NONPARAMETRIC IDENTIFICATION IN NONLINEAR SIMULTANEOUS EQUATIONS MODELS: THE CASE OF COVARIANCE RESTRICTIONS

PIERRE-ANDRE CHIAPPORI, IVANA KOMUNJER, AND DENNIS KRISTENSEN

Abstract. Nonlinear systems of simultaneous equations are ubiquitous in Economics. Despite their importance, relatively few results regarding the nonparametric identification and estimation in these models are available. Important advances by Matzkin (2015, ECMA) show that exclusion restrictions, as known in linear simultaneous equations models, can lead to nonparametric identification in nonlinear models. The point of departure of our analysis is a well-known fact (see, e.g., Fisher, 1966) that in linear simultaneous equations systems, traditional exclusion restrictions can under certain conditions be replaced with restrictions on the covariance structure of the unobservables. We extend those results to nonlinear systems with infinite dimensional parameters of interest. The key features of our nonparametric identification result are: (i) it is instrument free in that it does not use instrumental variables; and (ii) it is constructive, in the sense that it leads to a natural nonparametric estimator of the model parameters.

Keywords: simultaneity; endogeneity; exclusion restrictions; independence among unobservables.

Date: February 15, 2017

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. Kristensen: Department of Economics, University College London. Email: d.kristensen@ucl.ac.uk.

Acknowledgments. The authors would like to thank the seminar participants at NYU-AD for comments and suggestions.
Simultaneous equations systems are a signature tool in Social Sciences. They are used to analyze supply–demand models, games, and networks, to name a few examples in Economics. In Sociology, they are used to conduct studies of “peer influences.” Even in their simplest linear form, simultaneous equations systems suffer from a well-known identification problem: their structural parameters generally do not coincide with the reduced-form coefficients obtained from regressions among observables. The analysis of this problem has a long history in Economics, with the first account dating back to Wright (1928), and a systematic treatment given in the collective works by the Cowles Commission (Koopmans, 1950), and a monograph by Fisher (1966). The best known and today most widely applied solution to the identification problem involves finding observed “external factors” or what we would now call instrumental variables (IVs) which shift one of the equations in the system (say, the supply curve) without affecting the others (that is, the demand curve).\(^1\) While the IV solution is well established in the linear case (e.g. linear supply curve, linear demand curve), few results exist for simultaneous equations systems that are nonlinear. The purpose of our paper is to contribute to the relatively small literature focusing on nonlinear simultaneous equations and offer a novel nonparametric identification approach that relaxes the usual IV exclusion restrictions.

To describe our result more precisely, it is useful to revisit a linear example taken from Koopmans (1949): a competitive market for a single commodity whose price \(p\) and quantity \(q\) are determined through a linear supply-demand system,

\[
q + \alpha p + \beta i = u \quad \text{(demand)}
\]

\[
q + \gamma p + \delta r = v \quad \text{(supply)}.
\]

\(^1\)According to Wright (1928), “Such additional factors may be factors which (A) affect demand conditions without affecting cost conditions or which (B) affect cost conditions without affecting demand conditions.” (p.312)
Here, \( u \) and \( v \) are latent demand and supply shifters, respectively, drawn from an unknown probability distribution with mean values equal to zero, \( E[u] = 0 \) and \( E[v] = 0 \). \( i \) (income) and \( r \) (rainfall) are exogenous variables that are independent of \( u \) and \( v \), and known to be excluded from one of the equations. The parameters \( \alpha \), \( \beta \), \( \gamma \), and \( \delta \) (with \( \alpha \neq \gamma \)) are the structural parameters of interest. An application of the well-known rank condition (see, e.g., Koopmans and Reiersøl (1950); Fisher (1966)) shows that the above system is identified. The two exogenous variables \( i \) and \( r \) are the IVs that identify the supply and demand curves, respectively. Their key properties are: (i) exogeneity, \( i \perp (u,v) \), (ii) relevance (\( i \) appears in the demand equation or \( \beta \neq 0 \)), and (iii) exclusion (\( i \) does not appear in the supply equation), with analogous conditions for \( r \).

Now, relax the linearity assumption and consider instead a nonlinear supply-demand system of the form,

\[
\begin{align*}
f(q,p) + h(i,r) &= u \quad \text{(demand)} \\
g(q,p) + k(i,r) &= v \quad \text{(supply)},
\end{align*}
\]

in which the functions \( f(\cdot,\cdot) \), \( g(\cdot,\cdot) \), \( h(\cdot,\cdot) \) and \( k(\cdot,\cdot) \) are unknown (with the bivariate function \( f(\cdot,\cdot) \) invertible). Recent advances by Matzkin (2015) have shown that exclusion restrictions, similar to those found in the linear case, together with a normalization lead to the nonparametric identification of the structural functions \( f(\cdot,\cdot) \) and \( g(\cdot,\cdot) \). Specifically, Matzkin (2015) assumes that \( h(i,r) = i \) and \( k(i,r) = r \). In this sense, \( i \) and \( r \) are the IVs that continue to satisfy the same requirements as before: exogeneity, relevance, and exclusion.

The point of departure of our analysis is a well-known fact that in linear simultaneous equations systems, traditional exclusion restrictions can under certain conditions be replaced by restrictions on the covariance structure of the unobservables, \( u \) and \( v \). For an extensive treatment of this point, see Fisher (1966). Specifically, we propose replacing the exclusion restrictions, \( h(i,r) = i \) and \( k(i,r) = r \), with a normalization
condition on \((f, g)\) and the restriction that \(u \perp v\).\(^2\) In particular, the bivariate function \((h, k)\) is no longer set to be an identity function, and so becomes an additional structural function to identify. To achieve identification, some normalization of the model is needed and we therefore fix the the level and the partial derivatives of \((f, g)\) at two points; this normalization is implicitly imposed in Matzkin (2015) through the assumption that \((h, k)\) is the identity.

Exploiting the independence among unobservables to achieve identification is not a new idea. In linear simultaneous equations systems, the approach dates back to Fisher (1963, 1966). It is routinely used in macroeconomics to achieve identification in structural vector autoregression models; see, for example, Rubio-Ramirez, Waggoner, and Zha (2010). A recent application to semiparametric models with simultaneity can be found in Matzkin (2016). What is new, however, is how the independence is exploited. In the traditional analysis, one can think of the restriction \(u \perp v\) as providing an additional unobservable instrument, that can help identify one of the two equations (supply or demand), given that the other is identified by other means (e.g. using an IV). In our analysis, independence among unobservables is used to “de-couple” the nonlinear simultaneous equations system and write it as a system of seemingly unrelated nonlinear equations with a single endogenous variable and a single disturbance. Nonparametric identification can then proceed equation-by-equation, using existing tools for single equations.

Before proceeding, it is important to comment on the empirical plausibility of such independence restrictions. If the disturbances \(u\) and \(v\) are thought of as representing the variables that have been omitted in the model (as e.g. in the case of product unobservables commonly used in multiple choice models), it seems difficult to justify the restriction \(u \perp v\) on the grounds of some underlying economic theory. In such

\(^2\)The treatment in Fisher (1966) assumes a combination of exclusion (more generally linear parameter) restrictions and covariance restrictions. To keep our analysis focused, we work under covariance restrictions only; our results can however be extended to include a mix of exclusion and covariance restrictions.
situations therefore, the plausibility of our independence restrictions is perhaps not great. If on the other hand, one can think of the disturbances as corresponding to different agents in the model who act independently of each other, the independence assumption may seem more tenable.

Perhaps more importantly, independence is often implicitly imposed in empirical work to ease the burden in estimation. Examples include the assumptions of within-network independence of the error terms (\(?)\) MORE HERE

To explain how the approach works, return to the nonlinear supply-demand system above. Our first step is to show that given the normalization condition, \(u \perp v\) identifies the unknown function \(f(\cdot, \cdot)\) up to an unknown strictly monotonic transform \(\varphi\), that is, the demand equation can be written as,

\[
\varphi(m) + h(i, r) = u \quad \text{(demand)},
\]

where \(m = m(q, p)\) is a known function of the quantity \(q\) and price \(p\). We can therefore treat \(m\) as an observed endogenous variable, and the above equation is nothing but a single equation transformation model, in which the nonparametric identification of \(\varphi(\cdot), h(\cdot, \cdot)\), and the distribution of \(u\), is well established; see, e.g., Chiappori, Komunjer, and Kristensen (2015). This argument naturally generalizes to systems with \(G \geq 2\) equations. In such systems, independence among unobservables provides \(G - 1\) restrictions that identify the structural function \(f\) (which is a function of \(G\) endogenous variables) up to a strictly monotonic transform. Hence, the nonparametric identification argument can proceed as before. It is important to emphasize again, that our approach puts few restrictions on the structural function \(h\). In particular, it is not necessary that some of the exogenous variables be excluded from some of the equations in the system. In this sense, our approach is free of IVs. In applications where no natural exclusion restrictions are available, but independence between the errors is reasonable, our identification result will be useful.
The remainder of our paper is organized as follows. In Section 2, we present the model and introduce our assumptions. Section 3 derives the nonparametric identification result in two steps: first, we use our independence restriction to “de-couple” the original equations into a system of seemingly unrelated single endogenous variable equations with a single disturbance; second, in this decoupled system, identification proceeds equation-by-equation. Our last section concludes.

2. Model and Assumptions

The models that we consider take the form:

$$Y = \Lambda(V(X, Z) + \varepsilon, Z),$$

where $Y \in \mathbb{R}^G$, $X \in \mathcal{X} \subseteq \mathbb{R}^G$, and $Z \in \mathcal{Z} \subseteq \mathbb{R}^K$ are observed, $\varepsilon \in \mathbb{R}^G$ is latent, and the functions $\Lambda : \mathbb{R}^G \times \mathcal{Z} \to \mathbb{R}^G$ and $V : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}^G$ are unknown. While few assumptions shall be made on the supports $\mathcal{X}$ and $\mathcal{Z}$ of the regressors $X$ and $Z$, respectively, we hereafter maintain that the disturbances $\varepsilon$ have full support $\mathbb{R}^G$.\(^3\)

The regressors $X$ play a special role in that they are restricted to have the same dimension as $Y$, and to only enter the right hand side of the model through $V(X, Z)$. We will moreover require that $X$ has a continuous distribution and satisfies $\varepsilon \perp X \mid Z$. This will prove to be an important part of our identification argument.\(^4\) In other words, of the total set of regressors $(X, Z)$, we require that $G$ of these are continuously distributed and exogenous. The remaining regressors $Z$ are on the other hand allowed to be endogenous and potentially discrete, and may also affect $Y$ directly through the second argument of $\Lambda$. We shall work under the assumption

---

\(^3\)Under our assumptions on $\Lambda$ stated below, this in particular implies that $Y$ takes values in all of $\mathbb{R}^G$. Cases in which the support of $Y$ is a known strict subset of $\mathbb{R}^G$, such as for example the $G$-dimensional simplex (as is the case when $Y$ represents the vector of market shares, for example) are easily made part of our setup by transforming $Y$ appropriately.

\(^4\)The mathematical reason for this requirement is transparent from the invertibility condition on the matrix of partial derivatives of $V$ with respect to $X$, which we impose in the Normalization Condition stated later on.
that for all $z \in Z$, the mapping $\Lambda(\cdot, z)$ is twice continuously differentiable and invertible with everywhere nonvanishing Jacobian. The invertibility assumption on $\Lambda(\cdot, z)$ implies that the model in (1) can equivalently be expressed as:

\[
\Theta(Y, Z) = V(X, Z) + \varepsilon \quad \text{where} \quad \Theta(\cdot, z) \equiv \Lambda^{-1}(\cdot, z).
\]

This is a multivariate nonlinear regression model in which the vector of dependent variables $\Theta(Y, Z)$ is unknown.

To formally derive our results, we hereafter maintain the following assumptions.

**Assumption A1.** For all $z \in Z$, the conditional distribution of $X$ given $Z = z$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^G$.

**Assumption A2.** $\varepsilon \perp X \mid Z$ with a density $f_{\varepsilon|Z}(\cdot|z)$ that for all $z \in Z$ is strictly positive and twice continuously differentiable on $\mathbb{R}^G$ possibly except on a set of isolated points.

**Assumption A3.** For all $z \in Z$, the function $\Lambda(\cdot, z) : \mathbb{R}^G \to \mathbb{R}^G$ is invertible on $\mathbb{R}^G$, and twice continuously differentiable with Jacobian

\[
\det \left( \frac{\partial \Lambda(t, z)}{\partial t} \right) \neq 0 \quad \text{for all } t \in \mathbb{R}^G.
\]

**Assumption A4.** For all $z \in Z$, the function $V(\cdot, z) : \mathcal{X} \to \mathbb{R}^G$ is continuously differentiable on $\mathcal{X}$, possibly except on a set of isolated points, and for every $z \in Z$ there exists $x_z^* \in \mathcal{X}$ such that $\det (\partial V(x_z^*, z)/\partial x) \neq 0$.

**Assumption A5.** The components of $\varepsilon$ are mutually independent conditional on $Z$, so that for all $z \in Z$ and $t = (t_1, \ldots, t_G) \in \mathbb{R}^G$,

\[
f_{\varepsilon|Z}(t|z) = \prod_{g=1}^G f_{\varepsilon_g|Z}(t_g|z).
\]

**Assumption A6.** For all $z \in Z$, the inverse mapping $\Theta(\cdot, z) : \mathbb{R}^G \to \mathbb{R}^G$ is such that for every $1 \leq g \leq G$ there exists $1 \leq h \leq G$ such that

\[
\frac{\partial \Theta_h(y, z)}{\partial y_h} \neq 0 \quad \text{for every } y \in \mathbb{R}^G.
\]
Assumption A7. For all \( z \in Z \) and \( 1 \leq g \leq G \), the marginal conditional density \( f_{\epsilon|Z}(\cdot|z) \) satisfies
\[
\frac{\partial^2 \ln f_{\epsilon|Z}(t_g|z)}{\partial t_g^2} \neq 0 \quad \text{for almost every } t_g \in \mathbb{R}.
\]

Assumption A1 explicitly requires that the exogenous regressors \( X \) be continuously distributed given \( Z \). The reason why continuity is required is that our approach exploits the information in the partial derivatives of various observed quantities with respect to the components of \( X \). For those partial derivatives to be well defined, continuity of \( X \) is needed. Note, however, that no further restrictions are put on the support \( \mathcal{X} \) of \( X \). In particular, \( \mathcal{X} \) is allowed to be bounded. Moreover, no distributional assumptions are imposed on the remaining regressors, \( Z \). The next two assumptions are fairly standard, see, e.g., Matzkin (2015). Assumption A2 is the usual conditional independence (or exogeneity) restriction between the disturbances \( \epsilon \) and the regressors \( X \). The support of \( \epsilon | Z \) is assumed to be the entire Euclidean space \( \mathbb{R}^G \). Full support of \( \epsilon | Z = z \) together with the ontoness of the map \( \Lambda(\cdot, z) \) in Assumption A3 implies that (given \( Z = z \)) the dependent variables in \( Y \) have full support on \( \mathbb{R}^G \). Note that Assumption A3 implies that either \( \det (\partial \Lambda(\cdot, z) / \partial s) > 0 \) or \( \det (\partial \Lambda(\cdot, z) / \partial s) < 0 \). As we shall see later on, the sign of this Jacobian is identified, so we do not need to take a stand on its sign. It is worth pointing out that Assumption A2 allows \( \epsilon \) to have a conditional density that is non-differentiable at isolated points in \( \mathbb{R}^G \).

Under Assumptions A2 and A3, the conditional distribution of \( Y \) given \((X, Z)\) has a density of the form:
\[
(3) \quad f_{Y|(X,Z)}(y|x, z) = \left| \det \left( \frac{\partial \Theta(y, z)}{\partial y} \right) \right| f_{\epsilon|Z}(\Theta(y, z) - V(x, z)|z),
\]
which is everywhere strictly positive. To ensure that \( f_{Y|(X,Z)}(\cdot|z) \) is smooth enough so that the second cross-partial \( \partial^2 f_{Y|(X,Z)}(\cdot, z)/\partial y_g \partial x_h \) \((1 \leq g, h \leq G)\) all exist and are continuous, additional restrictions need to be made on \( V(\cdot, z) \). Assumption A4 ensures that \( V(\cdot, z) \) is continuously differentiable on \( \mathcal{X} \), possibly except on a
set of isolated points. Note that this smoothness requirement excludes piecewise linear regression functions. The second requirement in Assumption A4 regarding the rank of the matrix of partial derivatives $\partial V(\cdot, z)/\partial x$ in particular implies that the $G$ exogenous regressors need to be distinct. AND LINEARLY INDEPENDENT?

It is worth emphasizing that we only require the full rank condition to hold at one point $x^*_z$ (at which the function $V(\cdot, z)$ is obviously required to be differentiable). Perhaps more importantly, we shall demonstrate below that the existence of such a point $x^*_z$ is testable, and that moreover $x^*_z$ is easy to determine given the observable data.

Unlike the first four assumptions, Assumption A5 is new to the literature. Its role is to replace some of the usual exclusion restrictions on the regressors $X$ by mutual independence of the error components.

Assumption A6 strengthens some of the requirements in Assumption A3. Specifically, note that under A3 the Jacobian of $\Theta(\cdot, z)$ is everywhere non-vanishing. This implies in particular that for any realization $y$ of $Y$ and any component $g$ of $\Theta(\cdot, z)$, there exists a coordinate $h$ of $y$ such that the partial derivative $\partial \Theta_g(y, z)/\partial y_h \neq 0$. Otherwise, the $g$th column of the matrix of partial derivatives of $\Theta(y, z)$ would be equal to zero, which contradicts the nonzero determinant assumption. It is however possible under Assumption A3 that the index $h$ of the non-zero partial derivative changes with $y$. The goal of Assumption A6 is to preclude this possibility. Thus, in each column of the matrix of partial derivatives of $\Theta(\cdot, z)$ with respect to the components of $y$, there is at least one row that is almost nonzero.

Finally, Assumption A7 eliminates disturbance densities that are piecewise exponential (such as the double-exponential for example). The role of this assumption is to give bite to the independence restriction in Assumption A5. Indeed, under Assumption A5, $\ln f_{\varepsilon|Z}(t|z) = \sum_{g=1}^{G} \ln f_{\varepsilon_g|Z}(t_g, z)$ which is equivalent to the restriction that $\partial^2 \ln f_{\varepsilon|Z}(t, z)/\partial t_g \partial t_h = 0$ for $g \neq h$. This condition on the cross-partials will only have interesting implications if the second-order derivatives $\partial^2 \ln f_{\varepsilon|Z}(t, z)/\partial t^2_g$ are all non-zero.
Now, consider again $f_{Y\mid(X,Z)}(y \mid x, z)$ as derived in (3). Taking logarithms and twice differentiating with respect to $(y_g, x_h)$ for every $1 \leq (g, h) \leq G$ gives:

$$
\frac{\partial^2 \ln f_{Y\mid(X,Z)}(y \mid x, z)}{\partial y_g \partial x_h} = -\sum_{i=1}^{G} \frac{\partial V_i(x, z)}{\partial x_h} \frac{\partial \Theta_i(y, z)}{\partial y_g} \frac{\partial^2 \ln f_{\varepsilon_i \mid Z}(\Theta_i(y, z) - V_i(x, z), z)}{\partial t_i^2}.
$$

Note that in the above expression there are no cross partial derivatives with respect $t_i$ and $t_j$ ($j \neq i$); this is because under the assumed independence of the components of $\varepsilon$ their joint log-density is additively separable in the log-marginal densities. We can write the above $G \times G$ equations more compactly in matrix notation. Let $D(y, x, z)$ be the $G \times G$ matrix of second order derivatives of the log-conditional density of $\varepsilon$ given $Z$, $D(y, x, z) \equiv \text{diag} (D_1(y, x, z), \ldots, D_G(y, x, z))$, where

$$
D_g(y, x, z) \equiv \frac{\partial^2 \ln f_{\varepsilon_g \mid Z}(\Theta_g(y, z) - V_g(x, z), z)}{\partial t_g^2}, \quad 1 \leq g \leq G.
$$

Then,

$$
(4) \quad \left( \frac{\partial^2 \ln f_{Y\mid(X,Z)}(y \mid x, z)}{\partial y \partial x} \right) = -\left( \frac{\partial \Theta(y, z)}{\partial y} \right)' D(y, x, z) \left( \frac{\partial V(x, z)}{\partial x} \right).
$$

This matrix equality will prove to be crucial in obtaining our results. It is also useful for testing the rank requirement in Assumption A4. Indeed, fix any realization $z$ of $Z$. Then, any $x_z \in X$ such that the matrix of second cross-partial $\partial f_{Y\mid(X,Z)}(y \mid x, z)/\partial y \partial x$ is full rank for all $y \in \mathbb{R}^G$, possibly except on a set of isolated points, satisfies the full rank requirement in Assumption A4.

3. Identification

As usual, call the triple $\mathcal{S} \equiv (\Theta, V, f_{\varepsilon \mid Z})$ a structure, and let the model consist of all the structures that satisfy Assumptions A1 through A7. According to equation (3), each structure implies a unique conditional density of the observables; thus we can write $f_{Y\mid(X,Z)}(y \mid x, z; \mathcal{S})$. We shall say that two structures $\mathcal{S} = (\Theta, V, f_{\varepsilon \mid Z})$ and $\tilde{\mathcal{S}} \equiv (\tilde{\Theta}, \tilde{V}, \tilde{f}_{\varepsilon \mid Z})$ in the model are observationally equivalent, a property which we denote by $\mathcal{S} \sim \tilde{\mathcal{S}}$, if they imply the same conditional distribution of the observables,
i.e. if for almost every value of \((y, x, z)\) we have:

\[
f_{Y|X,Z}(y|x, z; \mathcal{S}) = f_{Y|X,Z}(y|x, z; \tilde{\mathcal{S}}).
\]

The structure \(\mathcal{S}\) is said to be identified within the model if \(\mathcal{S} \sim \tilde{\mathcal{S}}\) implies \(\tilde{\mathcal{S}} = \mathcal{S}\).\(^5\)

It is clear that without some normalization, the structure \(\mathcal{S}\) is not identifiable. Indeed, take any matrix of functions \(A : \mathcal{Z} \rightarrow \mathbb{R}^{G \times G}\) such that \(\det A(z) \neq 0\) for all \(z \in \mathcal{Z}\), and any two vector functions \(b : \mathcal{Z} \rightarrow \mathbb{R}^G\) and \(c : \mathcal{Z} \rightarrow \mathbb{R}^G\). Define a new structure \(\tilde{\mathcal{S}} \equiv (\tilde{\Theta}, \tilde{V}, \tilde{f}_{\varepsilon|Z})\) with

\[
\tilde{V}(x, z) = A(z)V(x, z) + b(z), \\
\tilde{\Theta}(y, z) = A(z)\Theta(y, z) + c(z), \text{ and} \\
\tilde{f}_{\varepsilon|Z}(t|z) = \frac{1}{|\det A(z)|} f_{\varepsilon|Z}(A(z)^{-1}(t + b(z) - c(z))|z).
\]

For appropriate choices of \(A(z)\), the structure \(\tilde{\mathcal{S}}\) will satisfy Assumptions A1-A7 and therefore be part of the model. It will then follow immediately from (3) that \(\mathcal{S} \sim \tilde{\mathcal{S}}\). To rule out such trivial transformations between observationally equivalent structures, some normalization conditions are needed that fix the location \((b(z), c(z))\) and scale \(A(z)\) functions.

To properly choose a normalization, we first need to discuss the choices of \(A(z)\) that will lead to disturbances \(\tilde{\varepsilon} \equiv A(z)\varepsilon\) with independent components. It is well-known ICA REFERENCE? that the answer depends on the number of Gaussian components of \(\varepsilon\). If at most one component of \(\varepsilon\) is Gaussian, then for any \(A(z) = D(z)P(z)\) with \(D(z) \in \mathbb{R}^{G \times G}\) diagonal and \(P(z) \in \mathbb{R}^{G \times G}\) a permutation matrix, \(\tilde{\varepsilon} = A(z)\varepsilon\) satisfies the independence Assumption A5 whenever \(\varepsilon\) does so. If \(\varepsilon\) is multivariate Gaussian, then taking \(A(z) = D(z)U(z)\) with \(U(z) \in \mathbb{R}^{G \times G}\) orthogonal leads to the same conclusion. Since \(A(z)\) has more free parameters in the latter case \((G(G + 1)/2)\) than in the former \((G)\), we impose the following.

\(^5\)Formally, equality of structures entails equality of conditional densities where the latter is to be understood to hold almost everywhere.
Normalization Condition. For every realization $z \in Z$, there exists known value $y^*_z \in \mathbb{R}^G$ such that the lower triangular part of $\partial \Theta(y^*_z, z)/\partial y$ is known.

For every realization $z \in Z$, there exist known values $\bar{x}_z$ and $x^*_z$ in $\mathcal{X}$, and functions $\bar{v}(z) \in \mathbb{R}^G$ and $V^*(z) \in \mathbb{R}^{G \times G}$ with $\det V^*(z) \neq 0$, such that:

$$V(\bar{x}_z, z) = \bar{v}(z) \quad \text{and} \quad \frac{\partial V(x^*_z, z)}{\partial x} = V^*(z).$$

Moreover, $E [\varepsilon | Z] = 0$.

Our normalization condition imposes $G(G + 1)/2 + 2G$ restrictions, one for each free component of $(A(z), b(z), c(z))$. There is a variety of normalization conditions that the researcher can choose from, mostly on an ad hoc basis. Some theoretical justification for choosing “smooth” conditions, such as those involving integrals of functions, rather than those involving the derivatives can be found in Chiappori, Komunjer, and Kristensen (2015). These results pertain however to single equation models, and tends to make the identification argument more involved. We therefore here choose to work with a normalization that fixes the values of the regression function $V(\cdot, z)$ and its derivative at a point. These conditions facilitate the comparison of Matzkin’s (2015) results with the ones obtained here. Omitting the variable $Z$, Matzkin (2015) imposes $V(x) = x$, for all $x \in \mathcal{X}$, which in turn implies $\partial V(x)/\partial x = \text{Id}$, for all $x \in \mathcal{X}$. Compared to this normalization, which pins down $V$ completely, we here impose a weaker one where the level and derivatives, respectively, of $V$ are only pinned down at single points, $\bar{x}$ and $x^*$.

Several remarks are in order. First, when $Z$ appears in the model, the normalization conditions on the regression function are imposed for every realization $z$ of $Z$. The points $\bar{x}_z$ and $x^*_z$ at which the level and derivatives of $V(\cdot, z)$ are normalized are allowed to vary with $z$. These could be fixed values, $\bar{x}$ and $x^*$, such that $V(\bar{x}, z) = \bar{v}(z)$ and $\partial V(x^*, z)/\partial x = V^*(z)$, but we here allow for added flexibility in the normalization. Second, fixing the level and the derivatives of the regression function only allow us to pin down the location $b(z)$ and scale $A(z)$ functions. In order to pin down the location $c(z)$, a normalization on either $\Theta(\cdot, z)$ or $f_{\varepsilon | Z}(\cdot | z)$
needs to be imposed. We here choose to normalize a particular integral of the latter, \( E[\varepsilon|Z] = 0 \), as such conditions are common in applied work.

Finally, it is useful to relate our conditions to those found in linear simultaneous equations models of the form \( \Theta Y = c + VX + \varepsilon \), in which \( c, \Theta \) and \( V \) are, respectively, \( G \times 1, G \times G \) and \( G \times G \) parameter matrices (see, e.g. Koopmans, 1950; Fisher, 1966). In these models, it is typically assumed that: (i) each row of \( \Theta \) has one coefficient normalized to 1 (\( G \) restrictions); (ii) for each equation, there exists at least \( G - 1 \) linear \textit{a priori} restrictions on the remaining components of \( \Theta, c, \) and \( V \) (\( G(G-1) \) restrictions); (iii) there is only one intercept \( c \) on the right hand side (\( G \) restrictions); and (iv) \( E(\varepsilon) = 0 \) (\( G \) restrictions). Thus, the total number of restrictions imposed is \( G(G+2) \). This, moreover is the minimal number of restrictions needed in order to identify the linear model. In our case, there are \( G(G+1)/2 + 2G \) restrictions imposed in the Normalization Condition, to which one should add \( G(G-1)/2 \) restrictions implied by the independence Assumption A7. Thus, the total number of restrictions imposed is again \( G(G+2) \) as in the linear case. In this sense, one could claim that the number of restrictions imposed in this paper is the minimal number that is necessary to achieve identification.

Given the assumptions presented in the previous section and having eliminated trivial transformations between observationally equivalent structures, we are now ready to state the main result of the paper:

**Theorem 1.** Under Assumptions A1 through A7 together with the Normalization Condition, the structure \( S \) is identified.

There exist two broad categories of results on identification. The first gives necessary and sufficient conditions for observational equivalence to imply equality of structures. The best known example in this category are the rank conditions for identification established in Koopmans (1949); see also the collective work by the Cowles Commission (Koopmans, 1950), as well as Fisher (1966). Though necessary and sufficient for identification, such conditions are not constructive in a sense that
they do not naturally lead to an estimator of the quantities of interest. The second family of results establishes identification by inverting the mapping from $\mathcal{S}$ to $f(y|x,z;\mathcal{S})$, i.e. by explicitly computing the structure $\mathcal{S}$ that generated the conditional density $f(y|x,z;\mathcal{S})$. Such approaches are constructive. As will be clear from the proof of the above result, ours falls in the latter category.

The rest of this section is devoted to the proof of the main result, which proceeds in two steps. We first show that the system of simultaneous equations can be decoupled in the following sense: There exist some identified real functions $\tilde{\Theta}_g(Y_1,\ldots,Y_G,Z)$ such that

$$(5) \quad \Theta_g(Y_1,\ldots,Y_G,Z) = \Psi_g(\tilde{\Theta}_g(Y_1,\ldots,Y_G,Z),Z),$$

for all $1 \leq g \leq G$, where $\Psi_g(\cdot,z) : \mathbb{R} \to \mathbb{R}$ is an (unknown) continuously differentiable strictly monotonic function. Put in words, every component $\Theta_g$ of the structural function $\Theta$ is identified up to an unknown strictly monotonic transformation $\Psi_g$. This step is achieved using the assumption of independence of the components $(\varepsilon_1,\ldots,\varepsilon_G)$ given $Z = z$. Defining $\bar{Y}_g \equiv \tilde{\Theta}_g(Y_1,\ldots,Y_G,Z)$ ($1 \leq g \leq G$), we can rewrite the original system (2) in terms of these new variables as

$$(6) \quad \Psi_g(\bar{Y}_g,Z) = V_g(X_1,\ldots,X_G,Z) + \varepsilon_g.$$ 

Effectively, in the first step we have identified $G$ transformations $\tilde{\Theta}_g$ ($1 \leq g \leq G$) of the original $G$ dependent variables $(Y_1,\ldots,Y_G)$ which de-couple the original nonlinear simultaneous equations system in (2), and write it as a set of $G$ seemingly unrelated univariate transformation models involving the new variables $(\bar{Y}_1,\ldots,\bar{Y}_G)$. These can be treated as observed since $\tilde{\Theta}$ is identified and therefore known to us.

In the second step, we consider each of the single equation transformation models (6) separately, and show that $\Psi_g, V_g$ and the distribution of $\varepsilon_g|Z$ are identified. This in turn implies that $\Theta_g(y,z) = \Psi_g(\tilde{\Theta}_g(y,z),z)$ is also identified and so Theorem 1 is established.
4. Proof

4.1. Step 1: Decoupling of the System. The set of simultaneous equations can be written as

\[ \Theta_1(Y_1, \ldots, Y_G, Z) = V_1(X_1, \ldots, X_G, Z) + \varepsilon_1, \]

\[ \vdots \]

\[ \Theta_G(Y_1, \ldots, Y_G, Z) = V_G(X_1, \ldots, X_G, Z) + \varepsilon_G. \]

We then wish to derive functions \( \bar{\Theta}_g \) and \( \Psi_g \) \((1 \leq g \leq G)\), so that equation (5) holds. To start, fix any value \( z \) of \( Z \) and consider again the matrix equality derived in (4). In the following, \( z \) will remain fixed and so we suppress any dependence on \( Z \) and \( z \).

Using the full rank condition in Assumption A4 and the Normalization Condition, evaluate (4) at \((x^*, y)\) and \((x^*, y^*)\). Combining the two sets of equations yields:

\[
\left( \frac{\partial \Theta(y)}{\partial y} \right)' D(y, x^*) D(y^*, x^*)^{-1} = \left( \frac{\partial^2 \ln f_{Y|X}(y|x^*)}{\partial y \partial x} \right) \left( \frac{\partial^2 \ln f_{Y|X}(y^*|x^*)}{\partial y \partial x} \right)^{-1} \left( \frac{\partial \Theta(y)}{\partial y} \right)'.
\]

Now, consider the terms in the last columns of both sides of the above system: the term in the \( i \)th row and \( G \)th column \((1 \leq i \leq G)\) is of the form

\[
\frac{\partial \Theta_G(y)}{\partial y_i} D_G(y, x^*) D_G(y^*, x^*)^{-1} = \left( \frac{\partial^2 \ln f_{Y|X}(y|x^*)}{\partial y_i \partial x} \right) \left( \frac{\partial^2 \ln f_{Y|X}(y^*|x^*)}{\partial y_i \partial x} \right)^{-1} \left( \frac{\partial \Theta_G(y)}{\partial y} \right)'.
\]

where \( (\partial^2 \ln f_{Y|X}/\partial y_i \partial x) \) denotes the \( i \)th row of the matrix of the second cross-partial, \( (\partial^2 \ln f_{Y|X}/\partial y \partial x) \), while \( (\partial \Theta_G/\partial y)' \) denotes the \( G \)th column of \( (\partial \Theta/\partial y)' \).

By the Normalization Condition, the right-hand side of Equation (7) is known. The idea then is to take ratios obtained for different values of \( i \) while keeping \( G \) fixed. The validity of this approach hinges on whether or not the quantities on lefthand side of (7) are nonzero. This is where Assumptions A6 and A7 play a key role. First, recall that under Assumption A7, the second order derivatives \( \partial^2 \ln f_{\varepsilon_g}(\cdot)/\partial t_g^2 \)
are nonzero almost everywhere on \( \mathbb{R} \); this, combined with the invertibility of \( \Theta(\cdot) \) guarantees that \( D_y(y, x^*) \neq 0 \) for almost every \( y \in \mathbb{R}^G \). 

**MAYBE WE NEED NONZERO EVERYWHERE BECAUSE WE ARE EVALUATING AT \( y^* \)?** Next, we use the fact (Assumption A6) that there exists at least one component \( h \) such that

\[
\frac{\partial \Theta_G(\cdot)}{\partial y_h} \neq 0,
\]

almost everywhere on \( \mathbb{R}^G \). Taking ratios is then justified and for every \( 1 \leq i \leq G \) we have:

\[
\frac{\partial \Theta_G(y) / \partial y_i}{\partial \Theta_G(y) / \partial y_h} = \frac{\partial^2 \ln f_{Y|X}(y|x^*) / \partial y_i \partial x}{\partial^2 \ln f_{Y|X}(y|x^*) / \partial y_h \partial x} \left( \partial \Theta_G(y^*) / \partial y \right)^{-1} \frac{\partial \Theta_G(y^*) / \partial y}{\partial \Theta_G(y^*) / \partial y_h},
\]

where the right-hand side is observed. We now show that the system of partial differential equations (8)-(9) determines \( \Theta_G(\cdot) \) up to a strictly increasing transformation that depends on \( z \). For this, let \( \tilde{\Theta}_G(y) \) denote any particular solution to (8)-(9). Then, \( \partial \tilde{\Theta}_G(y) / (\partial y_h) \neq 0 \) and

\[
\frac{\partial \Theta_G(y) / \partial y_i}{\partial \Theta_G(y) / \partial y_h} = \frac{\partial \tilde{\Theta}_G(y) / \partial y_i}{\partial \tilde{\Theta}_G(y) / \partial y_h}.
\]

Now, consider a change of variable \( \Gamma(y_1, \ldots, y_G) \equiv (y_1, \ldots, y_{h-1}, \tilde{\Theta}_G(y), y_{h+1}, \ldots, y_G) \). Note that \( \Gamma(\cdot) \) is continuously differentiable on \( \mathbb{R}^G \), \( \det (\partial \Gamma(y) / \partial y) = \partial \tilde{\Theta}_G(y) / \partial y_h \neq 0 \), and \( \Gamma(\cdot) \) is one-to-one on \( \mathbb{R}^G \). Hence, \( \Gamma(\cdot) \) is a coordinate transformation (or a \( C^1 \) diffeomorphism); see, e.g. Definition 15.9 in Apostol (1974). Performing the change of variable \( \Gamma(\cdot) \) we can write

\[
\Theta_G(y_1, \ldots, y_G) = \Theta_G(\Gamma^{-1}(\Gamma(y))) \equiv \Psi_G(\Gamma(y)),
\]
where the function $\Psi_G(\cdot) = \Theta_G (\Gamma^{-1}(\cdot))$ is continuously differentiable. Differentiating both sides of the last equation wrt $y_k$ ($k \neq h$) and $y_h$ we obtain

\begin{align*}
\frac{\partial \Theta_G}{\partial y_k}(y) &= \frac{\partial \Psi_G}{\partial y_k}(y-h, \bar{\Theta}_G) + \frac{\partial \Psi_G}{\partial \bar{\Theta}_G}(y-h, \bar{\Theta}_G) \frac{\partial \bar{\Theta}_G}{\partial y_k}(y) \\
\frac{\partial \Theta_G}{\partial y_h}(y) &= \frac{\partial \Psi_G}{\partial \bar{\Theta}_G}(y-h, \bar{\Theta}_G) \frac{\partial \bar{\Theta}_G}{\partial y_h}(y),
\end{align*}

where $\frac{\partial \Psi_G}{\partial \bar{\Theta}_G}$ denotes the partial derivative of $\Psi_G$ with respect to its $h$th argument, and we use $\Psi_G(y-h, \bar{\Theta}_G)$ to denote $\Psi_G(y_1, \ldots, y_{h-1}, \bar{\Theta}_G, y_{h+1}, \ldots, y_G)$. It follows from (8) and (12) that $\frac{\partial \Psi_G(y-h, \bar{\Theta}_G)}{\partial \bar{\Theta}_G} \neq 0$ almost everywhere on $\mathbb{R}^G$.

Moreover, taking the ratios of (11) and (12) we obtain

\[ \frac{\partial \Theta_G(y)}{\partial y_k} \frac{\partial y_k}{\partial \Theta_G(y)} = \frac{\partial \bar{\Theta}_G(y)}{\partial \bar{\Theta}_G(y)} + \frac{\partial \Psi_G(y-h, \bar{\Theta}_G)}{\partial \bar{\Theta}_G(y)} \frac{\partial \bar{\Theta}_G(y)}{\partial y_h} \frac{1}{\partial \Psi_G(y-h, \bar{\Theta}_G)} \]

Combining the above with (10) then gives $\partial \Psi_G(y-h, \bar{\Theta}_G)/\partial y_k = 0$ almost everywhere on $\mathbb{R}^G$, for all $k \neq h$. Thus, for a known function $\bar{\Theta}_G(\cdot)$, we have

\[ \Theta_G(y) = \Psi_G (\bar{\Theta}_G(y)), \]

with $\Psi_G : \mathbb{R} \to \mathbb{R}$ and $\partial \Psi_G(u)/\partial u \neq 0$ on $\mathbb{R}$. This establishes the identification of $\Theta_G$ up to a strictly monotonic transform $\Psi_G$.

**HERE!**

4.2. **Step 2: Identification of the Univariate Transformation Model.** For any given $1 \leq g \leq G$, consider the univariate transformation model (6). As before, we will suppress dependence on $Z$ since all arguments will be made conditional on $Z = z$, where $z$ remains fixed. As a first step, note that the Normalization Condition implies that: $V_g(\bar{x}) = \bar{v}_g$ and $\partial V_g(x^*)/\partial x = V^*_g$ where $V^*_g = (V^*_{g,1}, \ldots, V^*_{g,G})'$ is the $g$th column of $V^*$, and $\bar{x} (= \bar{x}_z)$, $x^* (= x^*_z)$ and $V^* (= V^*_z)$ are defined in the Normalization Condition. The full rank condition imposed on $V^*$ implies that $V^*_g \neq 0$ and so there exists at least one component $1 \leq h \leq G$ such that $V^*_{g,h} \neq 0$. Without loss of generality, we assume that this is the first component, $V^*_{g,1} \neq 0$. Now, since $g$
will also remain fixed in the following we also suppress the $g$ subscript and write the $g$th transformation model as:

\[(14) \quad \Psi(\bar{Y}) = V(X_1, \ldots, X_G) + \varepsilon,\]

where $\bar{Y} \in \mathbb{R}$ is a scalar observed variable, $\varepsilon \in \mathbb{R}$ is a scalar disturbance which is independent of $X$, while $\Psi(\cdot) : \mathbb{R} \to \mathbb{R}$ and $V(\cdot) : \mathcal{X} \to \mathbb{R}$ are both real valued.

To show identification of $\Psi$, $V$ and the distribution of $\varepsilon$, observe that the cumulative distribution function of $Y|X$ satisfies

\[
F_{Y|X}(y|x) = F_\varepsilon(\Psi(y) - V(x_1, \ldots, x_G))
\]

for every value $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Using that $V(\cdot)$ is continuously differentiable almost everywhere on $\mathcal{X}$ (Assumption A4), we can differentiate the above with respect to $y$ and $x_g$ ($1 \leq g \leq G$) to obtain

\[
\frac{\partial F_{Y|X}(y|x)}{\partial y} = \frac{\partial \Psi(y)}{\partial y} f_\varepsilon(\Psi(y) - V(x_1, \ldots, x_G)),
\]
\[
\frac{\partial F_{Y|X}(y|x)}{\partial x_g} = -\frac{\partial V(x)}{\partial x_g} f_\varepsilon(\Psi(y) - V(x_1, \ldots, x_G)).
\]

Evaluating the above equations at $X = x^*$, and taking the ratio of the partials wrt $y$ and $x_1$, which is allowed given our assumptions, gives

\[
\frac{\partial \Psi(y)}{\partial y} = -\frac{V_1^*}{r_1(y, x^*)},
\]

for almost every $y$, where $V_1^* = \partial V(x^*) / (\partial x_1) \neq 0$ due to the Normalization Condition and

\[
r_g(y, x) \equiv \frac{\partial F_{Y|X}(y|x) / \partial x_g}{\partial F_{Y|X}(y|x) / \partial y}
\]

is a function that is observed. Integrating with respect to $y$ yields

\[(15) \quad \Psi(y) = -V_1^* \int_{\bar{y}}^y \frac{1}{r_1(u, x^*)} du + \Psi(\bar{y}),\]

for any given $\bar{y} \in \mathbb{R}$. Note that the first term on the right-hand side above is a known function of $y$; thus, $\Psi(y)$ is identified up to an unknown constant $\Psi(\bar{y})$. In particular, the sign of its partial derivative with respect to $y$ is identified.
Next, using the obtained expression for $\partial \Psi(y)/\partial y$, and taking the ratio of the partials wrt $x_1$ and $y$ gives
\[
\frac{\partial V(x, z)}{\partial x_1} = V^*_1 \frac{r_1(y, x)}{r_1(y, x^*)},
\]
for almost every $x$. Integrating with respect to $x_1$ yields
\[
(16) \quad V(x, z) = V^*_1 \int_{\tilde{x}_1}^{x_1} \frac{r_1(y, s, x_2, \ldots, x_G)}{r_1(y, x^*)} ds + V(\tilde{x}_1, x_2, \ldots, x_G).
\]
Note again that the first term on the right-hand side above is known. The second term is an unknown function of $(x_2, \ldots, x_G)$. This term can further be determined using the information in the partials wrt to the other components of $x$. Specifically, taking the ratio of the partials wrt $x_2$ and $y$ gives
\[
\frac{\partial V(x, z)}{\partial x_2} = V^*_1 \frac{r_2(y, x)}{r_1(y, x^*)}.
\]
Combining the above with (16) then yields
\[
\frac{\partial V(\tilde{x}_1, x_2, \ldots, x_G)}{\partial x_2} = V^*_1 \left[ \frac{r_2(y, x)}{r_1(y, x^*)} - \int_{\tilde{x}_1}^{x_1} \frac{\partial r_1(y, s, x_2, \ldots, x_G)}{\partial x_2} ds \right].
\]
Note again that the right-hand side term is an observed function of $(x_2, \ldots, x_G)$, so integrating with respect to $x_2$ yields
\[
V(\tilde{x}_1, x_2, \ldots, x_G) = V^*_1 \int_{\tilde{x}_2}^{x_2} \left[ \frac{r_2(y, x_1, v, x_3, \ldots, x_G)}{r_1(y, x^*)} - \int_{\tilde{x}_1}^{x_1} \frac{\partial r_1(y, s, v, x_3, \ldots, x_G)}{\partial x_2} ds \right] dv
\]
\[
(17) \quad + V(\tilde{x}_1, \tilde{x}_2, x_3, \ldots, x_G).
\]
Combining (17) with (16) shows that $V(x)$ is now identified up to an unknown additive function of $(x_3, \ldots, x_G)$. Repeating the same reasoning for all the remaining components of $x$ shows that $V(x)$ is identified up to an unknown additive function of $z$ alone. The latter is then pinned down using the normalization condition $V(\bar{x}) = \bar{v}$.

It remains to establish that $\varepsilon$ is identified. For this, reintroduce the dependence on $Z$ and re-write (14) as $\Psi(Y, Z) = V(X, Z) + \varepsilon$. Recall that $V(x, z)$ is identified, and that $\Psi(y, z)$ has been identified up to an additive unknown function of $z$, $\Psi(\tilde{y}, z)$ in equation (15). Thus $\varepsilon$ is the sum of some known function of $(Y, X, Z)$ and an unknown function of $Z$. The normalization $E[\varepsilon|Z = z] = 0$ pins down $\Psi(\tilde{y}, z)$ and
thus establishes the identification of $\Psi$. Given that $\Psi$ and $V$ are now known, $\varepsilon$ can be treated as observed and so the distribution of $\varepsilon \mid Z$ is identified.

5. Conclusion

Just as in the case of linear simultaneous equations systems, see, e.g., Fisher (1966), restrictions on the covariance structure of the latent disturbances can be used to show identification in nonlinear systems. Those conditions replace the well-known exclusion restrictions.

MORE ABOUT HOW THIS IS CONSTRUCTIVE AND CAN LEAD TO ESTIMATORS?

References


Appendix A. Modified Proof of Step 2

We here modify the decoupling proof to avoid having to impose Assumption A6, and only assume that $\Lambda(\cdot, z)$ is invertible with inverse mapping $\Theta(\cdot, z)$ (Assumption A3). Recall that Assumption A6 is used to establish that equation (8) holds over all $y \in \mathbb{R}^G$. Instead we here split up $\mathbb{R}^G$ into $G$ non-overlapping subsets where on each subset $\frac{\partial \Theta_g(y)}{\partial y_h} \neq 0$. We then repeat the reasoning from before on each of these subsets.

First note that due to Assumption A3, $\frac{\partial \Theta(y)}{\partial y}$ has full rank for all $y$. Choose any $1 \leq g \leq G$, which is kept fixed in the following. For every $1 \leq h \leq G$, define

$$
\mathcal{Y}_{g,h} \equiv \left\{ y \in \mathbb{R}^G : \frac{\partial \Theta_g(y)}{\partial y_h} \neq 0 \right\},
$$

set $\tilde{\mathcal{Y}}_{g,1} \equiv \mathcal{Y}_{h,1}$, and for every $2 \leq h \leq G$ define

$$
\tilde{\mathcal{Y}}_{g,h} = \mathcal{Y}_{g,h} \setminus \bigcup_{i=1}^{h-1} \tilde{\mathcal{Y}}_{g,i}.
$$

Due to the full rank of $\frac{\partial \Theta(y)}{\partial y}$, then for any $y \in \mathbb{R}^G$, there exists at least one component $h$ such that $\frac{\partial \Theta_g(y)}{\partial y_h} \neq 0$. Thus, $\mathbb{R}^G = \bigcup_{h=1}^{G} \tilde{\mathcal{Y}}_{g,h}$, where $\tilde{\mathcal{Y}}_{g,h} \cap \tilde{\mathcal{Y}}_{g,i} = \emptyset$ for $h \neq i$.

We now show that there exists functions $\Psi_g$ and $\Theta_g$ so that (5) holds for $y \in \tilde{\mathcal{Y}}_{g,h}$, $h = 1, \ldots, G$. For a given choice of $h$, either $\tilde{\mathcal{Y}}_{g,h} = \emptyset$, in which case we skip it, or $\frac{\partial \Theta_g(y)}{\partial y_h} \neq 0$ for $y \in \tilde{\mathcal{Y}}_{g,h}$. Thus, taking ratios of (7) is then justified and for every $1 \leq i \leq G$ we have:

$$
\frac{\partial \Theta_g(y)/\partial y_i}{\partial \Theta_g(y)/\partial y_h} = \frac{(\partial^2 \ln f_{Y|X}(y|x^*)/\partial y_i \partial x) (V^{*-1})_g}{(\partial^2 \ln f_{Y|X}(y|x^*)/\partial y_h \partial x) (V^{*-1})_g},
$$

(18)
where the right-hand side is observed. We now show that the system of partial differential equations (8)-(9) determines \( \Theta_g(\cdot) \) up to a strictly increasing transformation that depends on \( z \). For this, let \( \bar{\Theta}_{g,h}(\cdot) \) denote any particular solution to (8)-(9) on \( \bar{Y}_{g,h} \). Then, for \( y \in \bar{Y}_{g,h} \), \( \partial \bar{\Theta}_{g,h}(y) / (\partial y_h) \neq 0 \) and

\[
\frac{\partial \Theta_g(y)/\partial y_i}{\partial \Theta_g(y)/\partial y_h} = \frac{\partial \bar{\Theta}_{g,h}(y)/\partial y_i}{\partial \bar{\Theta}_{g,h}(y)/\partial y_h}.
\]

Now, consider a change of variable \( \Gamma_h(\cdot) : (y_1, \ldots, y_G) \mapsto (y_1, \ldots, y_{h-1}, \Theta_g(y), y_{h+1}, \ldots, y_G) \). Note that \( \Gamma_h(\cdot) \) is continuously differentiable on \( \bar{Y}_{g,h} \), \( \det (\partial \Gamma_h(y)/\partial y) = \partial \bar{\Theta}_{g,h}(y)/\partial y_h \neq 0 \), and \( \Gamma(\cdot) \) is one-to-one on \( \bar{Y}_{g,h} \). Performing the change of variable \( \Gamma(\cdot) \) we can write

\[
\Theta_g(y_1, \ldots, y_G) = \Theta_g(\Gamma^{-1}_h(\Gamma_h(y))) = \Psi_{g,h}(\Gamma_h(y)),
\]

where the function \( \Psi_{g,h}(\cdot) = \Theta_g(\Gamma^{-1}_h(\cdot)) \) is continuously differentiable on \( \bar{Y}_{g,h} \). Differentiating on both sides of the last equation wrt \( y_k \) (\( k \neq h \)) and \( y_h \) we obtain

\[
\begin{align*}
\frac{\partial \Theta_g(y)/\partial y_k}{\partial \Theta_g(y)/\partial y_h} &= \frac{\partial \Psi_{g,h}(y_h, \Theta_{g,h})}{\partial y_k} + \frac{\partial \Psi_{g,h}(y_h, \Theta_{g,h})}{\partial \Theta_g} \frac{\partial \bar{\Theta}_{g,h}(y)/\partial y_k}{\partial \Theta_{g,h}(y)/\partial y_h} \\
\frac{\partial \Theta_g(y)/\partial y_h}{\partial \Theta_g(y)/\partial y_h} &= \frac{\partial \Psi_{g,h}(y_h, \Theta_{g,h})}{\partial y_h} \frac{\partial \bar{\Theta}_{g,h}(y)/\partial y_h}{\partial \Theta_{g,h}(y)/\partial y_h},
\end{align*}
\]

where \( \partial \Psi_g/(\partial \Theta_g) \) denotes the partial derivative of \( \Psi_g \) with respect to its \( h \)th argument, and we use \( \Psi_g(y_{-h}, \Theta_g) \) to denote \( \Psi_g(y_1, \ldots, y_{h-1}, \Theta_g, y_{h+1}, \ldots, y_G) \). It follows from (8) and (12) that \( \partial \Psi_g(y_{-h}, \Theta_g)/\partial \bar{\Theta}_g \neq 0 \) almost everywhere on \( \bar{Y}_{g,h} \). Moreover, taking the ratios of (11) and (12) we obtain

\[
\frac{\partial \Theta_g(y)/\partial y_k}{\partial \Theta_g(y)/\partial y_h} = \frac{\partial \bar{\Theta}_{g,h}(y)/\partial y_k}{\partial \bar{\Theta}_{g,h}(y)/\partial y_h} + \frac{\partial \Psi_{g,h}(y_{-g}, \Theta_g)/\partial y_k}{\partial \Psi_{g,h}(y_{-g}, \Theta_g)/\partial \Theta_g} \frac{1}{\partial \Theta_{g,h}(y)/\partial y_h}.
\]

Combining the above with (10) then gives \( \partial \Psi_{g,h}(y_{-g}, \bar{\Theta}_g)/\partial y_k = 0 \) almost everywhere on \( \bar{Y}_{g,h} \), for all \( k \neq h \). Thus, for a known function \( \bar{\Theta}_{g,h}(\cdot) \), we have

\[
\Theta_g(y) = \Psi_{g,h}(\bar{\Theta}_{g,h}(y)), \quad y \in \bar{Y}_{g,h},
\]

with \( \partial \Psi_{g,h}(u)/\partial u \neq 0 \) on \( \bar{Y}_{g,h} \). Defining \( \Psi_g(u) := \Psi_{g,h}(u) \) and \( \bar{\Theta}_g(y) = \bar{\Theta}_{g,h}(y) \) for \( y \in \bar{Y}_{g,h}, \ h = 1, \ldots, G \), equation (5) has been shown to hold.
**PROBLEM:** We know that $u \mapsto \Psi_{g,h}(u)$ is monotone on $\bar{Y}_{g,h}$, $h = 1, \ldots, G$, but this does not imply that $\Psi_{g}(u)$ as defined above is monotone.

**Appendix B. Alternative Proof of Step 2**

We wish to identify a function $\bar{\Theta}(y)$ so that $\bar{Y} = \bar{\Theta}(Y)$ satisfies

$$\Psi_{g}(\bar{Y}_{g}) = V(x) + \varepsilon,$$

for some strictly monotonic function $\Psi_{g}, g = 1, \ldots, G$. First note that any function $\bar{\Theta}(y) = (\bar{\Theta}_{1}(y), \ldots, \bar{\Theta}_{G}(y))$ on the form

$$\bar{\Theta}_{g}(y) = \Psi_{g}(\Theta_{g}(y))$$

will satisfy this restriction which in turn implies that the following relationship must hold,

$$\bar{\Theta}'(y) = \Theta'(y) D(y), \quad D(y) \equiv \text{diag}(\Psi_{1}'(\Theta_{1}(y)), \ldots, \Psi_{G}'(\Theta_{G}(y))).$$

Now observe that (see Section 3.1)

$$\Theta'(y) D(y) = -\left( \frac{\partial^{2} \ln f_{Y|X}(y,x^{*})}{\partial y \partial x} \right) V^{*-1}.$$ 

Thus, choosing

$$\bar{\Theta}'(y) = -\left( \frac{\partial^{2} \ln f_{Y|X}(y,x^{*})}{\partial y \partial x} \right) V^{*-1}$$

we have the desired result (assuming that we know $\Theta(y_{0})$ for some $y_{0}$).