



**Discussion Papers in Economics**

**A GENERAL RESULT ON OBSERVATIONAL EQUIVALENCE  
IN A CLASS OF NONPARAMETRIC STRUCTURAL EQUATIONS  
MODELS**

By

Giovanni Forchini  
(University of Surrey)

DP 01/14

Department of Economics  
University of Surrey  
Guildford  
Surrey GU2 7XH, UK  
Telephone +44 (0)1483 689380  
Facsimile +44 (0)1483 689548  
Web [www.econ.surrey.ac.uk](http://www.econ.surrey.ac.uk)  
ISSN: 1749-5075

# A General Result on Observational Equivalence in a Class of Nonparametric Structural Equations Models

G. Forchini<sup>12</sup>

University of Surrey

9 June 2014

## Abstract

This paper examines observational equivalence in a class of nonparametric structural equations models under weaker conditions than those currently available in the literature. It allows for several endogenous variables, does not impose differentiability or continuity of the equations which are part of the structure, and allows the unobserved errors to depend on the exogenous variables. The usefulness of the main result is illustrated by deriving observational equivalence conditions for some models including nonparametric simultaneous equations models, additive errors models, multivariate triangular models, etc.. Some of these yield well known results as special cases.

---

<sup>1</sup> Address for correspondence: Giovanni Forchini, School of Economics, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom. E-mail: [G.Forchini@surrey.ac.uk](mailto:G.Forchini@surrey.ac.uk). This research was partially supported by Australian Research Council grant DP0985432.

<sup>2</sup> I thank Federico Martellosio and Tang Srisuma for helpful comments.

# 1. Introduction

A fundamental concern for econometricians has been the possibility that some features of interest of a model may not be identified (see among others the excellent reviews of Hsiao (1983) and Matzkin (2007), and references therein). For a long time, great effort was devoted to understanding the conditions under which the coefficients of a system of linear simultaneous equations are identified (e.g. Hsiao (1983)). However, more recently, the interest has shifted to non-linear and non-parametric models. The study of the latter is considerably more difficult than that of a system of linear equations especially when the error terms do not appear in the models in an additive form. Seminal contributions include work by Brown (1983) and Roehrig (1988) which has been recently critically re-examined by Benkard and Berry (2006) and revisited by Matzkin (2008). See also Chesher (2003), Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007), Hoderlein and Mammen (2007), Matzkin (2008) and Imbens and Newey (2009)). Matzkin (2007) provides an excellent survey of existing results.

The classical approach to determine the identification of a feature of interest of a model is based on two steps. Firstly, a class of alternative models that produce the same distribution of the observed variables - and are thus observationally equivalent - is identified. Secondly, the researcher checks whether the feature of interest varies within the set of observationally equivalent models. If it does not change, the feature is identified, otherwise it is not. Methods used to establish identification are very varied and involve for example the conditional quantile approach (e.g. Chesher (2003), Matzkin (2003)), the control function approach (e.g. Imbens and Newey (2009)) as well as more direct approaches such as that of Matzkin (2008), (2010) and Berry and Haile (2013).

This paper derives a necessary and sufficient condition for the observational equivalence of two structures in a class of models introduced by Matzkin (2008) and further studied by Matzkin (2010) and Berry and Haile (2013). The strategy used in these papers has been to relate the conditional density of the endogenous variables conditional on the exogenous variables to the unobservable errors through a change of variable argument. This relationship is then used to understand what aspects of the model are identified.

This paper uses a different approach which exploits fundamental relationships between measurable functions expressed by the factorization lemma. Our conditions are weaker than those currently available in the literature because we allow for several endogenous variables, do not impose differentiability or continuity of the equations which are part of the structure, and allow the unobserved errors to depend on the exogenous variables. We illustrate the main result deriving observational equivalence conditions for some classes of models including nonparametric simultaneous equations models, additive models, multivariate triangular models, etc.. Some of them yield well known results as special cases but under weaker conditions.

The rest of this paper is organized as follows. Section 2 defines the set-up and derives the main result. Section 3 considers some special cases where the equations which are part of the structure are differentiable, and Section 4 concludes.

## 2. Model and main result

Suppose that an observable  $n$  dimensional continuous random vector  $y$  depends on an observable  $k$  dimensional random vector  $z$  and an unobservable  $n$  dimensional continuous random vector  $u$  through a relationship of the form (e.g. Matzkin (2008), (2010) and Berry and Haile (2013))

$$(1) \quad \eta(z, y) = u,$$

where  $\eta: \mathbb{R}^k \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $u$  has probability distribution  $P_{u|z}$  with respect to the Borel sigma-algebra  $\mathcal{U}^n = \mathcal{B}(\mathbb{R}^n)$ . Notice that the function  $\eta$  can be defined on a subset of  $\mathbb{R}^k \times \mathbb{R}^n$  to a subset of  $\mathbb{R}^n$  but to keep the notation as simple as possible we will not make this explicit. For each  $z \in \mathbb{R}^k$ , we regard  $\eta$  as a  $\mathcal{Y}^n / \mathcal{U}^n$  measurable map from the measurable space  $(\mathbb{R}^n, \mathcal{Y}^n)$  to  $(\mathbb{R}^n, \mathcal{U}^n)$  where the sigma-algebra  $\mathcal{Y}^n$  is such that  $Y \in \mathcal{Y}^n$  if  $\eta(z, Y) \in \mathcal{U}^n$ . We define the structure for  $y$  as the pair  $S = (\eta, P_{u|z})$ , and for each  $Y \in \mathcal{Y}^n$ ,  $P_{y|z, S}(Y) = P_{u|z}(\eta(z, Y))$  a.e. in  $z$ . Notice that the dimension of the unobservable random variable  $u$  is the same as the dimension of  $y$ .

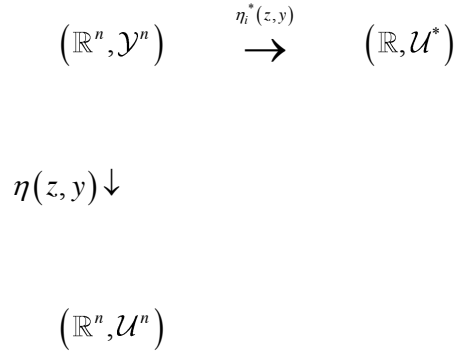
If the function  $\eta_z(y): \mathbb{R}^n \rightarrow \mathbb{R}^n$  where  $\eta_z(y) = \eta(z, y)$  is invertible, it is possible to “solve” for  $y$  in (1) and obtain the *reduced form*  $y = m(z, u)$  (e.g. Matzkin (2003)). Although we do not make use of this assumption, it is important to notice that it allows models of the form  $y = m(z, u)$  to be written as in equation (1). This is the case for the models studied by Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007) and Matzkin (2003). Other structures described by stochastic equations having the same form as equation (1) are: linear models and the additive error models of Newey, Powell and Vella (1999) and Newey and Powell (2003), some demand and supply models (e.g. Matzkin (2008)), models of differentiated product markets (e.g. Berry and Haile (2013)) and models of production functions in the presence of unobserved shocks to the marginal product of each input (e.g. Berry and Haile (2013)).

Notice that we allow the distribution of  $u$  to depend on  $z$ . One could argue that if this is the case, it may be possible to find another random variable  $\varepsilon$  - in general not necessarily of dimension  $n$  - having a distribution which does not depend on  $z$  such that  $u = s(z, \varepsilon)$  (e.g. Matzkin (2003)). However, in order to write the model as in equation (1) - with the observable endogenous and exogenous variables on the left-hand side and the unobserved errors  $\varepsilon$  on the right-hand side of (1) -

one would need  $s_z: \mathbb{R}^{\varepsilon} \rightarrow \mathbb{R}^n$  where  $s_z(\varepsilon) = s(z, \varepsilon)$  to be invertible, and this is not the case in general. Therefore, we will not impose this invertibility assumption in our main result. Notice, however, that the model considered allows  $u$  to satisfy  $u = s(z, \varepsilon)$  where the dimension of  $\varepsilon$  may even be larger than that of  $y$ . This case is also of interest (c.f. Theorem 3 of Matzkin (2003)).

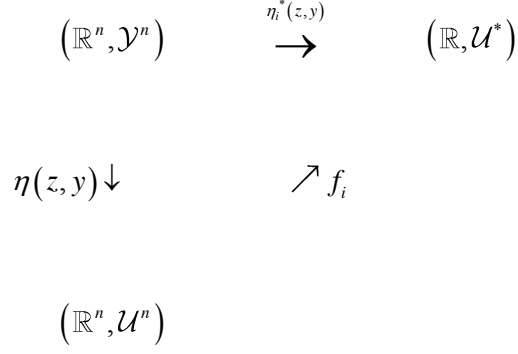
We now investigate the existence of another structure  $S^* = (\eta^*, P_{u^*|z})$  which is observationally equivalent to  $S = (\eta, P_{u|z})$  in the sense that  $P_{y|z, S}(Y) = P_{y|z, S^*}(Y)$  for all  $Y \in \mathcal{Y}^n$  a.e. in  $z$ . Notice that we need the map  $\eta^*: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n$  from  $(\mathbb{R}^n, \mathcal{Y}^n)$  to  $(\mathbb{R}^n, \mathcal{U}^{*n})$  to be  $\mathcal{Y}^n / \mathcal{U}^{*n}$  measurable so that  $Y \in \mathcal{Y}^n$  if  $\eta^*(z, Y) \in \mathcal{U}^{*n}$ . In this case  $\mathcal{U}^{*n} = \mathcal{B}(\mathbb{R}^n)$  is the Borel sigma-algebra on  $\mathbb{R}^n$ . The random variable  $u^*$  is defined by  $\eta^*(z, y) = u^*$ . In this notation we have also  $P_{y|z, S^*}(Y) = P_{u^*|z}(\eta^*(z, Y))$ .

Let  $\eta_i^*(z, y)$  denote the  $i$ -th component of  $\eta^*(z, y)$ . The following Diagram 1 illustrates the relationship between the quantities described above.



**Diagram 1**

The function  $\eta^*(z, y)$  must be  $\mathcal{Y}^n / \mathcal{U}^{*n}$  measurable for every  $z$ . It follows from the Factorization Lemma (e.g. Lemma 2.1 of Lehmann (1997)) that the  $i$ -th component of  $\eta^*(z, y)$ ,  $\eta_i^*(z, y)$ , is  $\mathcal{Y}^n / \mathcal{U}^*$  measurable if and only if there is a  $\mathcal{U}^n / \mathcal{U}^*$  measurable function  $f_i$  such that  $\eta_i^*(z, y) = f_i(\eta(z, y))$ . Thus, we can complete Diagram 1 above.



**Diagram 2**

Repeating this argument for all components of  $\eta^*(z, y)$ , and noticing that for each  $z$ ,  $\eta^*(z, y)$  is a map between two Euclidean spaces, we can conclude that  $\mathcal{Y}^n / \mathcal{U}^{*n}$  - measurability of  $\eta^*(z, y)$  implies the existence of a  $\mathcal{U}^n / \mathcal{U}^{*n}$  -measurable transformation  $f$  such that  $\eta^*(z, y) = f(\eta(z, y))$  (e.g. Exercise 7.5 of Schilling (2005)), or equivalently  $u^* = f(u)$ .

Now, suppose that there is a  $\mathcal{U}^n / \mathcal{U}^{*n}$  measurable function  $f$  such that  $\eta^*(z, y) = f(\eta(z, y))$ . Then,

$$\begin{aligned}
P_{y|z, S^*}(Y) &= P_{u^*|z}(\eta^*(z, Y)) && \text{by definition} \\
&= P_{u^*|z}(f(\eta(z, Y))) && \text{by assumption} \\
&= P_{u|z}(\eta(z, Y)) && \text{by measurability of the function } f \\
&= P_{y|z, S}(Y) && \text{by definition}
\end{aligned}$$

and the structures  $S = (\eta, P_{u|z})$  and  $S^* = (\eta^*, P_{u^*|z})$  are observationally equivalent.

Notice that the argument above shows that the structure  $S^* = (\eta^*, P_{u^*|z})$  is observationally equivalent to the structure  $S = (\eta, P_{u|z})$  if and only if there is a  $\mathcal{U}^n / \mathcal{U}^{*n}$  -measurable transformation  $f$  such that  $\eta^*(z, y) = f(\eta(z, y))$ . However, one can reverse the argument and conclude that the structure  $S = (\eta, P_{u|z})$  is observationally equivalent to the structure  $S^* = (\eta^*, P_{u^*|z})$  if and only if there is a  $\mathcal{U}^{*n} / \mathcal{U}^n$  -measurable transformation  $g$  such that  $\eta(z, y) = g(\eta^*(z, y))$ . Thus, the function  $f$  is invertible with inverse  $f^{-1} = g$ , so that  $f$  is a bijection.

We can summarize the above result as follows:

**Theorem 1.** A necessary and sufficient condition for the existence of a structure  $S^* = (\eta^*, P_{u^*|z})$ , where  $\eta^*(z, y)$  is  $\mathcal{Y}^n/\mathcal{U}^{*n}$  measurable for every  $z$ , which is observationally equivalent to  $S = (\eta, P_{u|z})$ , is the existence of a  $\mathcal{U}^n/\mathcal{U}^{*n}$ -measurable transformation  $f$  such that  $\eta^*(z, y) = f(\eta(z, y))$ . Moreover, the function  $f$  is bijective.

Notice that our result depends only on the measurability of the functions  $\eta$  and  $\eta^*$ . One can certainly restrict the class of observationally equivalent models by imposing extra conditions on them: for example,  $\eta$  and  $\eta^*$  could be taken to be differentiable. One could also impose restrictions on the distribution of the errors. For example, Matzkin (2003) and (2008) requires it not to depend on  $z$ . Newey, Powell and Vella (1999) and Newey and Powell (2003) impose moment conditions on the errors. These restrict the models allowed and thus facilitate the study of identification, however, Theorem 1 suggests that these are in fact not necessary. Theorem 1 also suggests that the assumptions on the error terms  $u$  are not as relevant as one may expect in establishing identification. This is due to the fact that  $u$  is essentially *defined* by  $u = \eta(z, y)$ .

Lemma 1 of Matzkin (2003) is the closest result to Theorem 1 available in the literature. This applies to the case where  $n=1$  with  $\eta$  being continuous and  $u$  having a continuous distribution with strictly increasing probability distribution. These assumptions imply that the function  $f$  is continuous and strictly increasing (and thus bijective). Our Theorem 1 shows that Lemma 1 of Matzkin (2003) essentially holds even if  $\eta$  is not continuous and  $u$  has an arbitrary continuous distribution.

Results for the multivariate case are given by Matzkin (2008) assuming, among other things, that the function  $\eta$  is twice continuously differentiable and  $u$  has a continuous distribution independent of  $z$ , with a density function which is continuously differentiable and has support on  $\mathbb{R}^n$ . Matzkin (2008) shows that observational equivalence between two structures is equivalent to requiring independence between  $z$  and  $u$ . By examining the derivation of this result by Matzkin (2008) one may be led to think that this result depends on the assumptions of differentiability of the various functions involved. However, it is possible to establish a link between observational equivalence and independence of  $u$  and  $z$  in a more general context. Suppose that  $P_{u|z} = P_u$  (an assumption used by Matzkin (2008)), and consider a structure  $S = (\eta, P_u)$ . According to Theorem 1, if there is a structure  $S^* = (\eta^*, P_{u^*|z})$  which is observationally equivalent to  $S = (\eta, P_u)$ , there must exist a  $\mathcal{U}^n/\mathcal{U}^{*n}$ -measurable transformation  $f$  such that  $\eta^*(z, y) = f(\eta(z, y))$ . This is equivalent to stating that  $u^* = f(u)$ . Since neither  $f$  nor the distribution of  $u$  depend on  $z$ , we must have  $P_{u^*|z} = P_{u^*}$ . Thus, we obtain the following result.

**Corollary 1.** If the structure  $S^* = (\eta^*, P_{u^*|z})$  is observationally equivalent to the structure  $S = (\eta, P_u)$ , then  $P_{u^*|z} = P_u$ .

This result generalizes the “necessary” part of Theorem 3.2 of Matzkin (2008). To recover the “sufficiency” in the result of Matzkin (2008) one has to introduce stronger assumptions on the variables involved to essentially make sure that for  $u$  and  $u^*$  one can write  $u^* = f(u)$ . This is certainly the case if these random vectors have continuous distributions.

Finally, Theorem 1, allows us to generate all structures observationally equivalent to  $S = (\eta, P_{u|z})$ . To do this we need to consider all bijective measurable functions  $f$  from a suitable subset of  $\mathbb{R}^k \times \mathbb{R}^n$  to a suitable subset of  $\mathbb{R}^n$ . Then the class of structures observationally equivalent to  $S = (\eta, P_{u|z})$  is  $S_f = (f \circ \eta, P_{f(u)|z})$  where  $\circ$  denotes function composition. Moreover, every functional of  $\eta$  which is unchanged by the transformation  $\eta \rightarrow f \circ \eta$  is uniquely determined in the whole class of observationally equivalent models, and thus it is identified.

### 3. Applications

Here are examples illustrating how Theorem 1 can be used to obtain identification conditions for models of practical interest. For all these applications the distribution of the errors does not play any major role.

#### 3.1. Simultaneous equations model

The first application is the simultaneous equations model for which the functions  $\eta(z, y)$  and  $\eta^*(z, y)$  are differentiable. Since from Theorem 1, the structures  $S = (\eta, P_{u|z})$  and  $S^* = (\eta^*, P_{u^*|z})$  are observationally equivalent if and only if  $\eta^*(z, y) = f(\eta(z, y))$ ,  $f$  must also be differentiable. Differentiating both sides we have

$$(3) \quad D_z \eta^* \underset{(n \times k)}{\quad} (z, y) = D_x f \underset{(n \times n)}{\quad} (x) \Big|_{x=\eta(z, y)} D_z \eta \underset{(n \times k)}{\quad} (z, y)$$

and

$$(4) \quad D_y \eta^* \underset{(n \times n)}{\quad} (z, y) = D_x f \underset{(n \times n)}{\quad} (x) \Big|_{x=\eta(z, y)} D_y \eta \underset{(n \times n)}{\quad} (z, y).$$



Assuming  $\text{rank}(D_y \eta(z, y)) = n$  and  $\text{rank}(D_y \eta^*(z, y)) = n$ , we can write

$$(5) \quad (D_y \eta^*(z, y))^{-1} D_z \eta^*(z, y) = (D_y \eta(z, y))^{-1} D_z \eta(z, y).$$

Using the implicit function theorem we can conclude (5) equals  $-D_z y(z)$  which is the same for the two observationally equivalent structures and is thus identified.

If  $n = k = 1$ , the structures above reduce to that considered by Matzkin (2008) in Section 5.1. In this case, equation (5) simplifies to

$$\frac{\frac{\partial \eta^*(z, y)}{\partial z}}{\frac{\partial \eta^*(z, y)}{\partial y}} = \frac{\frac{\partial \eta(z, y)}{\partial z}}{\frac{\partial \eta(z, y)}{\partial y}},$$

which is the same condition for observational equivalence obtained by Matzkin (2008) (see her equation (5.1)) and for the identification of  $\frac{dy(z)}{dz}$ .

### 3.2. Additive errors models

Newey, Powell and Vella (1999) have considered the identification of the following triangular system of equations  $y_1 - m(y_2, z_1) = u_1$  and  $y_2 - \pi(z_1, z_2) = u_2$  where  $y_2 \in \mathbb{R}^{n-1}$ ,  $z_1 \in \mathbb{R}^{k_1}$  and  $z_2 \in \mathbb{R}^{k_2}$ . Consider another structure for which  $y_1 - m^*(y_2, z_1) = u_1^*$  and  $y_2 - \pi^*(z_1, z_2) = u_2^*$  which is observationally equivalent to the previous one. We can assume all functions are differentiable. From Theorem 1,

$$(6) \quad y_1 - m^*(y_2, z_1) = f_1(y_1 - m(y_2, z_1), y_2 - \pi(z_1, z_2)),$$

and

$$(7) \quad y_2 - \pi^*(z_1, z_2) = f_2(y_1 - m(y_2, z_1), y_2 - \pi(z_1, z_2)).$$

Differentiating (6) with respect to  $z_2$  one has

$$D_x f_1(y_1 - m(y_2, z_1), x) \Big|_{x=y_2 - \pi(z_1, z_2)} \Big|_{(1 \times n-1)} D_{z_2} \pi(z_1, z_2) \Big|_{(n-1 \times k_2)} = 0,$$

where the dimensions are indicated in brackets. If  $\text{rank}(D_{z_2}\pi(z_1, z_2)) = n-1$ , then

$D_x f_1(y_1 - m(y_2, z_1), x) \Big|_{x=y_2 - \pi(z_1, z_2)} = 0$ . Similarly, differentiating (7) with respect to  $y_1$ ,

$D_y f_2(y, y_2 - \pi(z_1, z_2)) \Big|_{y=y_1 - m(y_2, z_1)} = 0$ . Thus (6) and (7) can be written as:

$$(8) \quad y_1 - m^*(y_2, z_1) = f_1(y_1 - m(y_2, z_1))$$

and

$$(9) \quad y_2 - \pi^*(z_1, z_2) = f_2(y_2 - \pi(z_1, z_2)).$$

Differentiating the first equation with respect to  $y_1$  we find that the derivative is one, so that  $f_1$  has the form  $f_1(x) = x + \text{constant}$ , from which it follows that  $m^*(y_2, z_1) = m(y_2, z_1) + \text{constant}$ .

Differentiating (7) with respect to  $y_2$ , one finds that  $f_2(x) = x + \text{vector of constants}$  so that  $\pi^*(z_1, z_2) = \pi(z_1, z_2) + \text{vector of constants}$ . Thus,  $m$  and  $\pi$  are identified up to additive constants (e.g. Theorem 2.3 of Newey, Powell and Vella (1999) and Theorem 4.5 of Matzkin (2007)).

### 3.3. Triangular models

We now consider a triangular system of equations of the form  $\eta_1(y_1, y_2, z_1) = u_1$  and  $\eta_2(y_2, z_1, z_2) = u_2$  where  $y_1 \in \mathbb{R}$ ,  $y_2 \in \mathbb{R}^{n-1}$ ,  $z_1 \in \mathbb{R}^{k_1}$  and  $z_2 \in \mathbb{R}^{k_2}$ . To find an observationally equivalent structure we consider  $\eta_1^*(y_1, y_2, z_1) = f_1(\eta_1(y_1, y_2, z_1), \eta_2(y_2, z_1, z_2))$  and  $\eta_2^*(y_2, z_1, z_2) = f_2(\eta_1(y_1, y_2, z_1), \eta_2(y_2, z_1, z_2))$ . Once again we assume differentiability of all functions involved. Notice that if  $n=2$ ,  $k_1=0$  and  $k_2=1$ , this reduces to the model considered by Matzkin (2008) in Section 5.2:  $\eta_1(y_1, y_2) = u_1$  and  $\eta_2(y_2, z) = u_2$  where  $y_1, y_2, z, u_1, u_2 \in \mathbb{R}$ .

Since  $\eta_1^*(y_1, y_2, z_2)$  does not depend on  $z_2$ ,

$$D_{z_2} f_1(\eta_1(y_1, y_2, z_1), \eta_2(y_2, z_1, z_2)) = D_x f(\eta_1(y_1, y_2, z_1), x) \Big|_{x=\eta_2(y_2, z_1, z_2)} D_{z_2} \eta_2(y_2, z_1, z_2) = 0.$$

If  $\text{rank}(D_{z_2} \eta_2(y_2, z_1, z_2)) = n-1 \leq k_2$ , then one must have  $D_x f(\eta_1(y_1, y_2, z_1), x) \Big|_{x=\eta_2(y_2, z_1, z_2)} = 0$  so that

$\eta_1^*(y_1, y_2, z_1) = f_1(\eta_1(y_1, y_2, z_1))$ . Using a similar argument

$$D_{y_1} \eta_2^*(y_2, z_1, z_2) = D_x f_2(x, \eta_2(y_2, z_1, z_2)) \Big|_{x=\eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1} = 0.$$

Assuming  $\frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1} \neq 0$ ,  $D_{y_1} \eta_2^*(y_2, z_1, z_2)$  is a multiple of  $D_x f_2(x, \eta_2(y_2, z_1, z_2)) \Big|_{x=\eta_1(y_1, y_2, z_1)}$ ,

which must therefore be zero. Thus,  $\eta_2^*(y_2, z_1, z_2) = f_2(\eta_2(y_2, z_1, z_2))$ . Differentiating the first equation we have

$$\begin{aligned} \frac{\partial \eta_1^*(y_1, y_2, z_1)}{\partial y_1} &= \frac{\partial f_1(\eta_1)}{\partial \eta_1} \frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1} \\ D_{y_2} \eta_1^*(y_1, y_2, z_1) &= \frac{\partial f_1(\eta_1)}{\partial \eta_1} D_{y_2} \eta_1(y_1, y_2, z_1) \\ D_{z_1} \eta_1^*(y_1, y_2, z_1) &= \frac{\partial f_1(\eta_1)}{\partial \eta_1} D_{z_1} \eta_1(y_1, y_2, z_1). \end{aligned}$$

Assuming that  $\frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1} \neq 0$ , it follows that

$$\begin{aligned} D_{y_2} \eta_1^*(y_1, y_2, z_1) &= \frac{\frac{\partial \eta_1^*(y_1, y_2, z_1)}{\partial y_1}}{\frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1}} D_{y_2} \eta_1(y_1, y_2, z_1) \\ D_{z_1} \eta_1^*(y_1, y_2, z_1) &= \frac{\frac{\partial \eta_1^*(y_1, y_2, z_1)}{\partial y_1}}{\frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1}} D_{z_1} \eta_1(y_1, y_2, z_1). \end{aligned}$$

If we use the implicit function theorem for fixed  $z$  to determine  $y_1$  as a function of  $y_2$  in

$\eta_1(y_1, y_2, z_1) = u_1$ , we obtain  $D_{y_2} y_1(y_2) = -\left(\frac{\partial \eta_1(y_1, y_2, z_1)}{\partial y_1}\right)^{-1} D_{y_2} \eta_1(y_1, y_2, z_1)$  so that  $D_{y_2} y_1(y_2)$  is

identified. This reduces to the identification condition for the ratio

$\frac{\partial y_1}{\partial y_2} = -\left(\frac{\partial \eta_1(y_1, y_2)}{\partial y_2}\right) / \left(\frac{\partial \eta_1(y_1, y_2)}{\partial y_1}\right)$  obtained by Matzkin (2008) if  $n=2$ ,  $k_1=0$  and  $k_2=1$ .

Differentiating  $\eta_2^*(y_2, z_1, z_2) = f_2(\eta_2(y_2, z_1, z_2))$  yields

$$\begin{aligned} D_{z_1} \eta_2^*(y_2, z_1, z_2) &= D_{\eta_2} f_2(\eta_2) D_{z_1} \eta_2(y_2, z_1, z_2) \\ D_{z_2} \eta_2^*(y_2, z_1, z_2) &= D_{\eta_2} f_2(\eta_2) D_{z_2} \eta_2(y_2, z_1, z_2) \\ D_{y_2} \eta_2^*(y_2, z_1, z_2) &= D_{\eta_2} f_2(\eta_2) D_{y_2} \eta_2(y_2, z_1, z_2). \end{aligned}$$

If  $\text{rank}(D_{y_2} \eta_2(y_2, z_1, z_2)) = \text{rank}(D_{y_2} \eta_2^*(y_2, z_1, z_2)) = n-1$ , we obtain

$$\begin{aligned} (D_{y_2} \eta_2^*(y_2, \bar{z}_1, \bar{z}_2))^{-1} D_{\bar{z}_1} \eta_2^*(y_2, \bar{z}_1, \bar{z}_2) &= (D_{y_2} \eta_2(y_2, \bar{z}_1, \bar{z}_2))^{-1} D_{\bar{z}_1} \eta_2(y_2, \bar{z}_1, \bar{z}_2) \\ (D_{y_2} \eta_2^*(y_2, \bar{z}_1, \bar{z}_2))^{-1} D_{\bar{z}_2} \eta_2^*(y_2, \bar{z}_1, \bar{z}_2) &= (D_{y_2} \eta_2(y_2, \bar{z}_1, \bar{z}_2))^{-1} D_{\bar{z}_2} \eta_2(y_2, \bar{z}_1, \bar{z}_2)' \end{aligned}$$

Therefore, using the implicit function theorem for  $\eta_2(y_2, \bar{z}_1, \bar{z}_2) = u_2$ , we observe that  $D_{(\bar{z}_1, \bar{z}_2)} y_2(\bar{z}_1, \bar{z}_2)$  is identified. In particular, if  $n=2$ ,  $k_1=0$  and  $k_2=1$ , the ratio  $\frac{\partial y_2}{\partial \bar{z}} = -\left(\frac{\partial \eta_2(y_2, \bar{z})}{\partial \bar{z}}\right) / \left(\frac{\partial \eta_2(y_2, \bar{z})}{\partial y_2}\right)$  is identified (c.f. Matzkin (2008)).

### 3.4. A simultaneous equations model of Berry and Haile (2013)

As a special case of the model considered in Section 3.1, suppose that  $\eta(y, z) = r(y) - g(z)$ ,  $y \in \mathbb{R}^n$  and  $z \in \mathbb{R}^k$  (e.g. Berry and Haile (2013) and Section 4.2 of Matzkin (2008)). We consider an alternative structure of the form  $\eta^*(y, z) = r^*(y) - g^*(z)$ . All observationally equivalent structures must be of the form  $\eta^*(y, z) = f(r(y) - g(z))$ . We assume differentiability of all functions, so that

$$\begin{aligned} D_y \eta^*(y, z) &= D_y r^*(y) = D_\eta f(\eta) \Big|_{\eta=r(y)-g(z)} D_y r(y) \\ D_z \eta^*(y, z) &= -D_z g^*(z) = -D_\eta f(\eta) \Big|_{\eta=r(y)-g(z)} D_z g(z). \end{aligned}$$

Notice that the right hand side of the first equation depends on  $z$  but the left hand side does not. Similarly, the right hand side of the second equation depends on  $y$  but the left hand side does not. This means that  $D_\eta f(\eta) \Big|_{\eta=r(y)-g(z)}$  must equal a constant matrix,  $F$  say. So  $\eta^*(y, z) = F(r(y) - g(z))$  and  $r^*(y) = F(r(y) + c)$  and  $g^*(z) = F(g(z) + c)$  where  $c$  is a vector of constants. Notice that  $F$  must be invertible to make the transformation  $f$  bijective. Assuming that  $D_y r(y)$  and  $D_y r^*(y)$  are non-singular (e.g. Assumption 1 of Berry and Haile (2013)) we can write

$$(D_y r^*(y))^{-1} D_z g^*(z) = (D_y r(y))^{-1} D_z g(z),$$

so that  $D_z y(z)$  is identified.

If  $n=k$  and  $g(z) = (g_1(z_1), g_2(z_2), \dots, g_n(z_n))$  as in equation (9) of Berry and Haile (2013), then  $F$  must be diagonal (or  $g^*(z)$  could not have the same form as  $g(z)$ ). By introducing the normalizations in equations (10), (11) and (12) of Berry and Haile (2013), one imposes  $c=0$  and  $F=I_n$  so that  $r^*(y) = r(y)$  and  $g^*(z) = g(z)$ . Notice once again that no restrictions on the errors have been imposed.

### 3.5. Control function separability

Blundel and Matzkin (2013) investigate under what conditions the structure  $(\eta, P_u)$  with  $\eta_1(y_2, y_2) = u_1$  and  $\eta_2(y_2, z) = u_2$  (their Model T) is observationally equivalent to a structure  $(\eta^*, P_u^*)$  with  $\eta_1^*(y_1, y_2) = \eta_1(y_1, y_2)$  and  $\eta_2^*(y_1, y_2, z) = u_2^*$  (their Model I). It follows immediately from Theorem 1 that  $\eta_2^*(y_1, y_2, z) = f(\eta_1(y_1, y_2), \eta_2(y_2, z))$ . Up to this point we have no assumption on the functions involved apart from measurability. Now we introduce some further assumptions as done by Blundel and Matzkin (2013). We assume that all functions are differentiable, and

$$\frac{\partial \eta_1(y_1, y_2)}{\partial y_1} > 0, \quad \frac{\partial \eta_2^*(y_1, y_2, z)}{\partial y_2} > 0, \quad \frac{\partial \eta_1(y_1, y_2)}{\partial y_2} \frac{\partial \eta_2^*(y_1, y_2, z)}{\partial y_1} < 0,$$

(see the proof of Lemma 1 in Blundel and Matzkin (2013)). Then,

$$\frac{\partial \eta_2^*(y_1, y_2, z)}{\partial y_1} = \frac{\partial f(\eta_1, \eta_2(y_2, z))}{\partial \eta_1} \Big|_{\eta_1 = \eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2)}{\partial y_1}.$$

Multiplying both sides by  $\frac{\partial \eta_1(y_1, y_2)}{\partial y_2}$  we obtain

$$\frac{\partial \eta_1(y_1, y_2)}{\partial y_2} \frac{\partial \eta_2^*(y_1, y_2, z)}{\partial y_1} = \frac{\partial f(\eta_1, \eta_2(y_2, z))}{\partial \eta_1} \Big|_{\eta_1 = \eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2)}{\partial y_1} \frac{\partial \eta_1(y_1, y_2)}{\partial y_2} < 0.$$

Since  $\frac{\partial \eta_1(y_1, y_2)}{\partial y_1} > 0$ , it must be that

$$(10) \quad \frac{\partial f(\eta_1, \eta_2(y_2, z))}{\partial \eta_1} \Big|_{\eta_1 = \eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2)}{\partial y_2} < 0.$$

Now we consider

$$\frac{\partial \eta_2^*(y_1, y_2, z)}{\partial y_2} = \frac{\partial f(\eta_1, \eta_2(y_2, z))}{\partial \eta_1} \Big|_{\eta_1 = \eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2)}{\partial y_2} + \frac{\partial f(\eta_1(y_1, y_2), \eta_2)}{\partial \eta_2} \Big|_{\eta_2 = \eta_2(y_2, z)} \frac{\partial \eta_2(y_2, z)}{\partial y_2} > 0,$$

so that

$$\frac{\partial f(\eta_1(y_1, y_2), \eta_2)}{\partial \eta_2} \Big|_{\eta_2 = \eta_2(y_2, z)} \frac{\partial \eta_2(y_2, z)}{\partial y_2} > - \frac{\partial f(\eta_1, \eta_2(y_2, z))}{\partial \eta_1} \Big|_{\eta_1 = \eta_1(y_1, y_2)} \frac{\partial \eta_1(y_1, y_2)}{\partial y_2}.$$

Notice that the right hand side is positive because of (10), so that  $\frac{\partial f(\eta_1, \eta_2)}{\partial \eta_2}$  and  $\frac{\partial \eta_2(y_2, z)}{\partial y_2}$  must

have the same sign and Theorem 1 of Blundel and Matzkin (2013) is satisfied without imposing conditions on the error terms.

## 4. Conclusions

This paper has studied necessary and sufficient conditions for two structures to be observationally equivalent. It has shown that observational equivalence can be established for very general structures defined by systems of equations that may not even be continuous. The main result of the paper and the applications tend to stress the importance of the system of structural equations over the distribution of the error terms. Some existing results have been shown to hold under much weaker conditions than previously thought. This suggests that others identification conditions available in the literature may be weakened.

## References

- Benkard, C. L. and S. 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. and P. A. Haile (2013). "Identification in a Class of Nonparametric Simultaneous Equations Models." *Cowles Foundation Discussion Paper No. 1787RR*.
- Blundel, R. and R. L. Matzkin (2013). "Control Functions in Nonseparable Simultaneous Equations Models." *cemmap working paper CWP28/10*.
- Brown, B. W. (1983). "The Identification Problem Systems of Nonlinear in the Variables." *Econometrica* **51**(1): 175-196.
- Chernozhukov, V. and C. Hansen (2005). "An IV Model of Quantile Treatment Effects." *Econometrica* **73**: 245-261.
- Chernozhukov, V., G. W. Imbens and W. K. Newey (2007). "Instrumental Variable Estimation of Nonseparable Models." *Journal of Econometrics* **139**: 4-14.
- Chesher, A. (2003). "Identification in Nonseparable Models." *Econometrica* **71**: 1405-1441.
- Hoderlein, S. and E. Mammen (2007). "Identification of Marginal Effects in Nonseparable Models without Monotonicity." *Econometrica* **75**(5): 1513-1518.
- Hsiao, C. (1983). Identification. *Handbook of Econometrics Volume 1*. Z. Griliches and M. D. Intriligator, North-Holland: 223-283.
- Imbens, G. W. and W. K. Newey (2009). "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." *Econometrica* **77**(5): 1481-1512.
- Lehmann, E. L. (1997). *Testing Statistical Hypotheses*. New York, Springer-Verlag.
- Matzkin, R. L. (2003). "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica* **71**: 1339-1376.
- Matzkin, R. L. (2007). Nonparametric Identification. *The Handbook of Econometrics*. J. J. Heckman and E. E. Leamer. Amsterdam, North-Holland.
- Matzkin, R. L. (2008). "Identification in Nonparametric Simultaneous Equations Models." *Econometrica* **76**(5): 945-978.
- Matzkin, R. L. (2010). "Estimation of Nonparametric Models with Simultaneity." *Discussion Papers UCLA*.
- Newey, W. K. and J. L. Powell (2003). "Instrumental Variable Estimation of Nonparametric Models." *Econometrica* **71**(5): 1565-1578.
- Newey, W. K., J. L. Powell and F. Vella (1999). "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica* **67**: 565-603.
- Roehrig, C. S. (1988). "Conditions for Identification in Nonparametric and Parametric Models." *Econometrica* **56**: 433-447.
- Schilling, R. L. (2005). *Measures, integrals and martingales*. Cambridge, Cambridge University Press.