CORRECT SPECIFICATION AND IDENTIFICATION OF NONPARAMETRIC TRANSFORMATION MODELS

PIERRE-ANDRE CHIAPPORI AND IVANA KOMUNJER

Abstract. This paper derives necessary and sufficient conditions for nonparametric transformation models to be (i) correctly specified, and (ii) identified. Our correct specification conditions come in a form of partial differential equations; when satisfied by the true distribution, they ensure that the observables are indeed generated from a nonparametric transformation model. Our nonparametric identification result is global; we derive it under conditions that are substantially weaker than full independence.

1. Introduction

A variety of structural econometric models comes in a form of transformation models containing unknown functions. One important class are models of binary choice in which the underlying random utilities are additively separable in the stochastic term as well as the unobserved attributes of the alternatives; a simple two-good version of a demand model à la Berry, Levinsohn, and Pakes (1995) is one specific example. Another class are hedonic models studied by Ekeland, Heckman, and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2005). Further examples of nonseparable econometric models that fall in the transformation model framework can be found in a recent survey by Matzkin (2007).

Date: November 2008

Affiliations and Contact Information: Chiappori: Department of Economics, Columbia University. Email: pc2167@columbia.edu. Komunjer: Department of Economics, University of California San Diego. Email: komunjer@ucsd.edu.

Paper presented at seminars at MIT and the University of Chicago. We thank participants and Victor Chernozhukov, Jim Heckman, Jerry Hausman and Whitney Newey for useful comments. Errors are ours.

The present paper focuses on the following two questions: first, is it possible to test whether a nonparametric transformation model is correctly specified? and, second, under what conditions is the correctly specified model also identified? Regarding the first question, our main contribution is twofold. We derive testable implications of nonparametric transformation models that come in a form of partial differential equations; in addition, we show that these equations are also sufficient for the models to be correctly specified. This means that any observed distribution satisfying these equations can indeed be derived from a nonparametric transformation model, the components of which can moreover be explicitly constructed. Regarding the second question, our main result is to show that transformation models are nonparametrically globally identified under conditions that are significantly weaker than full independence.

Extant literature offers few discussions on the subject of correct specification of nonparametric transformation models. The question that we ask is whether such models put restrictions on the distribution of the observables, restrictions which when violated would invalidate the assumption of the transformation model being correct. Conversely, we seek conditions on the observables which when satisfied guarantee that the transformation models are correctly specified. Among the few papers addressing this issue, one can mention Buera (2006) and Chiappori and Ekeland (2008) both of which deal explicitly with restrictions that take the form of partial differential equations.

We now discuss how the identification result of our paper relates to the literature. It is well-known that in nonparametric linear models $Y = g(X) + \epsilon$, the unknown function g can be identified from $E(\epsilon|Z) = 0$ w.p.1 if the conditional distribution of the endogenous regressor X given the instrument Z is *complete* (see, e.g., Darolles, Florens, and Renault, 2002; Blundell and Powell, 2003; Newey and Powell, 2003; Hall and Horowitz, 2005; Severini and Tripathi, 2006; d'Haultfoeuille, 2006). Given that the model is linear in g, this identification result is global in nature. Nothing is said, however, about the identification of the conditional distribution $F_{\epsilon|X}$ of the disturbance.

In this paper, we show that a similar completeness condition—when combined with independence—is sufficient for identification of T, g and $F_{\epsilon|X}$ in a nonparametric transformation model $Y = T\big(g(X) + \epsilon\big)$, where T is strictly monotonic. Specifically, we work in a framework in which X can be decomposed into an exogenous subvector X^I such that $\epsilon \perp X^I$, and an endogenous subvector X^{-I} whose conditional distribution given Z is complete. Our main assumption is that $E(\epsilon|Z) = 0$ w.p.1.

Even though the nonparametric transformation model is nonlinear in g and $F_{\epsilon|X}$, we obtain identification results that are global. We note that by letting $\theta \equiv (T, g)$ we can write the model as a special case of a nonlinear nonparametric instrumental variable model $E\big[\rho(Y, X, \theta)\big|Z\big] = 0$ w.p.1 where $\rho(Y, X, \theta) \equiv T^{-1}(Y) - g(X)$. For such models, Chernozhukov, Imbens, and Newey (2007) propose an extension of the completeness condition that guarantees θ to be locally nonparametrically identified. It is worth pointing out that their results are local in nature, and that nothing is being said about the identifiability of $F_{\epsilon|X}$.

Our results are close in spirit to those obtained by Ekeland, Heckman, and Nesheim (2004) who show that assuming $\epsilon \perp X$ is sufficient to establish nonparametric identifiability of T, g and the distribution F_ϵ of ϵ (up to unknown constants). In the same paper, the authors derive an additional result that relaxes the independence assumption and replaces it with $E(\epsilon|X) = 0$ w.p.1. They show that the latter is sufficient to identify general parametric specifications for $T(y, \phi)$ and $g(x, \theta)$ where ϕ and θ are finite dimensional parameters. Once $T(y, \phi)$ and $g(x, \theta)$ are specified, the results derived by Komunjer (2008) can be used to further check whether global GMM identification of ϕ and θ holds.

We extend Ekeland, Heckman, and Nesheim (2004) in two important directions: first, we prove nonparametric identification of the function T even when the regressor X contains an endogenous component; and second, we show that if there exists nonparametric instrumental variables Z such that the conditional distribution of

X^{-I} given Z is complete, then the conditional moment conditions $E(\epsilon|Z) = 0$ w.p.1 are sufficient to identify g nonparametrically.

The results of this paper are also related to the literature on nonparametric identification under monotonicity assumptions (see Matzkin, 2007, for a recent survey). For example, Matzkin (2003) provides conditions under which in models of the form $Y = m(X, \epsilon)$ with m strictly monotone, the independence assumption $\epsilon \perp X$ is sufficient to globally identify m and F_ϵ (see also Chesher, 2003, for additional local results). In a sense, our result shows that the independence condition can be substantially relaxed, if a certain form of separability between Y, X and ϵ holds, namely, if we have $T^{-1}(Y) = g(X) + \epsilon$.[1]

Even though we do not discuss the issue of nonparametric estimation of the transformation model, we point the reader to several related results. In a special case where $g(X) = \beta' X$, Horowitz (1996) develops $n^{1/2}$-consistent, asymptotically normal, nonparametric estimators of T and F_ϵ. Estimators of β are available since Han (1987). In a special case where the transformation T is finitely parameterized by a parameter ϕ, Linton, Sperlich, and Van Keilegom (2008) construct a mean square distance from independence estimator for the transformation parameter ϕ. Finally, it is worth pointing out that even though they do not provide primitive conditions for global nonparametric identification of θ in the model $E\big[\rho(Y, X, \theta)\big|Z\big] = 0$ w.p.1, the estimation methods developed in Ai and Chen (2003) and Chernozhukov, Imbens, and Newey (2007) yield consistent estimators for θ, and are readily applicable in our setup.

The remainder of the paper is organized as follows. Section 2 introduces the transformation model and recalls basic definitions. In Section 3, we derive necessary and sufficient conditions for the model to be correctly specified under two sets of assumptions: first, under a single independence restriction, and second, when several independence conditions are known to hold. In Section 4 we examine the conditions

[1]See also the discussion on page 24 in Blundell and Powell (2003).

under which the correctly specified model is also identified. All of our proofs are relegated to an Appendix.

2. Model

We consider a nonparametric transformation model of the form

$$Y = T\big(g(X) + \epsilon\big) \tag{1}$$

where Y belongs to $\mathcal{Y} \subseteq \mathbb{R}$, $X = (X_1, \ldots, X_{d_x})$ belongs to $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, ϵ is in $\mathbb{R}$, and $T : \mathbb{R} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathbb{R}$ are unknown functions. The variables Y and X are observed, while ϵ remains latent. We denote by $F_{\epsilon|X}$ the conditional distribution of ϵ given X.

Following the related literature (e.g., Koopmans and Reiersøl, 1950; Brown, 1983; Roehrig, 1988; Matzkin, 2003) we call *structure* a particular value of the triplet $(T, g, F_{\epsilon|X})$ where $T : \mathbb{R} \to \mathcal{Y}$, $g : \mathcal{X} \to \mathbb{R}$, and $F_{\epsilon|X} : \mathbb{R} \times \mathcal{X} \to \mathbb{R}$. Note that the model (1) simply corresponds to the set of all structures $(T, g, F_{\epsilon|X})$ that satisfy certain a priori restrictions. Each structure in the model induces a conditional distribution $F_{Y|X}$ of the observables, and two structures $(\tilde{T}, \tilde{g}, \tilde{F}_{\tilde{\epsilon}|X})$ and $(T, g, F_{\epsilon|X})$ are observationally equivalent if they generate the same $F_{Y|X}$.

When the set of all conditional distributions $F_{Y|X}$ generated by the model contains the true conditional distribution $F^0_{Y|X}$ we say that the model is correctly specified. In that case, the model contains at least one true structure $(T^0, g^0, F^0_{\epsilon|X})$ that induces $F^0_{Y|X}$. The model is then said to be identified, if the set of structures that are observationally equivalent to $(T^0, g^0, F^0_{\epsilon|X})$ reduces to a singleton. In what follows, we derive two sets of results. First, we provide necessary and sufficient conditions under which the model (1) is correctly specified. Second, we give sufficient conditions under which the correctly specified model is also identified.

3. Correct Specification Conditions

Hereafter, we restrict our attention to the transformations T in (1) that are smooth and strictly increasing from $\mathbb{R}$ onto $\mathcal{Y}$.

Assumption A1. *T is twice continuously differentiable on $\mathbb{R}$, $T'(t) > 0$ for every $t \in \mathbb{R}$, and $T(\mathbb{R}) = \mathcal{Y}$.*

In particular, the limit conditions $\lim_{t\to\{-\infty,+\infty\}} T(t) = \{\inf \mathcal{Y}, \sup \mathcal{Y}\}$ hold true under assumption A1. To simplify our analysis, we focus on the case in which the distributions $F_{\epsilon|X}$ in the model (1) are absolutely continuous.

Assumption A2. *For a.e. $x \in \mathcal{X}$, the conditional distribution $F_{\epsilon|X}$ of ϵ given $X = x$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, and has density $f_{\epsilon|X}$ that is continuously differentiable on $\mathbb{R}$ and satisfies:*

$$\int_{\mathbb{R}} f_{\epsilon|X}(t,x)dt = 1 \textit{ and } f_{\epsilon|X}(\cdot,x) > 0 \textit{ on } \mathbb{R}$$

Assumption A2 states that for almost every realization $x \in \mathcal{X}$ of X, the conditional density of ϵ given $X = x$ exists, is positive and continuously differentiable over its entire support $\mathbb{R}$.[2] This assumption, combined with the fact that T is a twice differentiable homeomorphism from $\mathbb{R}$ onto $\mathcal{Y}$, guarantees that the conditional distribution $F_{Y|X}$ of Y given $X = x$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, and has density $f_{Y|X}(\cdot, x)$ with support $\mathcal{Y}$ that is positive and continuously differentiable everywhere on $\mathcal{Y}$. Without loss of generality, we may assume that $\mathcal{Y}$ contains zero.

3.1. **Single Independence Restriction.** We now further restrict the dependence between ϵ and X by making the following assumption:

Assumption A3. *ϵ is independent of X_1.*

Assumption A3 states that at least one component of X is strongly exogenous; we may, with no loss of generality, assume it is X_1. In what follows, whenever $d_x > 1$, we denote by X^{-1} the remaining subvector of X, i.e. $X^{-1} \equiv (X_2, \dots, X_{d_x})$. The supports of X^1 and X^{-1} are denoted $\mathcal{X}^1$ and $\mathcal{X}^{-1}$, respectively.

[2]Each almost everywhere statement is to be understood with respect to the marginal distribution of the random variable in question.

The independence property in A3 has strong testable implications which we now derive. In what follows, $\Theta : \mathcal{Y} \to \mathbb{R}$ denotes the inverse mapping T^{-1}. Under A1, Θ is twice continuously differentiable and strictly increasing on $\mathcal{Y}$. Note that in addition $\Theta(\mathcal{Y}) = \mathbb{R}$. Equation (1) is equivalent to:

$$\epsilon = \Theta\left(Y\right) - g(X) \tag{2}$$

so by $\Theta' > 0$ and the independence of ϵ and X_1,

$$\begin{aligned} \Phi(y,x) &\equiv F_{Y|X}(y,x) \\ &= \Pr\left(Y \leqslant y \mid X = x\right) \\ &= \Pr\left(\epsilon \leqslant \Theta\left(y\right) - g(x) \mid X = x\right) \\ &= F_{\epsilon|X}\left(\Theta\left(y\right) - g(x), x^{-1}\right) \qquad (3) \end{aligned}$$

where $(y,x) \in \mathcal{Y} \times \mathcal{X}$, $\Phi : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$, and $x^{-1} \equiv (x_2, \ldots, x_{d_x})$. By assumption A2, $\Phi(\cdot, x)$ is twice continuously differentiable on $\mathcal{Y}$ for a.e. $x \in \mathcal{X}$. In order to ensure the existence of the other partial derivatives of Φ, we consider the following restrictions on g.

Assumption A4. *For a.e. $x \in \mathcal{X}$, the second-order partial derivative $\partial^2 g(x)/(\partial x_1)^2$ exists and is continuous; moreover, $\partial g(x)/\partial x_1 \neq 0$.*

Note that assumption A4 only restricts the behavior of the partial derivatives of g with respect to x_1. Nothing is being said about the behavior of g with respect to the remaining components x^{-1}. Under assumptions A2 and A4, the second-order partial derivatives $\partial^2\Phi(y,x)/\left(\partial y\right)^2$, $\partial^2\Phi(y,x)/\left(\partial y \partial x_1\right)$, and $\partial^2\Phi(y,x)/\left(\partial x_1\right)^2$ exist, are continuous and such that $\partial\Phi(y,x)/\partial y > 0$ and $\partial\Phi(y,x)/\partial x_1 \neq 0$ for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$.

We are now ready to state our first result.

Proposition 1. *If Φ is generated by a structure $(T, g, F_{\epsilon|X})$ that satisfies assumptions A1-A4, then for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$ we have:*

$$\textbf{Condition C}: \quad \frac{\partial^2}{\partial y \partial x_k}\left(\log\left|\frac{\partial\Phi(y,x)/\partial y}{\partial\Phi(y,x)/\partial x_1}\right|\right) = 0, \quad \textit{for every } 1 \leqslant k \leqslant d_x.$$

Proposition 1 shows that if the model (1) is correctly specified and such that any true structure $(T^0, g^0, F^0_{\epsilon|X})$ satisfies assumptions A1-A4, then $\Phi^0 \equiv F^0_{Y|X}$ necessarily satisfies condition C. We now examine whether condition C is also sufficient for the correct specification of the transformation model.

For this, we first note that under the assumptions of Proposition 1, any function Φ in (3) satisfies:

$$\lim_{y \to \inf \mathcal{Y}} \Phi(y,x) = 0 \quad \text{and} \quad \lim_{y \to \sup \mathcal{Y}} \Phi(y,x) = 1 \tag{4}$$

In addition, it holds that for a.e. $x \in \mathcal{X}$,

$$\lim_{y \to \{\inf \mathcal{Y}, \sup \mathcal{Y}\}} \int_0^y \left|\frac{\partial\Phi(u,x)/\partial y}{\partial\Phi(u,x)/\partial x_1}\frac{\partial\Phi(0,x)/\partial x_1}{\partial\Phi(0,x)/\partial y}\right| du = \{-\infty, +\infty\} \tag{5}$$

The converse to the implication in Proposition 1 is then as follows.

Proposition 2. *If $\Phi : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ is a mapping such that for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$, Φ has continuous third-order partial derivatives $\partial^3\Phi(y,x)/(\partial y \partial x_1 \partial x_k)$ and $\partial^3\Phi(y,x)/(\partial y^2 \partial x_k)$ ($1 \leqslant k \leqslant d_x$), satisfies the limit conditions* (4) *and* (5)*, and is such that $\partial\Phi(y,x)/\partial y > 0$, $\partial\Phi(y,x)/\partial x_1 \neq 0$, then condition C implies that there exists a structure $(\bar{T}, \bar{g}, \bar{F}_{\bar{\epsilon}|X})$ that satisfies assumptions A1-A4, and generates Φ.*

According to Propositions 1 and 2, if (and only if) the true distribution $F^0_{Y|X}$ satisfies condition C, the transformation model (1) is correctly specified. In other words, if $\Phi^0(y,x) = F^0_{Y|X}(y,x)$ satisfied condition C, then there exists a true structure $(T^0, g^0, F^0_{\epsilon|X})$ such that $\Phi^0(y,x) = F^0_{\epsilon|X}(\Theta^0(y) - g^0(x), x^{-1})$ for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$. Note, however, that condition C is by itself not sufficient to guarantee that this true structure is unique, i.e. that the model (1) is identified; this issue will be addressed in the next section. Before that, we examine the issue of correct specification in the case in which at least two components of X are strongly exogenous.

3.2. **Several Independence Restrictions.** If several of the X variables are independent of ϵ, additional restrictions on Φ are generated. Formally, let $X^I \equiv (X_1, \ldots, X_I)$ be a subvector of X containing the first I components of X where now $2 \leqslant I \leqslant d_x$.[3] We denote by $\mathcal{X}^I$ the support of X^I. Similarly, if $I < d_x$, we let X^{-I} denote the remaining subvector of X, i.e. $X^{-I} \equiv (X_{I+1}, \ldots, X_{d_x})$. The support of X^{-I} is denoted $\mathcal{X}^{-I}$.

We now assume the following:

Assumption A5. *ϵ is independent of X^I.*

Assumption A6. *For a.e. $x \in \mathcal{X}$ and every $1 \leqslant i, j \leqslant I$, the second-order partial derivatives $\partial^2 g(x)/\left(\partial x_i \partial x_j\right)$ exist and are continuous; moreover, $\partial g(x)/\partial x_i \neq 0$.*

Assumptions A5 and A6 strengthen our earlier assumptions A3 and A4, respectively. By the independence of ϵ and X^I, under A1 and A2 we now have:

$$\Phi\left(y, x\right) \equiv F_{\epsilon|X}\left(\Theta\left(y\right) - g(x), x^{-I}\right) \tag{6}$$

for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$. Similar to previously, A6 combined with A2 guarantees that for every $1 \leqslant i, j \leqslant I$ the second-order partial derivatives $\partial^2\Phi(y,x)/\left(\partial y\right)^2$, $\partial^2\Phi(y,x)/\left(\partial y \partial x_i\right)$, and $\partial^2\Phi(y,x)/\left(\partial x_i \partial x_j\right)$ exist, are continuous and such that $\partial\Phi(y,x)/\partial y > 0$ and $\partial\Phi(y,x)/\partial x_i \neq 0$ for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$. For any such (y, x), the independence between ϵ and X^I generates additional restrictions on Φ.

Proposition 3. *If Φ is generated by a structure $(T, g, F_{\epsilon|X})$ that satisfies assumptions A1-A2 and A5-A6, then for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$, Φ satisfies condition C, and in addition we have:*

$$\textbf{Condition S}: \quad \frac{\partial^2\Phi\left(y,x\right)}{\partial y \partial x_j}\frac{\partial\Phi\left(y,x\right)}{\partial x_i} = \frac{\partial^2\Phi\left(y,x\right)}{\partial y \partial x_i}\frac{\partial\Phi\left(y,x\right)}{\partial x_j}, \text{ for every } 1 \leqslant i, j \leqslant I.$$

[3] Of course, the ordering of the components of X is irrelevant.

It can be noted that at any point where $\partial\Phi(y,x)/\partial x_1$ does not vanish, condition S implies that for every $1 \leqslant i \leqslant I$, the ratio $(\partial\Phi(y,x)/\partial x_i)/(\partial\Phi(y,x)/\partial x_1)$ is a function of x only; indeed,

$$\begin{aligned}
&\frac{\partial}{\partial y}\left(\frac{\partial\Phi(y,x)/\partial x_i}{\partial\Phi(y,x)/\partial x_1}\right) \\
&= \left(\frac{1}{\partial\Phi(y,x)/\partial x_1}\right)^2\left(\frac{\partial^2\Phi\left(y,x\right)}{\partial y\partial x_i}\frac{\partial\Phi\left(y,x\right)}{\partial x_1}-\frac{\partial^2\Phi\left(y,x\right)}{\partial y\partial x_1}\frac{\partial\Phi\left(y,x\right)}{\partial x_i}\right) \\
&= 0 \text{ by condition S}
\end{aligned}$$

In other words, condition S is a standard separability property, expressing the fact that the marginal rate of substitution between x_i and x_1 along Φ does not depend on y. In particular, if one variable—here X_1—is known to be strongly exogenous, condition S provides a simple, nonparametric condition that can be used to decide whether any other variable X_i is also strongly exogenous.

Similar to previously, we now derive a converse to the implication in Proposition 3.

Proposition 4. *If $\Phi : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ is a mapping such that for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$, Φ has continuous third-order partial derivatives $\partial^3\Phi(y,x)/\left(\partial y\partial x_1\partial x_k\right)$ and $\partial^3\Phi(y,x)/\left(\partial y^2\partial x_k\right)$ ($1 \leqslant k \leqslant d_x$), satisfies condition C and the limit conditions* (4) *and* (5)*, and is such that $\partial\Phi(y,x)/\partial y > 0$, $\partial\Phi(y,x)/\partial x_i \neq 0$ ($1 \leqslant i \leqslant I$), then condition S implies that there exists a structure $(\tilde{T},\tilde{g},\tilde{F}_{\tilde{\epsilon}|X})$ that satisfies assumptions A1-A2 and A5-A6, and generates Φ.*

4. Identification Condition

We now address the identification problem, namely: If there exists a true structure $(T^0,g^0,F^0_{\epsilon|X})$ that generates $F^0_{Y|X}$, is it possible to find an alternative structure that is different from but observationally equivalent to $(T^0,g^0,F^0_{\epsilon|X})$? More formally, the structure $(T^0,g^0,F^0_{\epsilon|X})$ is globally identified if any observationally equivalent structure $(\tilde{T}^0,\tilde{g}^0,\tilde{F}^0_{\tilde{\epsilon}|X})$ satisfies: for every $t \in \mathbb{R}$, every $y \in \mathcal{Y}$, and a.e. $x \in \mathcal{X}$

$$\tilde{\Theta}^0(y) = \Theta^0(y), \quad \tilde{g}^0(x) = g^0(x), \quad \text{and} \quad \tilde{F}^0_{\tilde{\epsilon}|X}(t,x) = F^0_{\epsilon|X}(t,x).$$

In what follows, we maintain the assumption that X_1 is strongly exogenous, i.e. independent of ε. Regarding the other variables, X^{-1}, we assume the existence of an observable instrument Z that takes values in $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$, and whose relation to X^{-1} is specified below. As already stated, condition C—while sufficient for correct specification of the transformation model—is not sufficient to guarantee its identification. The problem we now examine can be restated as follows: to what extent is it possible to recover the functions $T^0 : \mathbb{R} \to \mathcal{Y}$, $g^0 : \mathcal{X} \to \mathbb{R}$, and $F^0_{\epsilon|X} : \mathbb{R} \times \mathcal{X}^{-1} \to \mathbb{R}$, which for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$ satisfy: $\Phi^0(y,x) = F^0_{\epsilon|X}(\Theta^0(y) - g^0(x), x^{-1})$, where as before $\Phi^0(y,x) = F^0_{Y|X}(y,x)$?

For one thing, it is clear from (1) that some normalization of the model is needed; indeed, for any $(\lambda, \mu) \in \mathbb{R}^2$, the transformation model (1) is equivalent to

$$Y = \tilde{T}\big(\lambda\, g(X) + \mu + \lambda\epsilon\big)$$

where $\tilde{T}$ is defined by $\tilde{T}(t) \equiv T\left((t-\mu)/\lambda\right)$. We therefore impose that any T in (1) satisfies the normalization condition:

$$T\,(0) = 0 \text{ and } T'\,(0) = 1 \tag{7}$$

An interpretation of (7) is discussed below.

In addition to the independence assumption A3, we now restrict the dependence between ϵ and X^{-1}.

Assumption A7. *For a.e. $z \in \mathcal{Z}$, $E(\epsilon|Z=z) = 0$ and the conditional distribution of X^{-1} given $Z = z$ is complete: for every function $h : \mathcal{X}^{-1} \to \mathbb{R}$ such that $E[h(X^{-1})]$ exists and is finite, $E[h(X^{-1}) \mid Z = z] = 0$ implies $h(x^{-1}) = 0$ for a.e. $x^{-1} \in \mathcal{X}^{-1}$.*

Recall from A3 that ϵ was assumed to be independent from X^1. The other components are on the other hand allowed to be endogenous provided there exists a vector of instruments Z with respect to which the distribution of X^{-1} is complete, and such that ϵ is mean independent of Z. Further discussion of the completeness condition can be found in Darolles, Florens, and Renault (2002), Blundell and Powell (2003), Newey and Powell (2003), Hall and Horowitz (2005), Severini and Tripathi (2006), and d'Haultfoeuille (2006), among others. For example, it is equivalent to requiring that for every function $h : \mathcal{X}^{-1} \to \mathbb{R}$ such that $E[h(X^{-1})] = 0$ and $\text{var}[h(X^{-1})] > 0$, there exist a function $g : \mathcal{Z} \to \mathbb{R}$ such that $E[h(X^{-1})g(Z)] \neq 0$ (see Lemma 2.1. in Severini and Tripathi, 2006).

The main identification result is provided by the following statement:

Proposition 5. *Assume that Φ^0 satisfies the conditions of Proposition 2. Let $(T^0, g^0, F^0_{\epsilon|X})$ and $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ be two observationally equivalent structures that generate $F^0_{Y|X}$, and satisfy assumptions A1-A4 and the normalization condition* (7). *Then, assumption A7 is necessary and sufficient to globally identify $(T^0, g^0, F^0_{\epsilon|X})$.*

Proposition 5 shows two results. First, that the completeness condition is sufficient to nonparametrically identify the transformation model (1). This identification result is global even though the model (1) is nonlinear in g and $F_{\epsilon|X}$. The second result of Proposition 5 is that the completeness condition is also necessary, in the following sense: assume that there exists some function $h : \mathcal{X}^{-1} \to \mathbb{R}$ that (i) does not vanish a.e., but (ii) is such that $E[h(X^{-1}) \mid Z = z] = 0$ for a.e. $z \in \mathcal{Z}$. Then there exists two different but observationally equivalent structures that generate $F^0_{Y|X}$ while satisfying assumptions A1-A4 and the normalization condition (7).

It is worth pointing out that the case of several strongly exogenous variables considered in Subsection 3.2 is a particular version of the setting above. Indeed, assume that the disturbance ϵ in the model (1) is known to be independent of X_i $(1 \leqslant i \leqslant I)$. Then, if $E(\epsilon) = 0$, it holds that w.p.1 $E(\epsilon|X_i) = 0$. It then suffices to include X_i in the vector of instruments Z for the conditional distribution of X^I to be complete with respect to Z.

Finally, we may briefly come back to the normalization condition (7). Its key role is to pin down an additive and a multiplicative constants in the identification of Θ^0. The same could be achieved by imposing the following set of restrictions:

$$E(\epsilon) = 0, \; E[g(X)] = 0, \; \text{and } \text{var}(\epsilon) = 1. \tag{8}$$

In other words, instead of normalizing the value of T at a given point (here, zero), we may require that both ϵ and $g(X)$ have mean zero; and instead of normalizing the value of the derivative T' at a given point (here, zero), we may require that ϵ have unit variance. We then have the following corollary to Proposition 5.

Corollary 6. *Proposition 5 remains true if we replace the normalization condition* (7) *with the one in* (8).

We conclude by noting that if the function g in the model (1) is further assumed to be bounded, then the completeness condition in assumption A7 can be replaced by a bounded completeness condition: for every *bounded* function $h : \mathcal{X}^{-1} \to \mathbb{R}$, $E[h(X^{-1}) \mid Z] = 0$ w.p.1 implies $h(x^{-1}) = 0$ for a.e. $x^{-1} \in \mathcal{X}^{-1}$. The bounded completeness condition is weaker then the completeness condition (see, e.g., Blundell, Chen, and Kristensen, 2007, for a discussion).

Appendix A. Proofs

Proof of Proposition 1. Consider a structure $(T, g, F_{\epsilon|X})$ that satisfies assumptions A1-A4, and generates $\Phi(y, x)$ in the sense of equation (3). Differentiating in y and x_1 gives:

$$\frac{\partial \Phi}{\partial y}(y,x) = \Theta'(y)\frac{\partial F_{\epsilon|X}}{\partial t}(\Theta(y) - g(x), x^{-1}) \tag{9}$$

$$\frac{\partial \Phi}{\partial x_1}(y,x) = -\frac{\partial g}{\partial x_1}(x)\frac{\partial F_{\epsilon|X}}{\partial t}(\Theta(y) - g(x), x^{-1}) \tag{10}$$

where Θ' is the derivative of Θ, and $\partial F_{\epsilon|X}/\partial t$ denotes the partial derivative of $F_{\epsilon|X}$ with respect to its first variable. Let $A \equiv \{x \in \mathcal{X} : \partial\Phi(y,x)/\partial y > 0 \text{ and } \partial\Phi(y,x)/\partial x_1 \neq 0 \text{ for every } y \in \mathcal{Y}\}$. From assumptions A2 and A4 the set A has probability one. Take any point $(x, y) \in A \times \mathcal{Y}$. Then $\partial\Phi(y,x)/\partial x_1 \neq 0$ and we have:

$$\frac{\partial\Phi(y,x)/\partial y}{\partial\Phi(y,x)/\partial x_1} = -\frac{\Theta'(y)}{\partial g(x)/\partial x_1} \tag{11}$$

Under assumption A1, Θ is twice continuously differentiable so we can differentiate the above with respect to y, which gives:

$$\frac{\partial}{\partial y}\left(\log\left|\frac{\partial\Phi(y,x)/\partial y}{\partial\Phi(y,x)/\partial x_1}\right|\right)=\frac{\Theta''(y)}{\Theta'(y)}$$

Given that the right hand side is a function of y alone, the above implies condition C. $\square$

Proof of Proposition 2. The proof is in three steps.

Step 1: Let $\Phi:\mathcal{Y}\times\mathcal{X}\to\mathbb{R}$ be a map such that for every $y\in\mathcal{Y}$ and a.e. $x\in\mathcal{X}$, Φ has continuous third-order partial derivatives $\partial^3\Phi(y,x)/\left(\partial y\partial x_1\partial x_k\right)$ and $\partial^3\Phi(y,x)/\left(\partial y^2\partial x_k\right)$ $(1\leqslant k\leqslant d_x)$, and $\partial\Phi(y,x)/\partial y>0$, $\partial\Phi(y,x)/\partial x_1\neq 0$. As before, the set $A=\{x\in\mathcal{X}:\partial\Phi(y,x)/\partial y>0$ and $\partial\Phi(y,x)/\partial x_1\neq 0$ for every $y\in\mathcal{Y}\}$ has probability one. Consider any $(y,x)\in\mathcal{Y}\times A$. If condition C is satisfied, then $\partial\log\left|(\partial\Phi(y,x)/\partial y)/(\partial\Phi(y,x)/\partial x_1)\right|/\partial y$ is a function of y only. Let then:

$$\phi\left(y\right)\equiv\frac{\partial}{\partial y}\left(\log\left|\frac{\partial\Phi(y,x)/\partial y}{\partial\Phi(y,x)/\partial x_1}\right|\right)\tag{12}$$

Note that ϕ is observable. Now, let $\bar{\Theta}:\mathcal{Y}\to\mathbb{R}$ be defined as a solution to the differential equation

$$\frac{\bar{\Theta}''\left(y\right)}{\bar{\Theta}'\left(y\right)}=\phi\left(y\right)$$

Integrating with respect to y on $\mathcal{Y}$, and using the fact that $0\in\mathcal{Y}$, a solution is:

$$\bar{\Theta}\left(y\right)=\int_0^y\exp\left(\int_0^u\phi\left(s\right)ds\right)du\tag{13}$$

Note that the function $\bar{\Theta}$ in (13) is defined over the whole space $\mathcal{Y}$, is twice continuously differentiable on $\mathcal{Y}$, and we have $\Theta'(y)>0$ on $\mathcal{Y}$. In addition, from the limit condition (5) we have $\lim_{y\to\inf\mathcal{Y}}\bar{\Theta}(y)=-\infty$ and $\lim_{y\to\sup\mathcal{Y}}\bar{\Theta}(y)=+\infty$, so $\bar{\Theta}(\mathcal{Y})=\mathbb{R}$. Letting $\bar{T}\equiv\bar{\Theta}^{-1}$, we then have that $\bar{T}$ satisfies assumption A1.

Step 2. Now, for any $(x,y)\in A\times\mathcal{Y}$, consider the partial differential equation:

$$\frac{\partial\bar{g}}{\partial x_1}(x)=-\frac{\partial\Phi(y,x)/\partial x_1}{\partial\Phi(y,x)/\partial y}\bar{\Theta}'(y)\tag{14}$$

Condition C implies that the right hand side is a function of x only. To see this, note that from (12) we have:

$$\log\left|\frac{\partial\Phi(y,x)/\partial x_1}{\partial\Phi(y,x)/\partial y}\right| = -\int_0^y \phi(s)ds + \alpha(x)$$

for some function α of x alone. So using (13),

$$\left|\frac{\partial\Phi(y,x)/\partial x_1}{\partial\Phi(y,x)/\partial y}\right| = \exp\left[-\int_0^y \phi\left(s\right)ds\right]\exp\left[\alpha(x)\right] = \frac{\exp\left[\alpha(x)\right]}{\bar{\Theta}'(y)}$$

which shows that $|(\partial\Phi/\partial x_1)/(\partial\Phi/\partial y)|\,\bar{\Theta}'$ is a function of x alone. Then, the same must hold for $[(\partial\Phi/\partial x_1)/(\partial\Phi/\partial y)]\,\bar{\Theta}'$. This is clearly true if for a.e. $x \in A$, the function $[(\partial\Phi(\cdot,x)/\partial x_1)/(\partial\Phi(\cdot,x)/\partial y)]\,\bar{\Theta}'(\cdot)$ keeps a constant sign on $\mathcal{Y}$; now, assume that for some $x^* \in \mathcal{X}$ there exist $(y_1,y_2) \in \mathcal{Y}^2$ such that we have $[(\partial\Phi(y_1,x^*)/\partial x_1)/(\partial\Phi(y_1,x^*)/\partial y)]\,\bar{\Theta}'(y_1) > 0$ and $[(\partial\Phi(y_2,x^*)/\partial x_1)/(\partial\Phi(y_2,x^*)/\partial y)]\,\bar{\Theta}'(y_2) < 0$; then, by continuity, there exists $y^* \in (\min\{y_1,y_2\},\max\{y_1,y_2\})$ such that $[(\partial\Phi(y^*,x^*)/\partial x_1)/(\partial\Phi(y^*,x^*)/\partial y)]\,\bar{\Theta}'(y^*) = 0$, which is in contradiction with $x^* \in A$; hence, the sign of $[(\partial\Phi/\partial x_1)/(\partial\Phi/\partial y)]\,\bar{\Theta}'$ cannot depend on y and is a function of x alone.

Given that the set $A\cap\mathcal{X}^1$ has probability one, one can integrate (14) with respect to x_1 on $\mathcal{X}^1$ to obtain:

$$\bar{g}\left(x\right) = \int_c^{x_1}\left(-\frac{\partial\Phi\left(y,u,x_2,\ldots,x_{d_x}\right)/\partial x_1}{\partial\Phi\left(y,u,x_2,\ldots,x_{d_x}\right)/\partial y}\bar{\Theta}'(y)\right)du \tag{15}$$

where $c \in \mathcal{X}^1$. Again, note that $\bar{g}$ is defined over the whole set $\mathcal{X}$. Moreover, from (14), we have that $\partial\bar{g}(x)/(\partial x_1)^2$ exists and $\partial\bar{g}(x)/\partial x_1 \neq 0$ on A, so that $\bar{g}$ satisfies assumption A4.

Step 3. Finally, for any $(y,x) \in \mathcal{Y}\times\mathcal{X}$, consider the following change in variables:

$$\Gamma:\ (y,x) \mapsto \left(\bar{\Theta}\left(y\right) - \bar{g}(x), x\right)$$

which maps $\mathcal{Y}\times\mathcal{X}$ onto $\mathbb{R}\times\mathcal{X}$. It is well defined since $\bar{\Theta}'\left(y\right) > 0$ over $\mathcal{Y}$; its inverse $\Gamma^{-1}:\mathbb{R}\times\mathcal{X}\to\mathcal{Y}\times\mathcal{X}$ is precisely:

$$\Gamma^{-1}:\ (t,x) \mapsto \left(\bar{T}\left(t+\bar{g}\left(x\right)\right), x\right)$$

The function Φ can therefore be written as:

$$\Phi\left(y,x\right)=\bar{F}\left(\bar{\Theta}\left(y\right)-\bar{g}(x),x\right) \tag{16}$$

where $\bar{F}\equiv\Phi\circ\Gamma^{-1}$. Given our assumptions on Φ, the mapping $\bar{F}:\mathbb{R}\times\mathcal{X}\to\mathbb{R}$ is such that for a.e. $x\in\mathcal{X}$, $\bar{F}(\cdot,x):\mathbb{R}\to\mathbb{R}$ is twice continuously differentiable on $\mathbb{R}$, and $\partial\bar{F}/\partial x_1$ exist and is continuous. From $\lim_{y\to\inf\mathcal{Y}}\bar{\Theta}(y)=-\infty$, $\lim_{y\to\sup\mathcal{Y}}\bar{\Theta}(y)=+\infty$, and the limit condition (4) we have:

$$\lim_{t\to-\infty}\bar{F}(t,x)=0 \quad\text{and}\quad \lim_{t\to+\infty}\bar{F}(t,x)=1$$

Moreover, differentiating equation (16) with respect to y and x_1, respectively, gives:

$$\frac{\partial\Phi}{\partial y}(y,x) = \bar{\Theta}'(y)\frac{\partial\bar{F}}{\partial t}(\bar{\Theta}(y)-\bar{g}(x),x) \tag{17}$$

$$\frac{\partial\Phi}{\partial x_1}(y,x) = -\frac{\partial\bar{g}}{\partial x_1}(x)\frac{\partial\bar{F}}{\partial t}\left(\bar{\Theta}(y)-\bar{g}(x),x\right)+\left.\frac{\partial\bar{F}}{\partial x_1}(t,x)\right|_{\bar{\Theta}(y)-\bar{g}(x)} \tag{18}$$

where $\partial\bar{F}/\partial t$ denotes the partial derivative of $\bar{F}$ with respect to its first variable. Noting that for every $(y,x)\in\mathcal{Y}\times A$, $\partial\Phi(y,x)/\partial y>0$ and $\bar{\Theta}'>0$, we have that for every $t=\bar{\Theta}\left(y\right)-\bar{g}(x)$:

$$\frac{\partial\bar{F}}{\partial t}(t,x)>0$$

Since $\bar{\Theta}$ is onto $\mathbb{R}$, the above holds for every $(t,x)\in\mathbb{R}\times A$. Hence $\bar{F}$ is a cumulative distribution function that satisfies assumption A2. Let $\bar{\epsilon}$ be a random variable whose conditional distribution given $X=x$ is given by $\bar{F}_{\bar{\epsilon}|X}(\cdot,x)\equiv\bar{F}(\cdot,x)$ for any $x\in\mathcal{X}$. We now show that $\bar{\epsilon}$ satisfies A3. For this, consider again any $(y,x)\in\mathcal{Y}\times A$ and take the ratio of (18) and (17):

$$\frac{\partial\Phi\left(y,x\right)/\partial x_1}{\partial\Phi\left(y,x\right)/\partial y}=-\frac{\partial\bar{g}(x)/\partial x_1}{\bar{\Theta}'(y)}+\frac{\partial\bar{F}\left(t,x\right)/\partial x_1|_{\bar{\Theta}(y)-\bar{g}(x)}}{\partial\Phi\left(y,x\right)/\partial y}$$

Since $\bar{\Theta}$ and $\bar{g}$ have been constructed to satisfy (14), it must be the case that $\partial\bar{F}(t,x)/\partial x_1=0$ whenever $t=\bar{\Theta}\left(y\right)-\bar{g}(x)$ i.e. $t\in\mathbb{R}$. Therefore $\bar{\epsilon}$ is independent of X_1 and we have:

$$\Phi\left(y,x\right)=\bar{F}_{\bar{\epsilon}|X}\left(\bar{\Theta}\left(y\right)-\bar{g}(x),x^{-1}\right)$$

which closes the proof. □

Proof of Proposition 3. Consider a structure $(T, g, F_{\epsilon|X})$ that satisfies assumptions A1-A2 and A5-A6, and generates Φ in the sense of equation (6). Differentiating in y and x_j $(1 \leqslant j \leqslant I)$ gives:

$$\frac{\partial \Phi}{\partial y}(y,x) = \Theta'(y)\frac{\partial F_{\epsilon|X}}{\partial t}(\Theta(y)-g(x),x^{-I}) \tag{19}$$

$$\frac{\partial \Phi}{\partial x_j}(y,x) = -\frac{\partial g}{\partial x_j}(x)\frac{\partial F_{\epsilon|X}}{\partial t}(\Theta(y)-g(x),x^{-I}) \tag{20}$$

where Θ' and $\partial F_{\epsilon|X}/\partial t$ are as in the proof of Proposition 1. Again differentiating equations (19) and (20) with respect to x_i $(1 \leqslant i \leqslant I)$ then gives:

$$\frac{\partial^2 \Phi}{\partial x_i \partial y}(y,x) = -\Theta'(y)\frac{\partial^2 F_{\epsilon|X}}{\partial t^2}(\Theta(y)-g(x),x^{-I})\frac{\partial g}{\partial x_i}(x) \tag{21}$$

$$\begin{aligned}\frac{\partial^2 \Phi}{\partial x_i \partial x_j}(y,x) = &-\frac{\partial^2 g}{\partial x_i \partial x_j}(x)\frac{\partial F_{\epsilon|X}}{\partial t}(\Theta(y)-g(x),x^{-I})\\ &+\frac{\partial g}{\partial x_j}(x)\frac{\partial^2 F_{\epsilon|X}}{\partial t^2}(\Theta(y)-g(x),x^{-I})\frac{\partial g}{\partial x_i}(x)\end{aligned} \tag{22}$$

where $\partial^2 F_{\epsilon|X}/(\partial t)^2$ denotes the second-order partial derivative of $F_{\epsilon|X}$ with respect to its first variable. Let now $A \equiv \{x \in \mathbb{R}^{d_x} : \partial\Phi(y,x)/\partial y > 0$ and $\partial\Phi(y,x)/\partial x_i \neq 0$ for every $1 \leqslant i \leqslant I$ and every $y \in \mathbb{R}\}$. Combining equations (21) and (22), then gives for any $(x,y) \in A \times \mathcal{Y}$:

$$\begin{aligned}\frac{\partial^2 g(x)}{\partial x_i \partial x_j} &= -\frac{1}{(\partial\Phi(y,x)/\partial y)^2}\left(\frac{\partial^2\Phi(y,x)}{\partial x_i\partial x_j}\frac{\partial\Phi(y,x)}{\partial y}-\frac{\partial^2\Phi(y,x)}{\partial x_i\partial y}\frac{\partial\Phi(y,x)}{\partial x_j}\right)\Theta'(y)\\ &= -\frac{1}{(\partial\Phi(y,x)/\partial y)^2}\left(\frac{\partial^2\Phi(y,x)}{\partial x_i\partial x_j}\frac{\partial\Phi(y,x)}{\partial y}-\frac{\partial^2\Phi(y,x)}{\partial x_j\partial y}\frac{\partial\Phi(y,x)}{\partial x_i}\right)\Theta'(y)\end{aligned}$$

where the second equality follows by interchanging the indices i and j in equations (19)-(22). Therefore

$$\frac{\partial^2\Phi(y,x)}{\partial x_i\partial y}\frac{\partial\Phi(y,x)}{\partial x_j} = \frac{\partial^2\Phi(y,x)}{\partial x_j\partial y}\frac{\partial\Phi(y,x)}{\partial x_i} \tag{23}$$

which is condition S. □

Proof of Proposition 4. The proof is in three steps.

Step 1. Let $\Phi : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ be a map such that for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$, Φ has continuous third-order partial derivatives $\partial^3\Phi(y,x)/\left(\partial y \partial x_1 \partial x_k\right)$ and $\partial^3\Phi(y,x)/\left(\partial y^2 \partial x_k\right)$ $(1 \leqslant k \leqslant d_x)$, satisfies condition C and the limit conditions (4) and (5), and is such that $\partial\Phi(y,x)/\partial y > 0$, $\partial\Phi(y,x)/\partial x_i \neq 0$ $(1 \leqslant i \leqslant I)$. Note that any such Φ satisfies the requirements of Proposition 2. Let then $\tilde{T} \equiv \bar{T}$ with $\bar{T}$ as defined in Step 1 of the proof of Proposition 2; this $\tilde{T}$ satisfies assumption A1.

Step 2. We now prove the following Lemma:

Lemma 1. *There exists a function $\tilde{g} : \mathcal{X} \to \mathbb{R}$ such that for every $1 \leqslant i, j \leqslant I$ and a.e. $x \in \mathcal{X}$ the second-order partial derivatives $\partial\tilde{g}/(\partial x_i \partial x_j)$ exist and are continuous; moreover, for every $1 \leqslant i \leqslant I$ and a.e. $x \in \mathcal{X}$:*

$$\frac{\partial \tilde{g}}{\partial x_i}(x) = -\frac{\partial\Phi(y,x)/\partial x_i}{\partial\Phi(y,x)/\partial y}\Theta'(y) \tag{24}$$

Proof of Lemma 1. As before, the set $A = \{x \in \mathcal{X} : \partial\Phi(y,x)/\partial y > 0 \text{ and } \partial\Phi(y,x)/\partial x_1 \neq 0 \text{ for every } y \in \mathcal{Y}\}$ has probability one. Take any $(y,x) \in \mathcal{Y} \times A$; then from condition C, the function $[(\partial\Phi/\partial x_1)/(\partial\Phi/\partial y)]\Theta'$ does not depend on y (c.f. Step 2 in the proof of Proposition 2). For some $c_1 \in \mathcal{X}^1$, define

$$\tilde{g}_1(x) \equiv \int_{c_1}^{x_1} -\frac{\partial\Phi\left(y, u, x_2, \ldots, x_{d_x}\right)/\partial x_1}{\partial\Phi\left(y, u, x_2, \ldots, x_{d_x}\right)/\partial y}\Theta'(y)du \tag{25}$$

Then $\tilde{g}_1 : \mathcal{X} \to \mathbb{R}$, and under the assumptions of Proposition 4, the second-order partial derivatives $\partial\tilde{g}_1/(\partial x_i \partial x_j)$ exist and are continuous $(1 \leqslant i, j \leqslant I)$. Moreover, it is clear that for any $x \in A$

$$\frac{\partial \tilde{g}_1}{\partial x_1}(x) = -\frac{\partial\Phi(y,x)/\partial x_1}{\partial\Phi(y,x)/\partial y}\Theta'(y) \tag{26}$$

We now show that one can find some $\tilde{g}_2 : \mathcal{X}^{-1} \to \mathbb{R}$ whose second-order partial derivatives $\partial\tilde{g}_2/(\partial x_i \partial x_j)$ exist and are continuous $(2 \leqslant i, j \leqslant I)$, and such that the function $(\tilde{g}_1 + \tilde{g}_2) : \mathcal{X} \to \mathbb{R}$ satisfies, for every $x \in A$,

$$\frac{\partial\left(\tilde{g}_1 + \tilde{g}_2\right)}{\partial x_2}(x) = -\frac{\partial\Phi(y,x)/\partial x_2}{\partial\Phi(y,x)/\partial y}\Theta'(y) \tag{27}$$

This requires that

$$\frac{\partial \tilde{g}_2}{\partial x_2}(x^{-1}) = -\frac{\partial \Phi(y,x)/\partial x_2}{\partial \Phi(y,x)/\partial y}\Theta'(y) - \frac{\partial \tilde{g}_1}{\partial x_2}(x) \tag{28}$$

where as before $x^{-1} = (x_2, \ldots, x_{d_x})$. But by conditions C and S, the right hand side of (28) depends neither on x_1 nor on y. Indeed, regarding y, we know that $\tilde{g}_1(x)$ does not depend on y; moreover

$$\frac{\partial \Phi(y,x)/\partial x_2}{\partial \Phi(y,x)/\partial y}\Theta'(y) = \left[\frac{\partial \Phi(y,x)/\partial x_1}{\partial \Phi(y,x)/\partial y}\Theta'(y)\right]\left[\frac{\partial \Phi(y,x)/\partial x_2}{\partial \Phi(y,x)/\partial x_1}\right]$$

and by condition C the first term of the right hand side does not depend on y (c.f. Step 2 in the proof of Proposition 2); similarly, by condition S, the second term of the right hand side does not depend on y. Hence, the right hand side of (28) does not depend on y. Regarding x_1, from equation (26) we have that

$$\frac{\partial^2 \tilde{g}_1}{\partial x_1 \partial x_2}(x) = \frac{\partial}{\partial x_2}\left(-\frac{\partial \Phi(y,x)/\partial x_1}{\partial \Phi(y,x)/\partial y}\Theta'(y)\right)$$

which implies, by condition S, that the partial derivative with respect to x_1 of the right hand side of (28) equals zero. We can therefore write that:

$$-\frac{\partial \Phi(y,x)/\partial x_2}{\partial \Phi(y,x)/\partial y}\Theta'(y) - \frac{\partial \tilde{g}_1}{\partial x_2}(x) = \gamma(x^{-1})$$

and define $\tilde{g}_2 : \mathcal{X}^{-1} \to \mathbb{R}$ as

$$\tilde{g}_2(x^{-1}) \equiv \int_{c_2}^{x_2} \gamma\left(v, x_3, \ldots, x_{d_x}\right) dv$$

where c_2 is a constant that belongs to the support of X_2. Note that as previously, under the assumptions of Proposition 4, the second-order partial derivatives $\partial \tilde{g}_2/(\partial x_i \partial x_j)$ exist and are continuous ($2 \leqslant i, j \leqslant I$); moreover, for every $x \in A$ we have both (27) and

$$\frac{\partial\left(\tilde{g}_1 + \tilde{g}_2\right)}{\partial x_1}(x) = \frac{\partial \tilde{g}_1}{\partial x_1}(x) = -\frac{\partial \Phi(y,x)/\partial x_1}{\partial \Phi(y,x)/\partial y}\Theta'(y)$$

The same method applies, by induction, for all indices $1 \leqslant i \leqslant I$. If for every $1 \leqslant j \leqslant i-1$, we have constructed $\tilde{g}_j\left(x_j, \ldots, x_{d_x}\right)$, such that

$$\frac{\partial\left(\tilde{g}_1 + \ldots + \tilde{g}_{i-1}\right)}{\partial x_j}(x) = -\frac{\partial \Phi(y,x)/\partial x_j}{\partial \Phi(y,x)/\partial y}\Theta'(y) \text{ for all } 1 \leqslant j \leqslant i-1$$

then we can observe that, by conditions C and S, the expression

$$-\frac{\partial\Phi(y,x)/\partial x_i}{\partial\Phi(y,x)/\partial y}\Theta'(y)-\frac{\partial\left(\tilde{g}_1+\ldots+\tilde{g}_{i-1}\right)}{\partial x_i}(x)$$

does not depend on $y, x_1, \ldots, x_{i-1}$. Letting $x^{-(i-1)} \equiv (x_i, \ldots, x_{d_x})$, we can let $\gamma(x^{-(i-1)})$ denote the expression above, and for some constant c_i that belongs to the support of X_i, we can define

$$\tilde{g}_i\left(x_i,\ldots,x_{d_x}\right)\equiv\int_{c_i}^{x_i}\gamma\left(v,x_{i+1},\ldots,x_{d_x}\right)dv$$

Then, we have that for every $1 \leqslant j \leqslant i-1$

$$\frac{\partial\left(\tilde{g}_1+\ldots+\tilde{g}_i\right)}{\partial x_j}(x)=\frac{\partial\left(\tilde{g}_1+\ldots+\tilde{g}_{i-1}\right)}{\partial x_j}(x)$$

and

$$\frac{\partial\left(\tilde{g}_1+\ldots+\tilde{g}_i\right)}{\partial x_i}(x)=-\frac{\partial\Phi(y,x)/\partial x_i}{\partial\Phi(y,x)/\partial y}\Theta'(y)$$

Ultimately, the function $\tilde{g} : \mathcal{X} \to \mathbb{R}$ defined by $\tilde{g} \equiv \sum_{i=1}^{I} \tilde{g}_i$ has continuous second-order partial derivatives $\partial\tilde{g}/(\partial x_i \partial x_j)$ $(1 \leqslant i, j \leqslant I)$; moreover $\tilde{g}$ satisfies Property (24). $\square$

Note that under the assumptions of Proposition 4, we have for a.e. $x \in \mathcal{X}$, $\partial\tilde{g}/\partial x_i \neq 0$ $(1 \leqslant i, j \leqslant I)$, so that $\tilde{g}$ satisfies assumption A6.

Step 3. Now, for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$ consider the change in variables:

$$\Gamma : (y,x) \mapsto (\tilde{\Theta}(y)-\tilde{g}(x), x)$$

which maps $\mathcal{Y} \times \mathcal{X}$ onto $\mathbb{R} \times \mathcal{X}$. Its inverse $\Gamma^{-1} : \mathbb{R} \times \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$, $\Gamma^{-1} : (t, x) \mapsto (\tilde{T}(t + \tilde{g}(x)), x)$, is again well defined since $\tilde{\Theta}' > 0$ on $\mathcal{Y}$. Write then Φ as a function of the new variables:

$$\Phi(y,x)=\tilde{F}\left(\tilde{\Theta}(y)-\tilde{g}(x),x\right)$$

where $\tilde{F} \equiv \Phi \circ \Gamma^{-1}$, so $\tilde{F} : \mathbb{R} \times \mathcal{X} \to \mathbb{R}$. From the limit conditions (4) and $\lim_{y\to\inf\mathcal{Y}} \tilde{\Theta}(y) = -\infty$ and $\lim_{y\to\sup\mathcal{Y}} \tilde{\Theta}(y) = +\infty$, we have that for every $x \in \mathcal{X}$, $\lim_{t\to-\infty} \tilde{F}(t, x) = 0$ and $\lim_{t\to+\infty} \tilde{F}(t, x) = 1$. Under the assumptions of Proposition 4, for a.e. $x \in \mathcal{X}$ the mapping $\tilde{F}(\cdot, x) : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable on

$\mathbb{R}$. Moreover, the partial derivatives $\partial\tilde{F}/\partial x_i$ ($1 \leqslant i \leqslant I$) exist and are continuous. Then, for every $1 \leqslant i \leqslant I$ and every $(y, x) \in \mathcal{Y} \times \mathcal{X}$

$$\frac{\partial\Phi}{\partial y}(y, x) = \tilde{\Theta}'(y)\frac{\partial\tilde{F}}{\partial t}(\tilde{\Theta}(y) - \tilde{g}(x), x) \tag{29}$$

$$\frac{\partial\Phi}{\partial x_i}(y, x) = -\frac{\partial\tilde{g}}{\partial x_i}(x)\frac{\partial\tilde{F}}{\partial t}(\tilde{\Theta}(y) - \tilde{g}(x), x) + \frac{\partial\tilde{F}}{\partial x_i}(t, x)\Big|_{t=\tilde{\Theta}(y)-\tilde{g}(x)} \tag{30}$$

From (29), we conclude that for every $t = \tilde{\Theta}(y) - \tilde{g}(x)$ and a.e. $x \in \mathcal{X}$, $\partial\tilde{F}(t, x)/\partial t > 0$. Since $\tilde{\Theta}(\mathcal{Y}) = \mathbb{R}$, the statement holds for every $t \in \mathbb{R}$, so $\tilde{F}$ satisfies assumption A2. Now, consider any $(y, x) \in \mathcal{Y} \times A$ and take the ratio of (30) and (29). Since $\tilde{g}$ satisfies (24), it must be the case that for every $t = \tilde{\Theta}(y) - \tilde{g}(x)$ and every $i \leqslant i \leqslant I$ we have $\partial\tilde{F}(t, x)/\partial x_i = 0$. Since $\tilde{\Theta}(\mathcal{Y}) = \mathbb{R}$, we have that for any $t \in \mathbb{R}$ and a.e. $x \in \mathcal{X}$, $\tilde{F}(t, x) = \tilde{F}(t, x^{-I})$. The proof then concludes by letting $\tilde{\epsilon}$ be a random variable whose conditional distribution given $X = x$ is given by $\tilde{F}_{\tilde{\epsilon}|X}(\cdot, x) \equiv \tilde{F}(\cdot, x^{-I})$, and noting that $\tilde{\epsilon}$ satisfies the independence condition in A5. $\square$

Proof of Proposition 5. The proof of sufficiency is based on that of Proposition 2. Specifically, it is done in three steps. The fourth and last step shows necessity.

Step 1. We have seen in Step 1 of the proof of Proposition 2 that Θ^0 must satisfy the equation:

$$\frac{\Theta^{0\prime\prime}(y)}{\Theta^{0\prime}(y)} = \phi^0(y)$$

where

$$\phi^0(y) \equiv \frac{\partial}{\partial y}\left(\log\left|\frac{\partial\Phi^0(y, x)/\partial y}{\partial\Phi^0(y, x)/\partial x_1}\right|\right)$$

This determines Θ^0 up to two constants $K_1 \in \mathbb{R}$ and $K_2 > 0$: for any $y \in \mathcal{Y}$

$$\Theta^0(y) = K_1 + K_2\int_0^y \exp\left(\int_0^u \phi^0(s)ds\right)du$$

From the normalization condition (7) we have: $\Theta^0(0) = T^{0^{-1}}(0) = 0$ and $\Theta^{0\prime}(0) = 1/[T^{0\prime}(\Theta^0(0))] = 1$, which pins down the constants K_1 and K_2; finally for any $y \in \mathcal{Y}$,

$$\Theta^0(y) = \int_0^y \exp\left(\int_0^u \phi^0(s)\,ds\right)du \equiv \bar{\Theta}^0(y) \tag{31}$$

(where $\bar{\Theta}^0 : \mathcal{Y} \to \mathbb{R}$ is defined in analogy to $\bar{\Theta}$ in (13) by replacing ϕ with ϕ^0) is the *only* solution that satisfies the normalization. Hence for any $y \in \mathcal{Y}$ we have:

$$\Theta^0(y) = \tilde{\Theta}^0(y).$$

Step 2. Let A^0 be a set of probability one, defined as $A^0 \equiv \{x \in \mathcal{X} : \partial\Phi^0(y,x)/\partial y > 0 \text{ and } \partial\Phi^0(y,x)/\partial x_1 \neq 0 \text{ for every } y \in \mathcal{Y}\}$. Again from Step 2 of the proof of Proposition 2 we know that for every $x \in A^0$, g^0 satisfies:

$$\frac{\partial g^0}{\partial x_1}(x) = -\frac{\partial\Phi^0(y,x)/\partial x_1}{\partial\Phi^0(y,x)/\partial y}\Theta^{0\prime}(y) \tag{32}$$

In analogy with the particular solution $\bar{g}$ defined in (15), a particular solution $\bar{g}^0 : \mathcal{X} \to \mathbb{R}$ to (32) is

$$\bar{g}^0(x) \equiv \int_c^{x_1} \left(-\frac{\partial\Phi^0(y,u,x_2,\ldots,x_{d_x})/\partial x_1}{\partial\Phi^0(y,u,x_2,\ldots,x_{d_x})/\partial y}\bar{\Theta}^{0\prime}(y) \right) du \tag{33}$$

where $c \in \mathcal{X}^1$. Obviously, any solution to (32) must have the same partial in x_1 as $\bar{g}^0$ in (33); it must therefore be of the form:

$$g^0(x) = \bar{g}^0(x) + \beta^0(x^{-1}) \tag{34}$$

for some function $\beta^0 : \mathcal{X}^1 \to \mathbb{R}$.

Step 3. Now let g^0 be an arbitrary solution, and consider $E(\epsilon|Z)$ where $\epsilon = \Theta^0(Y) - g^0(X)$. Letting $F_{Y|Z}$ and $F_{X|Z}$ denote the conditional distributions of Y given Z and of X given Z, respectively, we have:

$$\begin{aligned} E(\epsilon|Z=z) &= \int_{\mathcal{Y}} \Theta^0(y)dF_{Y|Z}(y,z) - \int_{\mathcal{X}} g^0(x)dF_{X|Z}(x,z) \\ &= \int_{\mathcal{Y}} \bar{\Theta}^0(y)dF_{Y|Z}(y,z) - \int_{\mathcal{X}} [\bar{g}^0(x) + \beta^0(x^{-1})]dF_{X|Z}(x,z) \end{aligned} \tag{35}$$

Now, consider a structure $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ that is observationally equivalent to $(T^0, g^0, F^0_{\epsilon|X})$ and has the same properties as $(T^0, g^0, F^0_{\epsilon|X})$. It follows from (35) that for a.e. $z \in \mathcal{Z}$:

$$E(\epsilon|Z=z) = 0 = E(\tilde{\epsilon}|Z=z) \Rightarrow \int_{\mathcal{X}} [\beta^0(x^{-1}) - \tilde{\beta}^0(x^{-1})] dF_{X|Z}(x,z) = 0$$
$$\Leftrightarrow E\big([\beta^0(X^{-1}) - \tilde{\beta}^0(X^{-1})]\big|Z=z\big) = 0$$

where $\tilde{\epsilon} = \tilde{\Theta}^0(Y) - \tilde{g}^0(X)$. Then, the completeness assumption A7 implies $\beta^0(x^{-1}) = \tilde{\beta}^0(x^{-1})$ for a.e. $x^{-1} \in \mathcal{X}^{-1}$. Combined with equation (36), this implies that for a.e. $x \in \mathcal{X}$,

$$g^0(x) = \tilde{g}^0(x)$$

From $F^0_{\epsilon|X}(\Theta^0(y) - g^0(x), x^{-1}) = \tilde{F}_{\tilde{\epsilon}|X}(\Theta^0(y) - g^0(x), x^{-1})$ for every $y \in \mathcal{Y}$ and a.e. $x \in \mathcal{X}$, and the fact that $\Theta^0(\mathcal{Y}) = \mathbb{R}$, we conclude that for every $t \in \mathbb{R}$ and a.e. $x^{-1} \in \mathcal{X}^{-1}$, $F^0_{\epsilon|X}(t, x^{-1}) = \tilde{F}^0_{\tilde{\epsilon}|X}(t, x^{-1})$.

Step 4. Finally, assume that the completeness condition is violated, in the sense that there exists some function $h : \mathcal{X}^{-1} \to \mathbb{R}$ that (i) does not vanish a.e., but (ii) is such that $E[h(X^{-1}) \mid Z = z] = 0$ for a.e. $z \in \mathcal{Z}$. Let $(T^0, g^0, F^0_{\epsilon|X})$ be a structure generating Φ^0, that satisfies assumptions A1-A4 and the normalization condition (7). Define $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ by

$$\begin{aligned}
\tilde{\Theta}^0(y) &\equiv \Theta^0(y) \\
\tilde{g}^0(x) &\equiv g^0(x) + h(x^{-1}) \\
\tilde{F}^0_{\tilde{\epsilon}|X}(t,x) &\equiv F^0_{\tilde{\epsilon}|X}\left(t + h(x^{-1}), x^{-1}\right)
\end{aligned}$$

for every $y \in \mathcal{Y}$, every $t \in \mathbb{R}$, and a.e. $x \in \mathcal{X}$. Then, the structure $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ satisfies the normalization condition (7), as well as assumptions A1-A4. Note that assumption A4 only requires $\tilde{g}^0$ to be smooth with respect to the first component x_1; hence, it is satisfied even if the function $h(x^{-1})$ is discontinuous. Since the structure $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ is observationally equivalent to $(T^0, g^0, F^0_{\epsilon|X})$, $(T^0, g^0, F^0_{\epsilon|X})$ is not identified. $\square$

Proof of Corollary 6. The proof is similar to that of Proposition 5. From Step 1, we know that Θ^0 is determined up to two constants $K_1 \in \mathbb{R}$ and $K_2 > 0$: for any $y \in \mathcal{Y}$

$$\Theta^0(y) = K_1 + K_2\bar{\Theta}^0(y)$$

where $\bar{\Theta}^0$ is as given in (31). Then, from Step 2, any solution to (32) is of the form:

$$g^0(x) = K_2[\bar{g}^0(x) + \beta^0(x^{-1})] \tag{36}$$

for some function $\beta^0 : \mathcal{X}^1 \to \mathbb{R}$, where $\bar{g}^0$ is as defined in (33). From the normalization condition $E(\epsilon) = E[g(X)] = 0$ we have $E[\Theta^0(Y)] = 0$, so

$$K_1 = -K_2 E[\bar{\Theta}^0(Y)]$$

Now, consider two observationally equivalent structures $(\tilde{T}^0, \tilde{g}^0, \tilde{F}^0_{\tilde{\epsilon}|X})$ and $(T^0, g^0, F^0_{\epsilon|X})$, and let $\tilde{K}_1 \in \mathbb{R}$ and $\tilde{K}_2 > 0$ denote the two constants defining $\tilde{\Theta}^0$. We then have:

$$E(\epsilon|Z) = K_2\Big[E[\bar{\Theta}^0(Y)|Z] - E[\bar{\Theta}^0(Y)] - E[\bar{g}^0(X)|Z] - E[\beta^0(X^{-1})|Z]\Big]$$
$$E(\tilde{\epsilon}|Z) = \tilde{K}_2\left[E[\bar{\Theta}^0(Y)|Z] - E[\bar{\Theta}^0(Y)] - E[\bar{g}^0(X)|Z] - E[\tilde{\beta}^0(X^{-1})|Z]\right]$$

so, since $K_2 > 0$ and $\tilde{K}_2 > 0$, $E(\epsilon|Z) = 0 = E(\tilde{\epsilon}|Z)$ w.p.1 if and only if $E\big([\beta^0(X^{-1}) - \tilde{\beta}^0(X^{-1})]|Z\big) = 0$ w.p.1. Again, by the completeness condition this implies $\beta^0(x^{-1}) = \tilde{\beta}^0(x^{-1})$ for a.e. $x^{-1} \in \mathcal{X}^{-1}$. Therefore, we can write that:

$$\tilde{\epsilon} = \frac{\tilde{K}_2}{K_2}\,\epsilon$$

By using the second normalization condition $\mathrm{var}(\epsilon) = 1 = \mathrm{var}(\tilde{\epsilon})$ we then get that $\tilde{K}_2 = K_2$, which combined with the above gives $\tilde{K}_1 = K_1$. This completes the sufficiency part of the proof. The necessity is same as in the proof of Proposition 5. $\square$

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

Blundell, R., and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Vol. II*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge University Press.

Brown, B. W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175–196.

Buera, F. (2006): "Non-Parametric Identification and Testable Implications of the Roy Model," mimeo, Northwestern University.

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.

Chiappori, P. A., and I. Ekeland (2008): "Identifiability Through Partial Differential Equations," mimeo, Columbia University.

Darolles, S., J. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Centre de Recherche et Développement Économique, 05-2002.

d'Haultfoeuille, X. (2006): "On the Completeness Condition in Nonparametric Instrumental Problems," INSEE, Documents de Travail du CREST 2006-32.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *The Journal of Political Economy*, 112, S60–S109.

Hall, P., and J. L. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904–2929.

Han, A. K. (1987): "A Non-Parametric Analysis of Transformations," *Joumal of Econometrics*, 35, 191–209.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2005): "Estimation and Simulation of Hedonic Models," in *Frontiers in Applied General Equilibrium*, ed. by T. Kehoe, T. Srinivasan, and J. Whalley. Cambridge University Press, Cambridge.

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

Komunjer, I. (2008): "Global Identification in Nonlinear Semiparametric Models," mimeo, University of California San Diego.

Koopmans, T. C., and O. Reiersøl (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21, 165–181.

Linton, O., S. Sperlich, and I. Van Keilegom (2008): "Estimation of a Semiparametric Transformation Model," *The Annals of Statistics*, 36, 686–718.

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

——— (2007): "Nonparametric Identification," in *Handbook of Econometrics, Vol. 6B*, ed. by J. J. Heckman, and E. Leamer, pp. 5307–5368. Elsevier.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.

Severini, T. A., and G. Tripathi (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory*, 22, 258–278.