal{A}(c;S) \\ u^- \in \mathcal{A}(c;S)}} u_j^+ - u_j^- \qquad c \in (\underline{c}, \overline{c}].$$

So, because $H(\overline{c}) = 0$, given any $\epsilon > 0$ there must exist $c^0 \in (\underline{c}, \overline{c})$ such that the rectangle

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

lies in $\mathcal{B}(u^*, \epsilon)$ (Lemma 11 ensures that each interval is nonempty). Thus, $\mathcal{A}(c^0; S) \subset \mathcal{U} \subset \mathcal{B}(u^*, \epsilon)$. To complete the proof, we show that $\ln f$ is regular on $\mathcal{U}$. By construction $u^* \in \mathcal{A}(c^0; S) \subset \mathcal{U}$. Now take arbitrary j and any $u_j \neq u_j^*$ such that $(u_j, u_{-j}) \in \mathcal{U}$ for some u_{-j}. By Lemma 12 (part b) and the definition of $\mathcal{U}$, there must also exist $\tilde{u}_{-j}$ such that $(u_j, \tilde{u}_{-j}) \in \mathcal{A}(c^0; S)$. Let $\hat{u}(j, u_j)$ solve

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

This solution must lie in $\mathcal{A}(c^0; S) \subset \mathcal{U}$ 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 $\partial \ln f(\hat{u}(j, u_j))/\partial u_j \neq 0$. $\qquad \square$

C   Other Proofs Omitted from the Text

Proof of Lemma 1.   With (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 $\mathbb{Y}$ 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 an open subset of $\mathbb{R}^J$. $\square$

Proof of Proposition 1.   Fix an arbitrary $x \in \mathbb{X}$. By (13),

$$\frac{\partial \phi(y^*\left(x\right)|x)}{\partial x_j} = 0 \tag{C.1}$$

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

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

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

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

Satisfaction of these conditions is observable. Thus, part (ii) is verifiable. $\square$

Proof of Lemma 4.   Recall that $d\left(x, y\right)^\mathsf{T} = \left(1, -\frac{\partial \ln \phi(y|x)}{\partial x_1}, \dots, -\frac{\partial \ln \phi(y|x)}{\partial x_J}\right)$. Suppose first that (32) holds for nonzero $c = \left(c_0, c_1, \dots, c_J\right)^\mathsf{T}$. Differentiating (32) with respect to x yields (33), with $\tilde{c} = \left(c_1, \dots, c_J\right)^\mathsf{T}$. 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 $d\left(x, y\right)$ is nonzero and $d\left(x, y\right)^\mathsf{T} c = 0$, we must have $c_j \neq 0$ for some $j > 0$. Thus (33) must hold for some nonzero $\tilde{c}$. Now suppose (33) holds for nonzero $\tilde{c} = \left(c_1, \dots, c_J\right)^\mathsf{T}$. 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 $c = \left(c_0, c_1, \dots, c_J\right)^\mathsf{T}$, $d\left(x^0, y\right)^\mathsf{T} c = 0$ by construction. Since the first component of $d\left(x, y\right)$ equals 1 for all $\left(x, y\right)$ and $\mathbb{X}$ is an open connected subset of $\mathbb{R}^J$, (33) implies that $\frac{\partial}{\partial x_j}\left[d\left(x, y\right)^\mathsf{T} c\right] = 0$ for all j and every $x \in \mathbb{X}$. Thus (32) holds for some nonzero c. $\square$

Proof of Claim 5.   Fix any $\epsilon > 0$. From (38),

$$\ln f_\lambda\left(u\right) - \ln f\left(u\right) = \lambda_{\tau(u)} \prod_j p\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right) + \ln \kappa. \tag{C.2}$$

So because $\ln \kappa \in [-\lambda, 0]$ (see (39) and the discussion that follows in the text) and $\lambda_{\tau(u)} \prod_j p\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right) \in [0, \lambda]$,

$$\sup_{u \in \mathbb{R}^J} |\ln f_\lambda(u) - \ln f(u)| \le \lambda. \tag{C.3}$$

Further, differentiating (C.2) (see Appendix D) we have

$$\frac{\partial \ln f_\lambda(u)}{\partial u_j} - \frac{\partial \ln f(u)}{\partial u_j} = \lambda_{\tau(u)} \left[\prod_{\ell \ne j} p\left(\frac{u_\ell - \dot{u}_{\tau(u)\ell}}{w/2}\right)\right] p'\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right) \left(\frac{w}{2}\right)^{-1} \tag{C.4}$$

$$\frac{\partial^2 \ln f_\lambda(u)}{\partial u_j^2} - \frac{\partial^2 \ln f(u)}{\partial u_j^2} = \lambda_{\tau(u)} \left[\prod_{\ell \ne j} p\left(\frac{u_\ell - \dot{u}_{\tau(u)\ell}}{w/2}\right)\right] p''\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right) \left(\frac{w}{2}\right)^{-2} \tag{C.5}$$

while for $j \ne k$,

$$\frac{\partial^2 \ln f_\lambda(u)}{\partial u_k \partial u_j} - \frac{\partial^2 \ln f(u)}{\partial u_k \partial u_j} = \lambda_{\tau(u)} \left[\prod_{\ell \ne j, k} p\left(\frac{u_\ell - \dot{u}_{\tau(u)\ell}}{w/2}\right)\right] p'\left(\frac{u_k - \dot{u}_{\tau(u)k}}{w/2}\right) p'\left(\frac{u_j - \dot{u}_{\tau(u)j}}{w/2}\right) \left(\frac{w}{2}\right)^{-2}. \tag{C.6}$$

The function p is bounded, as are its first and second derivatives (see Appendix D). So because $\lambda_{\tau(u)} \in (0, \lambda]$, (C.3)–(C.6) demonstrate that

$$\sup_{u \in \mathbb{R}^J} |\ln f_\lambda(u) - \ln f(u)| + \max_{j \in \{1,\dots,J\}} \sup_{u \in \mathbb{R}^J} \left|\frac{\partial \ln f_\lambda(u)}{\partial u_j} - \frac{\partial \ln f(u)}{\partial u_j}\right| +$$
$$\max_{j,k \in \{1,\dots,J\}} \sup_{u \in \mathbb{R}^J} \left|\frac{\partial^2 \ln f_\lambda(u)}{\partial u_j \partial u_k} - \frac{\partial^2 \ln f(u)}{\partial u_j \partial u_k}\right| \quad < \quad \epsilon$$

for all sufficiently small $\lambda > 0$. $\qquad\square$

Proof of Lemma 6. For any $\mathfrak{h} \in C^2(K)$ and $u \in K$ define $\rho_{\mathfrak{h}}(u) = \sum_j \left|\frac{\partial \mathfrak{h}(u)}{\partial u_j}\right| + \left|\det\left(\frac{\partial^2 \mathfrak{h}(u)}{\partial u \partial u^\mathsf{T}}\right)\right|$. A function $\mathfrak{h} \in C^2(K)$ has no degenerate critical point on K if and only if $\rho_{\mathfrak{h}}(u) > 0$ for all $u \in K$. So by the hypothesis of the Lemma, $\rho_{\mathfrak{f}}(u) > 0$ for all $u \in K$. Because K is compact and $\rho_{\mathfrak{f}}$ is continuous, there must exist $\delta > 0$ such that $\rho_{\mathfrak{f}}(u) > \delta$ for all $u \in K$. If $||\mathfrak{f} - \mathfrak{g}||_{C_K^2} < \epsilon$, then

$$\left|\frac{\partial \mathfrak{f}(u)}{\partial u_j} - \frac{\partial \mathfrak{g}(u)}{\partial u_j}\right| \quad < \quad \epsilon \quad \forall j, \forall u \in K$$

$$\left|\frac{\partial^2 \mathfrak{f}(u)}{\partial u_j \partial u_k} - \frac{\partial^2 \mathfrak{g}(u)}{\partial u_j \partial u_k}\right| \quad < \quad \epsilon \quad \forall j, k, \forall u \in K.$$

For sufficiently small $\epsilon > 0$ these imply

$$\sum_j \left|\left|\frac{\partial \mathfrak{f}(u)}{\partial u_j}\right| - \left|\frac{\partial \mathfrak{g}(u)}{\partial u_j}\right|\right| \quad < \quad \frac{\delta}{2} \qquad \forall u \in K$$

$$\left|\left|\det\left(\frac{\partial^2 \mathfrak{f}(x)}{\partial x \partial x^\mathsf{T}}\right)\right| - \left|\det\left(\frac{\partial^2 \mathfrak{g}(u)}{\partial u \partial u^\mathsf{T}}\right)\right|\right| \quad < \quad \frac{\delta}{2} \qquad \forall u \in K$$

33

which require

$$\sum_j \left| \frac{\partial \mathfrak{g}(u)}{\partial u_j} \right| > \sum_j \left| \frac{\partial \mathfrak{f}(u)}{\partial u_j} \right| - \frac{\delta}{2} \qquad \forall u \in K \qquad (C.7)$$

$$\left| \det \left(\frac{\partial^2 \mathfrak{g}(u)}{\partial u \partial u^\intercal} \right) \right| > \left| \det \left(\frac{\partial^2 \mathfrak{f}(x)}{\partial x \partial x^\intercal} \right) \right| - \frac{\delta}{2} \qquad \forall u \in K \qquad (C.8)$$

Summing (C.7) and (C.8), for all $u \in K$ we have $\rho_\mathfrak{g}(u) > \rho_\mathfrak{f}(u) - \delta > 0$. $\qquad\square$

D Triweight Perturbation on a Square

The proof of Lemma 5 uses a particular perturbation of a log density on a square in $\mathbb{R}^J$. Here we provide some additional discussion of this perturbation and derive some elementary properties referenced in the proof.

Recall that for $v \in \mathbb{R}$ we defined

$$p(v) = 1\{|v| \le 1\}(1 - v^2)^3.$$

The function $p(v)$ is equal to zero at -1 and 1, strictly increasing for $v \in (-1, 0)$, and strictly decreasing for $v \in (0, 1)$. It attains a maximum (of 1) at $v = 0$. For $v \in [-1, 1]$, its first and second derivatives are given by

$$p'(v) = -6v\left(1 - v^2\right)^2$$
$$p''(v) = 24v^2\left(1 - v^2\right) - 6(1 - v^2)^2,$$

which are continuous and bounded. The first and second derivatives of p at -1, 0, and 1 are given in Table 1 below.

Table 1: Some Values of $p(v)$ and Its Derivatives

v	$p(v)$	$p'(v)$	$p''(v)$
-1	0	0	0
0	1	0	-6
1	0	0	0

Let s denote a (closed) square $\times_j \left[\underline{u}_j, \overline{u}_j\right]$ in $\mathbb{R}^J$, with $\overline{u}_j - \underline{u}_j = \overline{w} > 0$ for all $j = 1, \ldots, J$. Let $\dot{u}_s = \left(\frac{\overline{u}_1 + \underline{u}_1}{2}, \ldots, \frac{\overline{u}_J + \underline{u}_J}{2} \right)$ denote the center of this square. Let $\ln f$ be a twice continuously differentiable log density defined on $\mathbb{R}^J$, with $f = \exp(\ln f)$ its associated probability density function. Given any finite scalar $\lambda_s > 0$ and $\kappa > 0$, let

$$f_{\lambda_s}(u) = \kappa f(u) \exp\left[\lambda_s \prod_{j=1}^J p\left(\frac{u_j - \dot{u}_{sj}}{\overline{w}/2}\right)\right] \qquad u \in s.$$

34

Then, on the square s, $\ln f_{\lambda_s}$ is equal to the sum of $\ln f$, the rescaled multivariate triweight function $\lambda_s \prod_{j=1}^{J} p\left(\frac{u_j - \dot{u}_{sj}}{\bar{w}/2}\right)$, and the constant $\ln(\kappa)$. Observe that $\lambda_s \prod_{j=1}^{J} p\left(\frac{u_j - \dot{u}_{sj}}{\bar{w}/2}\right)$ takes values in the interval $[0, \lambda_s]$, attaining λ_s only at the center of the square. Figure 4 illustrates the scaled multivariate triweight function for the case $J = 2$ with $\lambda_s = 1$ and $\kappa = 0$.

Figure 4: Plot of a Scaled Bivariate Triweight Function
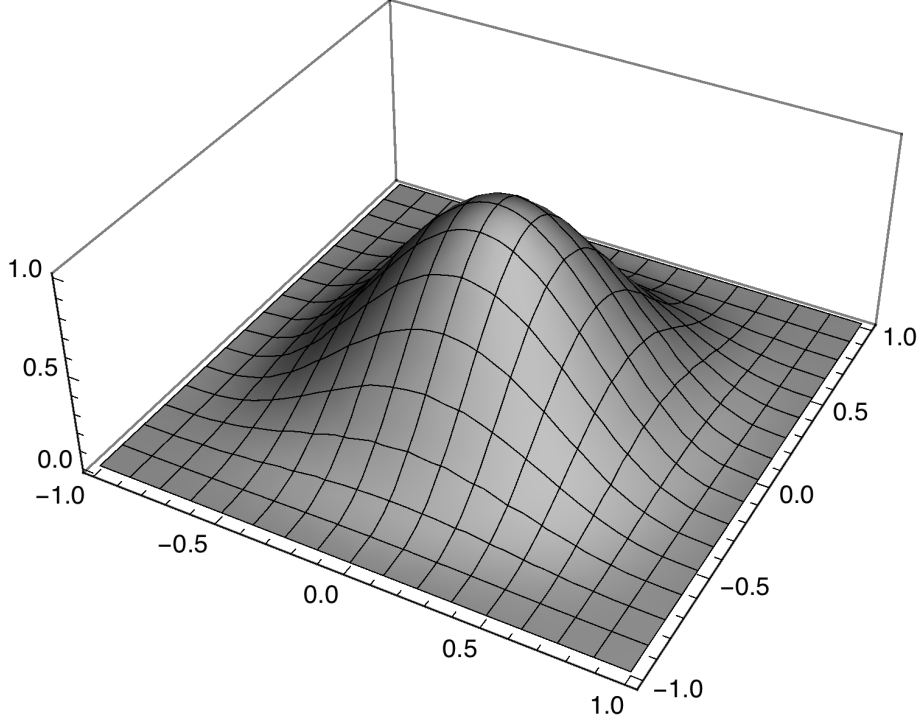
Recalling Table 1, observe that, regardless of λ_s, for any u on the boundary of the square s we have

$$
\begin{aligned}
\ln f_{\lambda_s}(u) &= \ln f(u) + \ln(\kappa) \\
\frac{\partial}{\partial u_j} \ln f_{\lambda_s}(u) &= \frac{\partial}{\partial u_j} \ln f(u) \quad \forall j \\
\frac{\partial^2}{\partial u_j \partial u_k} \ln f_{\lambda_s}(u) &= \frac{\partial^2}{\partial u_j \partial u_k} \ln f(u) \quad \forall j, k.
\end{aligned}
$$

These properties ensure that when we perturb $\ln f$ on adjacent squares—potentially with different scaling factors λ_s for each square (but the same κ for all squares)—the perturbed log density function will remain twice continuously differentiable, even on the boundaries of the squares.

E   Further Falsifiability Results

Here we provide additional falsifiability results—two for the linear index model and one for the full model. First, recalling (29) and the definition

$$b_k\left(y\right) = \left(\frac{\partial \ln|\mathsf{J}\left(y\right)|}{\partial y_k}, \frac{\partial r_1(y)}{\partial y_k}, \ldots, \frac{\partial r_J(y)}{\partial y_k}\right)^{\mathsf{T}},$$

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

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

Proposition 6. *Under the hypotheses of Theorem 3, the model defined by (24) and Assumption 1 is falsifiable.*

Suppose now that 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 $y \in \mathbb{Y}$, $\mathbb{X}$ contains two sets of points, $\tilde{\mathbf{x}} = \left(\tilde{x}^0, \ldots, \tilde{x}^J\right)^{\mathsf{T}}$ and $\tilde{\tilde{\mathbf{x}}} = \left(\tilde{\tilde{x}}^0, \ldots, \tilde{\tilde{x}}^J\right)^{\mathsf{T}}$, such that (i) $\tilde{\mathbf{x}} \neq \tilde{\tilde{\mathbf{x}}}$ and (ii) $D\left(\tilde{\mathbf{x}}, y\right)$ and $D\left(\tilde{\tilde{\mathbf{x}}}, y\right)$ have full rank. Then the model defined by (24) and Assumption 1 is falsifiable.*

Proof. By Lemma 3, $\partial r(y)/\partial y_k$ is identified for all k using the derivatives of $\phi\left(y|x\right)$ at points x in $\tilde{\mathbf{x}}$ (only) or in $\tilde{\tilde{\mathbf{x}}}$ (only). Letting $\partial r(y)/\partial y_k\left[\tilde{\mathbf{x}}\right]$ 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 $\partial r(y)/\partial y_k\left[\tilde{\mathbf{x}}\right] = \partial r(y)/\partial y_k\left[\tilde{\tilde{\mathbf{x}}}\right]$ 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 $x \in \mathbb{X}$ and the associated $y^*\left(x\right)$ defined by (17). It was then demonstrated that for some open rectangle $\mathcal{X}\left(x\right) \ni x$ the ratios

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

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 (assuming the model is correctly specified) $r(y^{**}(\tilde{x})) = g(\tilde{x}) + u^{**}$ 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 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 $(x^0, x') \in \{\mathcal{X}(x) \cap \mathcal{X}(\tilde{x})\}$. $\qquad\qquad\square$

F   Differenced Derivatives

In section 4.2 we relied on second derivatives of the log density $\ln f$. It is straightforward to extend our arguments to cases without twice differentiability, replacing any matrix of second derivatives with differences of the first derivatives. To see this, suppose that (32) holds for some nonzero $c = (c_0, c_1, \ldots, c_J)$. This implies that for any x and x' in $\mathbb{X}$,

$$[d(y, x) - d(y, x')]^\mathsf{T} c = 0.$$

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

$$\begin{bmatrix} \frac{\partial \ln \phi(y|x)}{\partial x_1} - \frac{\partial \ln \phi(y|x')}{\partial x_1} \\ \vdots \\ \frac{\partial \ln \phi(y|x)}{\partial x_J} - \frac{\partial \ln \phi(y|x')}{\partial x_J} \end{bmatrix}^\mathsf{T} \tilde{c} = 0 \qquad \forall x \in \mathbb{X},\, x' \in \mathbb{X}, \tag{F.1}$$

where $\tilde{c} = (c_1, \ldots, c_J)$. Thus, for identification to (possibly) fail there must exist a nonzero vector $\tilde{c}$ satisfying (F.1).

G   Discrete Support

In this appendix we relax Assumption 1 (retaining only its requirements of injective r and exogenous X) in order to provide an initial exploration of identification when X and U have discrete support. Here we limit attention to the case of a linear index function. Let f now denote the probability mass function for U. Let $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{Y}$ denote the supports of X, U, and Y, respectively.

G.1   Point Identification with Large Discrete Support

We first consider a version of our Theorem 2, replacing its assumption

$$\mathcal{X} = \mathcal{U} = \mathbb{R}^J \tag{G.1}$$

with

$$\mathcal{X} = \mathcal{U} = \mathbb{Z}^J. \tag{G.2}$$

Thus we now have discrete supports for X and U but retain the assumption that variation in X can compensate one-for-one for all variation in U. The support condition (G.2) may be more appealing than (G.1); however, the following result demonstrates that continuous variation in the instruments is not itself essential, even for point identification.

Proposition 9. *Suppose r is injective, U is independent of X, and $\mathcal{X} = \mathcal{U} = \mathbb{Z}^J$. Then in the linear index model, r is identified on $\mathcal{Y}$.*

Proof. The proof is nearly identical to that of Theorem 2. Here the change of variables simplifies to

$$\Pr(y|x) = f(r(y) - x). \tag{G.3}$$

The LHS of (G.3) is observed for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. Letting $F_1(\cdot)$ denote the marginal CDF of U_1, for any $y \in \mathcal{Y}$ we have

$$\lim_{T \to \infty} \sum_{\hat{x}_1 = x_1}^{T} \sum_{\hat{x}_2 = -T}^{T} \cdots \sum_{\hat{x}_J = -T}^{T} \Pr(y|\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_J) = F_1(r_1(y) - x_1).$$

So the value of $F_1(r_1(y) - x_1)$ is identified for all $y \in \mathcal{Y}$ and x_1 in its support. Recall that by our normalizations,

$$r_1(\dot{y}) - \dot{x}_1 = 0 \tag{G.4}$$

for known (arbitrary) $\dot{y}$ and $\dot{x}_1$. Thus, for any y we can find the unique value $\overset{o}{x}_1(y)$ such that

$$F_1\left(r_1(y) - \overset{o}{x}_1(y)\right) = F_1(0),$$

implying

$$r_1(y) = x_1^0(y).$$

An analogous argument applies to all $j \neq 1$, yielding identification of $r_j(y)$ for all y and j at all y. $\qquad\square$

G.2 Partial Identification with an Injective PMF

Now consider arbitrary finite support for X and U. Let $\mathcal{U} = (u^1, \ldots, u^K)$, with $K < \infty$. Let $f^k = f(u^k)$ and suppose that f is injective.[58] Let $\mathcal{Y}_k$ denote the support of $Y|\{U = u^k\}$.

   Once again, the change of variables takes the form (G.3). So by injectivity of f, for every x and every $k = 1, \ldots K$ we can find the (unique) vector $y^k(x)$ such that

$$f\left(r\left(y^k(x)\right) - x\right) = f^k,$$

implying

$$r\left(y^k(x)\right) - x = u^k. \tag{G.5}$$

Although the value of each u^k is unknown, for any $\hat{x}$ and x in $y^k(\mathcal{X})$ we can take differences of (G.5) to obtain

$$r_j\left(y^k(\hat{x})\right) - r\left(y^k(x)\right) = \hat{x}_j - x_j.$$

For every k, this gives point identification of the first differences of r on $\mathcal{Y}_k$. This can be interpreted as identification of approximations to the first derivatives of r at each point in $\mathcal{Y}$ (cf. Lemma 3).

   If $\mathcal{Y}_k \cap \mathcal{Y}_{k'}$ is nonempty for some $k' \neq k$, the first differences on $\mathcal{Y}_k$ and $\mathcal{Y}_{k'}$ can be linked to deliver identification of the first differences within on a larger subset of $\mathcal{Y}$. In some cases this can again yield point identification of r on all of $\mathcal{Y}$. Suppose, for example, that there is a permutation (potentially with repetition) $p = 1, \ldots, P$ of the indices $k = 1, \ldots, K$ such that $\mathcal{Y}_p \cap \mathcal{Y}_{p+1}$ is nonempty for each $p = 1, \ldots, P - 1$ (this hypothesis is verifiable under the injectivity of f). Then we have identification of the first differences of r on $\cup_k \mathcal{Y}_k$, i.e., on $\mathcal{Y}$. Recalling the location normalization $r_j(\dot{y}) - \dot{x}_j = 0$ for all j, this implies identification of r.

---

[58]Injectivity is generic in the set of K-dimensional probability mass functions. The set $\mathcal{M}$ of probability mass functions f with support $(u^1, \ldots, u^K)$ is the relative interior of the $(K-1)$-simplex. The subset $\hat{\mathcal{M}} \subset \mathcal{M}$ for which $\hat{f}(u^k) = \hat{f}(u^{k'})$ for at least one pair $(k, k' \neq k)$ is a closed (relative to $\mathcal{M}$) set such that for any $\epsilon > 0$ and any $\hat{f} \in \mathcal{M}$, there exists $f \in \{\mathcal{M} \backslash \hat{\mathcal{M}}\}$ satisfying $\|\hat{f} - f\| < \epsilon$. (For example, with $K = 2$, $\mathcal{M}$ is the relative interior of a line segment in $\mathbb{R}^2$ and $\hat{\mathcal{M}}$ is a point on this segment. With $K = 3$, $\mathcal{M}$ is the relative interior of a triangle in $\mathbb{R}^3$ and $\hat{\mathcal{M}}$ is the union of three half-open line segments on the face of the triangle). Thus, the set $\{\mathcal{M} \backslash \hat{\mathcal{M}}\}$ is an open (relative to $\mathcal{M}$) dense subset of $\mathcal{M}$.

References

ANDERSON, R. M., AND W. R. ZAME (2001): "Genericity with Infinitely Many Parameters," *Advances in Economic Theory, B.E. Journals in Economics*, 1, 1–62.

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

BANYAGA, A., AND D. HURTUBISE (2004): *Lectures on Morse Homology.* Springer Science+Business Media, Dordrecht.

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.

——— (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.

CHESHER, A., AND A. ROSEN (2015): "Characterizations of identified sets delivered by structural econometric models," working paper CWP 63/15, Cemmap, Institute for Fiscal Studies UCL.

CHIAPPORI, P.-A., AND I. KOMUNJER (2009a): "Correct Specification and Identification of Nonparametric Transformation Models," Discussion paper, University of California San Diego.

———— (2009b): "On the Nonparametric Identification of Multiple Choice Models," Discussion paper, University of California San Diego.

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 14.* 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.

HUNT, B. R., T. SAUER, AND J. A. YORKE (1992): "Prevalence: A Translation-Invariant "Almost Every" On Infinite-Dimensional Spaces," *Bulletin of the American Mathematical Society*, 27(2), 217–238.

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

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

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

KOOPMANS, T. C., AND O. REIERSOL (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21, 165–181.

KOOPMANS, T. C., H. RUBIN, AND R. B. LEIPNIK (1950): "Measuring the Equation Systems of Dynamic Economics," in *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10*, ed. by T. C. Koopmans. 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, Part 2," *Sankhya*, 15, 219–236.

MAS-COLELL, A. (1985): *The Theory of General Equilibrium: A Differentiable Approach*, Econometric Society Monographs. Cambridge University Press, Cambridge.

MASTEN, M. (2015): "Random Coefficients on Endogenous Variables in Simultaneous Equations Models," Discussion paper, CEMMAP working paper CWP25/15.

MATSUMOTO, Y. (2002): *An Introduction to Morse Theory*. American Mathematical Society, Providence, RI, Translation by Kiki Hudson and Masahico Saito.

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

———— (2015): "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1–66.

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.

SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

STINCHCOMBE, M. B. (2002): "Some Genericity Analysis in Nonparametric Statistics," Discussion paper, University of Texas.

TORGOVITZKY, A. (2015): "Identification of Nonseparable Models Using Instruments with Small Support," *Econometrica*, 83, 1185–1197.

VAN MILL, J. (2002): *The Infinite-Dimensional Topology of Function Spaces*. Elsevier.