NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nikolay Arefiev

STRUCTURAL MODELS WITH TESTABLE IDENTIFICATION

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 79/EC/2014

STRUCTURAL MODELS WITH TESTABLE IDENTIFICATION

NIKOLAY AREFIEV

Abstract. For linear Gaussian simultaneous equations models with orthogonal structural shocks, I show that, if appropriate instruments are available, there exists a set of inclusion and exclusion restrictions sufficient for the full identification, such that each identification restriction from this set is testable. This result does not depend on the assumption whether the model is recursive or cyclical, although the causal representation of cyclical models is not unique. To prove this, I provide a reduced form rank condition for the identification of simultaneous equations models, propose a graphical interpretation of the rank condition, provide graphical interpretations of various sufficient conditions for identification of structural vector autoregressions, and formulate new conditional independence tests.

Keywords: Identification, instrumental variables, data-oriented identification, sparse structural models, structural vector autoregression, SVAR, simultaneous equations model, SEM, probabilistic graphical model, PGM.

JEL codes: C30, E31, E52.

1. Introduction

A common problem in econometrics is to measure the causal effects and structural shocks that have produced observed covariances or more general comovements in a given dataset. In applications where controlled experiments are too expensive or not possible, this problem is usually solved using identification assumptions, which presume the existence of some causal relationships in the true data-generating model and an absence of others. To make such assumptions, however, a strong theoretical argument is required, which would explain why the presumably excluded causal effects cannot be present in the true model, but this theory may be currently unavailable. The question of this paper is, therefore, whether it is possible to measure some causal effects using only testable identification restrictions.

---

*November 27, 2014.

National Research University Higher School of Economics. E-mail: n.arefiev@gmail.com. Address: office 708A, Shabolovka 31, 119049, Moscow, Russia. Cell phone: +7 926 825 7147.

**This research was inspired by online course at Coursera "Probabilistic Graphical Models" lectured by Daphne Koller, see Koller (2009). I thank Alina Arefeva, Antoine d'Autume, Svetlana Bryzgalova, Jean-Bernard Chatelain, Jean-Pierre Drugeon, Boris Demishev, Bulat Gafarov, James Hamilton, Oleg Itskhoki, Maarten Janssen, Ramis Khabibulin, Sergey Kusnetzov, Jessie Li, Judea Pearl, Christopher Sims, Alain Trognon, Ilya Voskoboynikov, and Mark Watson for helpful comments and suggestions. I also thank participants of groups 7inR and Macroteam, and other colleagues from University Paris-1 Panthéon-Sorbonne and from The Higher School of Economics for fruitful discussions. This study (research grant No 14-01-0088) was supported by The National Research University Higher School of Economics' Academic Fund Program in 2014/2015.

For Gaussian simultaneous equations models (SEMs) with orthogonal structural shocks, I show that if appropriate instruments are available, then testable identification is possible. Namely, given appropriate instruments, there exists a set of inclusion and exclusion restrictions, satisfying the following properties. On the one hand, each restriction from this set can be tested either as the null or as the alternative hypothesis. On the other hand, taken together, these restrictions suffice for the full identification of the structural model. This result does not depend on the assumption whether the model is recursive or cyclical, although cyclical models have multiple SEM and causal representations, and require stronger instruments for testable identification.

To demonstrate the existence of testable identification, I derive the following results. First, I propose graphical interpretations of rank condition, of the sufficient condition for identification of structural vector autoregression (SVAR) models (Rubio-Ramírez et al. (2010)), and of the theory of partial identification (Christiano et al. (1999)). An example demonstrating the graphical interpretations of these conditions can be found in Section 2.1, formal propositions are presented in Section 3 and proven in Appendices A and B. Second, I formulate and provide a reduced form rank condition for the identification of simultaneous equations models. An example demonstrating why the reduced form rank condition is helpful for testable identification is presented in Section 2.1, a formal proposition is formulated in Section 4.2 and proven in Appendix C. Finally, I show that if each structural equation includes a variable, which is present in this equation and absent in the others, then in almost all parameter points, there exists an empirical procedure, which asymptotically exactly identifies the set of variables entering into each structural equation. The procedure uses results from Section 4.1, and is presented in Section 5. The power of the proposed tests I discuss in Section 6.

This paper stems from the literature on probabilistic graphical models (PGM) (reviewed in Koller (2009); Pearl (2009)). Chen and Pearl (2014) provide a review of many identification criteria for intricate causal models. Most of these criteria, however, deal only with recursive models. The graphical interpretations of the various conditions for identifying the simultaneous equations models and structural vector autoregressions (SVARs) provided in this paper, are a powerful new tool for identifying cyclical models.

The idea of using PGM for testable identification of SVARs is not new (Kwon and Bessler (2011); Bryant and Bessler (2011); Hoover (2005); Oxley et al. (2009); Reale and Wilson (2001); Wilson and Reale (2008)). This literature usually only considers the PGM, where the influence of the predetermined variables has been concentrated out of the covariance matrix for the contemporaneous variables. This approach may help to identify the model, but it can never achieve testable identification without non-testable identification

restrictions. Instead, I only concentrate out the process that has generated the values of the predetermined variables, but not the predetermined variables themselves. The advantage of my approach is that it may suffice for the full identification of the structural model, even without any non-testable assumptions.

There are various alternative approaches proposed in the literature for testable identification (Klein and Vella (2010); Li and Müller (2009); Lowbel (2012); Magnusson and Mavroeidis (2014); Rossi (2005); Rigobon (2013)), although these approaches never suffice for testable identification without non-testable assumptions. My paper complements this literature, and offers a fresh way of approaching identification tools.

2. Method of Testable Identification in Two Examples

Before unpacking the formal theorems and proofs, I start with two examples that demonstrate how the method of testable identification can be applied in practice. The first example deals with a recursive model, and the second shows how testable identification can be achieved in a cyclical model. Some definitions and propositions required for these examples are intuitively introduced in this section, and elaborated in later sections.

2.1. **Testable identification of a recursive model.**

Consider the following simultaneous equations model:

$$y_1 = c_1 + b_{11} z_1 + \varepsilon_1 \tag{1a}$$

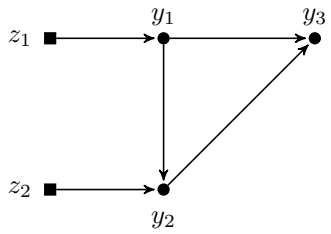
$$y_2 = c_2 + a_{21} y_1 + b_{22} z_2 + \varepsilon_2 \tag{1b}$$

$$y_3 = c_3 + a_{31} y_1 + a_{32} y_2 + \varepsilon_3 \tag{1c}$$

where y_1, y_2, and y_3 are endogenous variables, z_1 and z_2 are exogenous or predetermined variables, referred to further as instruments, ε_1, ε_2, and ε_3 are independent structural shocks, and a_{ij}, b_i, and c_i are parameters of the model.

It is well-known that recursive models with orthogonal structural shocks like model (1) are fully identified; a heuristic argument is that I can estimate equations in (1) one at a time, using, for example, the ordinary least squares regression, to achieve a consistent estimator of the parameters. The identification in this model is achieved using appropriate inclusion and exclusion restrictions. In (1a), for example, z_1 is included in the equation for y_1, and this is an inclusion restriction, but y_2 is excluded from the equation for y_1, and this is an exclusion restriction. The questions that I pose are the following. First, which of these restrictions are testable? For example, can I test the assumption that y_2 does not enter into the equation for y_1? Second,

a) Conditional causal diagram



b) Partial moral graph

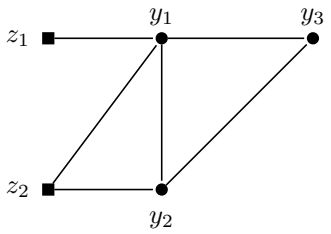


FIGURE 1. Conditional causal diagram and partial moral graph for model (1).

does the set of testable inclusion and exclusion restrictions suffice for the full or partial identification of the structural model? To answer these questions, I propose the following five-step procedure.

Step 1. Draw the conditional causal diagram. The *conditional causal diagram* is a directed graph, where the nodes are the random variables of the structural model, and where the edges are defined by the inclusion restrictions. The conditioning is made on Z, so the random process generating Z is not represented in the conditional causal diagram. A formal definition of the conditional causal diagram is provided in Section 3.

In model (1), I have five random variables, z_1, z_2, y_1, y_2, and y_3, so I have drawn five respective vertices, see Figure 1a. In (1a), z_1 is included into the equation for y_1, so in the causal diagram in Figure 1a, z_1 directly influences y_1. Using the language of graph theory, I can say equivalently that z_1 is a parent of y_1, and y_1 is a child of z_1. In (1b), y_1 and z_2 are included into the equation for y_2, so in the causal diagram y_1 and z_2 directly influence y_2. Finally, in (1c), y_1 and y_2 are included into the equation for y_3, so in the causal diagram y_1 and y_2 directly influence y_3.

Step 2. Draw the partial moral graph. To draw the *partial moral graph*, moralize and disorient the conditional causal diagram. "To moralize" means to marry all parents of each child. Node y_1 in the conditional causal diagram has just one parent, z_1, so moralization is not required. Node y_2 has two parents, y_1 and z_2, so I need to "marry" them, that is, connect them with an undirected edge. Node y_3 has parents y_1 and y_2, but they are already connected in the conditional causal diagram with edge $y_1 \to y_2$, so additional moralization is not required. Finally, disorient the graph, which means disregarding all directions. The resulting moral graph is depicted in Figure 1b. A formal definition of the partial moral graph is provided in Section 4.1.

Step 3. Draw the map of testable exclusion restrictions produced by the partial moral graph. The partial moral graph drawn in the previous step is useful for testable identification because of the following result, which is acknowledged in the literature on probabilistic graphical models to be true for

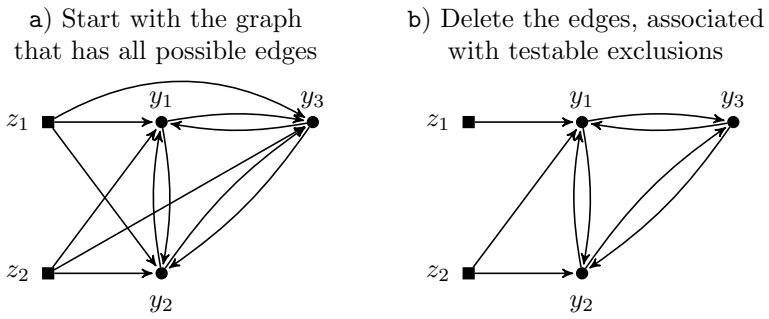


FIGURE 2. Drawing the map of testable exclusion restrictions produced by the partial moral graph.

full moral graphs (Koller (2009); Pearl (2009)). However, I show that this result also holds for some pairs of nodes in the partial moral graph, see Section 4.1. Consider a pair of endogenous variables (y_i, y_j), or one endogenous variable and one instrument, (y_i, z_j). In almost all parameter points, these variables are associated with adjacent vertices in the partial moral graph if and only if the partial correlation between them with conditioning on all the other variables of the structural model is not zero.

Using this result, I can draw the map of testable exclusion restrictions produced by the partial moral graph in the following way. Begin with the directed graph that has all possible edges, see Figure 2a. Within the framework of this paper, I assume that the instruments are known to be exogenous, and so this assumption is not tested. For this reason, there are no edges in Figure 2a directed from endogenous variables to instruments. Observe that in the partial moral graph in Figure 1b there is no edge z_1y_2, so the partial correlation between z_1 and y_2 with conditioning on z_2, y_1, and y_3 is zero. If edge z_1y_2 were to be present in the conditional causal diagram, this edge would also be present in the partial moral graph, and the partial correlation would not be zero. Therefore, I have a testable exclusion restriction, the restriction that z_1 does not enter into the structural equation for y_2, which is associated with the testable property of the joint probability distribution function that $\text{corr}(z_1, y_2|z_2, y_1, y_3) = 0$; I can delete edge z_1y_2 from the map of testable exclusions.

Similarly, since edge z_1y_3 is absent in the partial moral graph, I have another testable restriction that z_1 does not enter into the equation for y_3, which is associated with the testable property of the joint distribution function that $\text{corr}(z_1, y_3|z_2, y_1, y_2) = 0$, so I can delete edge z_1y_3 from the map of testable exclusions. Finally, there is no edge z_2y_3 in the partial moral graph, so I have the third testable exclusion, and I can delete the respective edge from the map of exclusions. I therefore produced the map of exclusions depicted in Figure 2b.

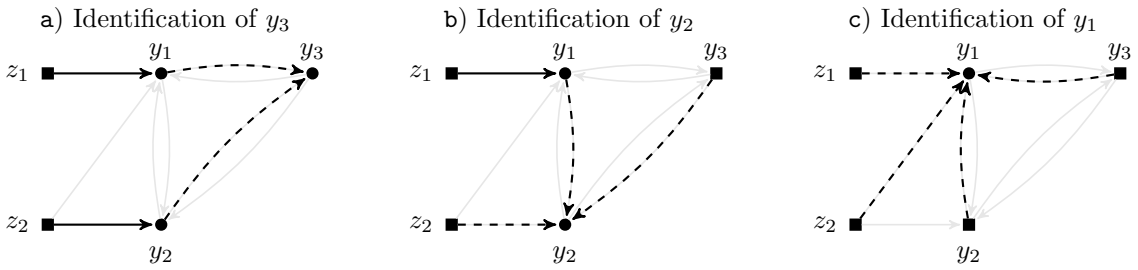


FIGURE 3. Graphical sufficient condition for identification

A map of exclusions is formally defined in Section 4.1. The exclusion restrictions, formulated in the way demonstrated in this example, are referred to hereafter as the *directly testable exclusions*.

Step 4. Verify whether the map of exclusions suffices for identification. A natural question is whether the map of exclusions depicted in Figure 2b suffices for identification. To answer this question, in Propositions 1 and 2 of Section 3 below, I propose graphical interpretations of various sufficient conditions for the identification, including the rank condition, the Rubio-Ramírez et al. (2010) sufficient condition, and the theory of partial identification. Given this, I prove that in almost all parameter points, a sufficient condition for the identification of all parameters in the structural equation for y_i is that each parent of y_i has an independent identifying path in the conditional causal diagram. An identifying path for a parent of y_i is a paths starting either with an instrument, or with any variable whose equation has been identified, or with any non-descendant of y_i, and that reaches the parent. The identifying paths for different parents must be independent, which means that they must not intersect on any node.

Using the above results, I verify whether the exclusion restrictions from Figure 2b suffice for the full identification. I begin with node y_3, see Figure 3a. This node has two parents, y_1 and y_2, so I need two independent identifying paths for the identification of the third structural equation. These paths do in fact exist, both in the map of testable exclusions and in the conditional causal diagram. Indeed, the identifying path for y_1 is z_1y_1, which by the definition of identifying path starts with instrument z_1 and reaches the parent. The identifying path for y_2 is z_2y_2, which starts with instrument z_2 and reaches the parent. These paths do not intersect on any node, so they are independent. Therefore, node y_3 is identified, which means that all parameters in equation (1c) are identified by the map of testable exclusions.

Now consider node y_2, see Figure 3b. The parents of y_2 are z_2, y_1, and y_3, so I need three independent identifying paths for the identification of the second equation. Node z_2 creates an identifying path of length 1 for itself, the path starts with z_2 in the role of instrument and it reaches z_2 in the role of parent. In the same manner, y_3 creates an identifying path for itself, the path starts with y_3 in the role of a node, which

has in the previous step been proven to be identified, and reaches y_3 in the role of parent of y_2. Finally, the identifying path for y_1 is z_1y_1, so node y_2 is also identified. In the same way, it is possible to show that y_1 is also identified, see Figure 3c. Therefore, the exclusion restrictions represented by the map of exclusions depicted in Figure 2b suffice for the full identification of the structural model.

Step 5. Test the required inclusion restrictions. Now I have a set of testable exclusion restrictions, which suffices for identification, but I have not tested whether the inclusion restrictions required for identification are satisfied. Indeed, the conclusion about identification depends on the assumption of the existence of edges z_1y_1 and z_2y_2 in the conditional causal diagram. If at least one of these edges is absent, there are no two independent identifying paths for the parents of y_3, and in this case no parameters in the structural model are identified. The map of testable exclusions that I use for identification, however, does not guarantee the presence of any edges. To achieve testable identification, therefore, I need to test the assumption that two independent paths connecting sets of nodes $\{z_1, z_2\}$ and $\{y_1, y_2\}$ exist.

To test the required inclusions, I propose the following procedure. Consider regressions of each variable from $\{y_1, y_2\}$ onto each instrument $\{z_1, z_2\}$:

$$y_1 = \pi_{10} + \pi_{11} \cdot z_1 + \pi_{12} \cdot z_2 + u_1 \tag{2a}$$

$$y_2 = \pi_{20} + \pi_{21} \cdot z_1 + \pi_{22} \cdot z_2 + u_2 \tag{2b}$$

Put the coefficients of these regressions into matrix $\mathbf{\Pi}(y_1, y_2|z_1, z_2)$:

$$\mathbf{\Pi}(y_1, y_2|z_1, z_2) = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \tag{3}$$

In Section 4.2, I prove that if the rank of $\mathbf{\Pi}(y_1, y_2|z_1, z_2)$ is two, then two independent paths connecting sets $\{z_1, z_2\}$ and $\{y_1, y_2\}$ exist. Using this result, I can test the inclusion restrictions, which are required for identification.

To test the rank of Π, I can use the following variation of the Johansen (1991) rank test. First, I estimate regressions of $\{y_1, y_2\}$ against $\{z_1, z_2\}$, and put the estimated coefficients into matrix $\mathbf{\Pi}(y_1, y_2|z_1, z_2)$. Second, I estimate regressions of $\{z_1, z_2\}$ against $\{y_1, y_2\}$, and put the estimated coefficients into matrix $\mathbf{\Pi}(z_1, z_2|y_1, y_2)$. Third, I calculate the degree of freedom (df), which is equal to the number of columns minus the number of rows of $\mathbf{\Pi}(y_1, y_2|z_1, z_2)$ plus 1, which in the considered example is 1. Finally, I calculate the df smallest eigenvalues $\lambda_1, \ldots, \lambda_{\text{df}}$ of product $\mathbf{\Pi}(y_1, y_2|z_1, z_2) \times \mathbf{\Pi}(z_1, z_2|y_1, y_2)$ and calculate the

TABLE 1. Testable identification restrictions for model (1)

Identification restriction	Testable property of PDF
$z_1 \not\to y_2$, so $b_{21} = 0$	$\text{corr}(z_1, y_2 \mid z_2, y_1, y_3) = 0$
$z_1 \not\to y_3$, so $b_{31} = 0$	$\text{corr}(z_1, y_3 \mid z_2, y_1, y_2) = 0$
$z_2 \not\to y_3$, so $b_{32} = 0$	$\text{corr}(z_2, y_3 \mid z_1, y_1, y_2) = 0$
There are 2 independent paths connecting $\{z_1, z_2\}$ with $\{y_1, y_2\}$	$\text{rank}\left(\mathbf{\Pi}\left(y_1, y_2 \mid z_1, z_2\right)\right) = 2$

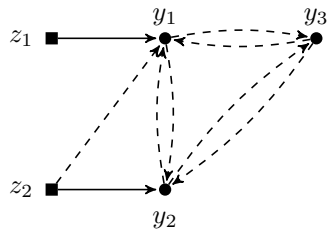


FIGURE 4. Summary of testable inclusion and exclusion restrictions.

statistic:

$$s = T \sum_{j=1}^{\text{df}} \ln\left(1 - \lambda_j\right) \tag{4}$$

where T is the number of observations. Under the null hypothesis that $\text{rank}\left(\mathbf{\Pi}\left(y_1, y_2 \mid z_1, z_2\right)\right) < 2$, the statistic is asymptotically distributed as $\chi^2(\text{df})$.

Table 1 and Figure 4 summarize the testable inclusion and exclusion restrictions sufficient for the full identification of the structural model. Each absent edge in Figure 4 is associated with a testable exclusion restriction, each solid edge is associated with a testable inclusion restriction, and the existence of the dashed edges is not important for identification, since the model is fully identified whether or not these edges are present in the causal diagram.

Now compare the true structural model:

$$\begin{pmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & -a_{32} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \tag{5}$$

with the estimated model:

$$\begin{pmatrix} 1 & -a_{12} & -a_{13} \\ -a_{21} & 1 & -a_{23} \\ -a_{31} & -a_{32} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ 0 & b_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \tag{6}$$

a) First representation b) Second representation c) Partial moral graph for both representations

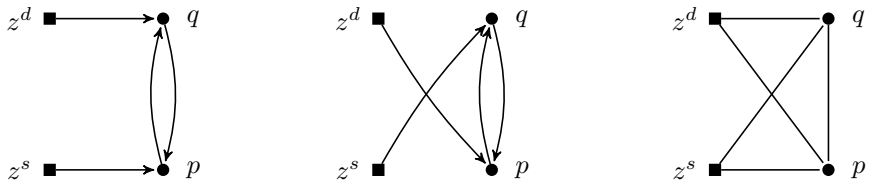


FIGURE 5. Equivalent causal representations of cyclical model (7) and the partial moral graph.

The estimated model is more complicated than the true model. In particular, the true model is triangular, whilst the estimated model is cyclical. However, the advantage of the estimated model is that it can be identified using only testable identification restrictions.

2.2. **Testable identification of a cyclical model.** Consider now an example, which demonstrates how a testable identification can be achieved for cyclical models. One difficulty with cyclical models is that the same model has several SEM and causal representations. In some cases, one representation is intuitively clearer than others, but generally, even intuition may not help to choose the only "correct" model.

Consider a market, where the demand and supply curves are given by the following equations:

$$q + \alpha p = c_1 + \gamma z^d + \varepsilon^d \tag{7a}$$

$$q - \beta p = c_2 + \delta z^s + \varepsilon^s \tag{7b}$$

where q is the log quantity of sales, p is the log price, $(-\alpha)$ and β are respectively the demand and supply elasticities, z^d is a determinant for the demand, z^s is a determinant for the supply, ε^d and ε^s are independent structural shocks, α, β, γ, δ, c_1 and c_2 are the estimated parameters.

There are two SEM representations of model (7), in which the demand and supply equations are identified: the first is where q is derived from (7a) and p from (7b), and the second is where p is obtained from (7a) and q from (7a). The conditional causal diagrams associated with these representations are depicted in parts a and b of Figure 5. This is not possible to justify that one representation is better than the other, neither from the economic theory, nor from any empirical tests. The moral graph associated with model (7) is the same for both representations (see Figure 5c).

In Section 4.1, I show how to test the null hypothesis that two given instruments do not have any common children within the endogenous variables in the conditional causal diagram. Using this test, I can distinguish the models in Figures 5a and 5b from the other observationally equivalent models, but I cannot distinguish

between them. This, however, suffices for the full identification of the model as written in (7). The required exclusion restrictions can be tested in the same way as in the previous example.

In contrast to the directly testable exclusions considered in the example in Section 2.1, the exclusion considered in this section is an *indirectly testable exclusion*. By definition, indirectly testable exclusions are associated with the test of the null hypothesis that there is no moralization effect between instruments. As I discuss in Section 6, indirect tests require stronger instruments than direct tests.

3. A Graphical Method of Identification

In this section, I formulate and prove various sufficient conditions for identification, which I have already applied in the previous section. Consider the following simultaneous equations model (SEM):

$$\mathbf{A}Y = \mathbf{B}Z + \mathcal{E} \tag{8}$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices of parameters, Y is an $n \times 1$ vector of the centralized endogenous variables, Z is an $m \times 1$ vector of the centralized exogenous or predetermined variables, and $\mathcal{E}$ is an $n \times 1$ vector of the unobservable Gaussian disturbances uncorrelated with Z, $\mathcal{E} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Most of the paper assumes that the structural shocks are independent, so the covariance matrix $\mathbf{\Sigma}$ is diagonal. This assumption, however, is not used in Propositions 1 and 4 below, where $\mathbf{\Sigma}$ is assumed to be a symmetric positive definite matrix without any identifying assumptions imposed. The constant term is omitted in (8) because all variables have been centralized, so the term is zero. Matrix $\mathbf{A}$ is nonsingular, and the matrices of parameters $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Sigma}$ are normalized so that for each $i = 1, 2, \ldots, n : a_{i,i} > 0$ and $\sigma_{ii} = 1$, where $a_{i,i}$ and σ_{ii} are the respective elements of $\mathbf{A}$ and $\mathbf{\Sigma}$. The variables of vector Z are referred to hereafter as the *primary instruments*. Primary instruments may be correlated with each other, but they are all independent of $\mathcal{E}$. I assume that there are enough observations and that there is a sufficient variance of Z to estimate the conditional probability distribution function $f(Y|Z)$ generated by (8).

Assume that Z is generated using a process $\mathbf{S}Z = \mathcal{E}_Z$ such that $\mathbf{S}$ is not singular and $\mathbb{E}(\mathcal{E}_Z \mathcal{E}_Z^T) = \mathbf{I}$. Then the whole model can be written as:

$$\mathbf{P}X = \mathcal{E}_X, \tag{9}$$

where

$$\mathbf{P} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \qquad X = \begin{pmatrix} Y \\ Z \end{pmatrix} \qquad \mathcal{E}_X = \begin{pmatrix} \mathcal{E} \\ \mathcal{E}_Z \end{pmatrix}$$

If no identification constraints are imposed on (8), then this model is not identified, which means that many different parameter points ($\mathbf{A}$ $\mathbf{B}$) exist, producing the same conditional probability distribution function $f(Y|Z)$ (see Appendices A.2 and B.1 for a brief review). To identify the model, I consider only those identification constraints, which restrict particular parameters to zero. All identification constraints are summarized by the conditional causal diagram, which was intuitively introduced in Section 2.1, and whose formal definition is:

Definition 1 (Conditional and unconditional causal diagrams)**.** A causal diagram is a directed graph, where the nodes are the random variables of the structural model, and where the edges are defined by the inclusion restrictions: edge $x_i \to x_j$ is present in the causal diagram if and only if $p_{ji} \neq 0$, where p_{ji} is the respective element of $\mathbf{P}$.

- The conditional causal diagram represents only the edges associated with matrices $\mathbf{A}$ and $\mathbf{B}$;
- The unconditional causal diagram represents edges associated with all entries of $\mathbf{P}$.

If edge $y_j \to y_i$ exists in the conditional causal diagram, then y_j is said to be a parent of y_i, and y_i is a child of y_j. If there is path $y_{j_1} \to y_{j_2} \to \cdots \to y_{j_N}$, then y_{j_i} is ancestor of y_{j_k} if $i < k$, and y_{j_i} is descendant of y_{j_k} if $i > k$. Two paths are independent if they do not intersect on any node. Each node is interpreted as a path of length 1.

Definition 2 (Primary identifying path)**.** A path in the conditional causal diagram is a primary identifying path for a parent y_j of node y_i if it starts with a primary instrument and reaches y_j.

Definition 3 (Identified node)**.** Node y_i said to be identified by the conditional causal diagram if all parameters in the i^{th} lines of $\mathbf{A}$ and $\mathbf{B}$ are identified.

In empirical studies, where the structural shocks may be not independent and no constraints are imposed on $\mathbf{\Sigma}$, the identification of a given parameter is usually verified in the literature using the *rank condition*, which is briefly reviewed in Appendix A.2. In this section, I propose the following graphical interpretation of this condition:

Proposition 1 (Graphical interpretation of rank condition)**.** *Assume that $\mathbf{\Sigma}$ is a symmetric positive definite matrix, and no identification constraints are imposed $\mathbf{\Sigma}$.*

- *If node y_i is identified in a given parameter point by the constraints, which are summarized by the conditional causal diagram, then for each parent of y_i there exists an independent primary identifying path in the conditional causal diagram.*

- *If for each parent of y_i there exists an independent primary identifying path in the conditional causal diagram, then node y_i is identified in almost all parameter points by the constraints, summarized by the conditional causal diagram.*

Proof. See Appendix A. □

Most of this paper concerns models with orthogonal structural shocks, in which case the rank condition is only a sufficient, but not a necessary condition for identification. Consider again the example depicted in Figure 3. The rank condition suffices for the identification of y_3, but it is not sufficient for the identification of y_1 or y_2. Indeed, for each parent of y_3 there is an independent primary identifying path, which starts with a primary instrument and reaches the parent (see Figure 3a), so y_3 is identified. The rank condition, however, does not suffice for the identification of y_2, because y_2 has 3 parents, but only two primary instruments are available; since it is not possible to draw three independent paths starting with two nodes, the rank condition is not satisfied for y_2. Nor is the rank condition satisfied for y_1.

Assume now that the structural shocks are orthogonal, so $\mathbf{\Sigma}$ is diagonal. When the independence assumption is made, some endogenous variables may possess the same properties as the primary instruments, so they can produce additional identifying paths and identify additional parameters. I introduce two kinds of instruments, recursive instruments and respective instruments. A *Recursive instrument* is defined as any endogenous node, which has been identified using other instruments. Node y_j is said to be a *respective instrument* for y_i if y_j is not a descendant of y_i.

Definition 4 (Recursive identifying path)**.** In a model with orthogonal structural shocks, a path in the conditional causal diagram is a recursive identifying path for a parent y_j of node y_i if it starts with an identified node and reaches y_j.

Definition 5 (Respective identifying path)**.** In a model with orthogonal structural shocks, a path in the conditional causal diagram is a respective identifying path for a parent y_j of node y_i if it starts with a non-descendant of y_i and reaches y_j.

Proposition 2 below uses Rubio-Ramírez et al.'s (2010) sufficient condition for identification, to prove that recursive instruments can be used for the identification of structural models in the same manner as primary instruments. To prove the sufficiency of respective instruments in the same proposition, I use the theory of partial identification, as reviewed in Christiano et al. (1999).

Proposition 2 (Recursive condition for identification)**.** *Assume that the structural shocks are independent, so* $\mathbf{\Sigma}$ *is a positive diagonal matrix. If for each parent of* y_i *in the conditional causal diagram there is an independent primary, recursive or respective identifying path, then* y_i *is globally identified by the causal diagram in almost all parameter points.*

Proof. See Appendix B. □

Comparing the recursive condition for identification, as formulated in Proposition 2, with the rank condition formulated in Proposition 1, I note that the recursive condition, on the one hand, requires a shock independence assumption, but on the other hand, permits the use of recursive and respective instruments in addition to the primary instruments permitted by Proposition 1. An example of application of Proposition 2 can be found in Section 2.1.

4. Testable Identification Restrictions

In this section, I provide definitions and propositions, which I have already applied in Section 2 to formulate testable exclusion and inclusion restrictions.

4.1. **Testable exclusions.** Consider concentration matrix $\mathbf{C}$, also know as the precision matrix, which is defined as the inverse covariance matrix of X: $\mathbf{C} = \left(\mathbb{E}\left(XX^T\right)\right)^{-1}$. Since each variable in Z is exogenous or predetermined, and the covariance matrices for $\mathcal{E}$ and $\mathcal{E}_Z$ are normalized to the identity matrices, I have: $\mathbb{E}(\mathcal{E}_X\mathcal{E}_X^T) = \mathbf{I}$. Observe that:

$$\begin{aligned}\mathbf{I} &= \mathbb{E}(\mathcal{E}_X\mathcal{E}_X^T) = \mathbb{E}(\mathbf{P}XX^T\mathbf{P}^T)\\ &= \mathbf{P}\mathbf{C}^{-1}\mathbf{P}^T,\end{aligned}$$

from which I get:

$$\begin{aligned}\mathbf{C} &= \mathbf{P}^T\mathbf{P}\\ &= \begin{pmatrix}\mathbf{A}^T\mathbf{A} & -\mathbf{A}^T\mathbf{B}\\ -\mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B}+\mathbf{S}^T\mathbf{S}\end{pmatrix}\end{aligned} \tag{10}$$

The concentration matrix, on the one hand, can be estimated from the data without any prior identification assumptions. On the other hand, it gives estimators for $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{B}$, which are helpful for finding testable exclusions. To make the right-bottom block of the concentration matrix more informative, the value of $S^TS = \left(\mathbb{E}\left(ZZ^T\right)\right)^{-1}$, which can be estimated separately, is subtracted from this block. I then

obtain matrix $\hat{\mathbf{C}}$ referred to hereafter as the *partial concentration matrix*:

$$\begin{aligned}\hat{\mathbf{C}} &= \mathbf{C} - \begin{pmatrix} \mathbf{0}_{n\times n} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{m\times n} & \left(\mathbb{E}\left(ZZ^T\right)\right)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}^T\mathbf{A} & -\mathbf{A}^T\mathbf{B} \\ -\mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B} \end{pmatrix} \end{aligned} \tag{11}$$

As I see from the obtained equation, in comparison to the full concentration matrix, the partial concentration matrix also gives an estimator for $\mathbf{B}^T\mathbf{B}$.

Definition 6 (Partial concentration network)**.** The partial concentration network is an undirected graph, which spans the random variables of the model, where x_i and x_j are adjacent if and only if $\hat{c}_{ji}$ is not zero, where $\hat{c}_{ij}$ is the respective element of $\hat{\mathbf{C}}$.

The partial concentration network is useful for testable identification, because it is closely related to the partial moral graph. Before formally defining the partial moral graph and showing this relationship, I define relatives and strangers:

Definition 7 (Relatives and strangers)**.** Vertices x_i and x_j are *relatives* in the conditional causal diagram if and only if at least one of the following conditions holds:

(1) x_i is a child of x_j, $x_i \leftarrow y_j$;

(2) x_i is a parent of x_j, $x_i \rightarrow y_j$;

(3) there is a vertex x_k such that x_k is a common child of x_i and x_j: $x_i \rightarrow x_k \leftarrow y_j$.

vertices x_i and x_j are *strangers* if and only if they are not relatives.

Using this definition, the partial moral graph can be redefined as follows:

Definition 8 (Partial moral graph)**.** A partial moral graph is an undirected graph, where the nodes are the random variables of the model, and where any two nodes are adjacent if and only if they are relatives in the conditional causal diagram.

Proposition 3 (Partial moral graph and partial concentration network)**.** *Assume that the structural shocks are independent, so matrix* $\mathbf{\Sigma}$ *is diagonal.*

- *If an edge is absent in the partial moral graph, this edge is also absent in the partial concentration network.*

- *In almost all parameter points, if an edge is absent in the partial concentration network, this edge is also absent in the partial moral graph.*

Before proving Proposition 3, I will introduce an assumption. Let $\hat{\mathbf{P}}$ be the matrix obtained from $\mathbf{P}$ by substituting the bottom m lines with zeros:

$$\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \underset{m\times n}{\mathbf{0}} & \underset{m\times m}{\mathbf{0}} \end{pmatrix}$$

.

Assumption 1 (Generic assumption for parameter point $\hat{\mathbf{P}}$)**.** *Parameter point $\hat{\mathbf{P}}$ satisfies **generic assumption** if for each i and j, $i \neq j$, the existence of k such that $[\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$ implies $\sum_k [\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$.*

The generic assumption excludes edge-of-the-knife cases, where different causal effects precisely offset each other in equilibrium. For example, in a model with two variables y_1 and y_2, where y_1 positively influences y_2 and y_2 negatively influences y_1, the generic assumption excludes the case where the parameters are such that y_1 and y_2 are entirely uncorrelated in equilibrium. Since the generic assumption is not satisfied only in the subspace of parameters with a lower number of degrees of freedom than the full space of parameters, it is satisfied in almost all parameter points.

Proof of Proposition 3. First, I prove that x_i and x_j are relatives if and only if there exists index k such that $[\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$. This result directly follows from the fact that x_i and x_j can be relatives only in one of the following cases:

- x_i is a child of x_j. In this case $[\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$ for $k = i$ (recall the normalization rule $a_{ii} > 0$).
- x_i is a parent of x_j, then $[\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$ for $k = j$.
- x_i and x_j have a child in common. Let this child be x_k. Then $[\hat{\mathbf{P}}]_{ki} \cdot [\hat{\mathbf{P}}]_{kj} \neq 0$.

Now observe that $\hat{c}_{ij} = \left[\hat{\mathbf{P}}^T\hat{\mathbf{P}}\right]_{ij} = \sum_k [\hat{\mathbf{P}}]_{ik}[\hat{\mathbf{P}}]_{jk}$. If x_i and x_j are strangers, then $\forall k = 1, 2, \ldots, n+m$: $[\hat{\mathbf{P}}]_{ik} \cdot [\hat{\mathbf{P}}]_{jk} = 0$, therefore $\hat{c}_{ij} = 0$. If x_i and x_j are relatives, then there exists k such that $[\hat{\mathbf{P}}]_{ik} \cdot [\hat{\mathbf{P}}]_{jk} \neq 0$, and through the generic assumption, which is satisfied in almost all parameter points, I obtain $\hat{c}_{ij} \neq 0$. □

The following two corollaries stem from Proposition 3. Consider two nodes in the partial moral graph; node x_i is associated with a variable in vector Y or Z, and node y_i is associated with a variable in Y.

Corollary 1 (Directly testable exclusion $x_i \not\to y_j$)**.** *If edge $x_i - y_j$ is absent in the partial moral graph, there is directly testable exclusion restriction that x_i does not enter into the structural equation for y_j,which*

is associated with the testable property of the probability distribution function that the respective element of the partial concentration matrix is zero.

Consider two primary instruments, z_i and z_j.

Corollary 2 (Indirectly testable exclusion restrictions)**.** *If edge $z_i - z_j$ is absent in the partial moral graph, there is indirectly testable exclusion that z_i and z_j do not have any common children among the endogenous variables, which is associated with the testable property of the probability distribution function that the respective element of the partial concentration matrix is zero.*

In Section 2.1 I use partial correlations instead of the partial concentration network to test the exclusions. This is possible because the full concentration matrix is closely related to the matrix of partial correlations. The partial correlation between x_i and x_j with conditioning on the other variables of the model $X_{(-i,-j)}$ is defined as the correlation between the residuals of the regressions of x_i and x_j on $X_{(-i,-j)}$. Knowing the matrix of concentration, the partial correlations can be calculated using the following formula:

$$\operatorname{corr}(x_i, x_j | X_{(-i,-j)}) = \frac{c_{ij}}{\sqrt{c_{ii} c_{jj}}}, \tag{12}$$

where c_{ij}, c_{ii}, and c_{jj} are the respective elements of matrix $\mathbf{C}$. Therefore, element c_{ij} of matrix $\mathbf{C}$ is zero if and only if x_i and x_j are partially uncorrelated, which in the Gaussian case is true if and only if x_i and x_j are conditionally independent with conditioning on $X_{(-i,-j)}$.

Partial correlations, however, cannot be directly used to test indirectly testable exclusions. In this case, I can test the hypothesis that the respective element of the partial concentration matrix is zero. In Section 2.2, for example, the hypothesis that z_1 and z_2 do not have common children among the endogenous variables is associated with a testable property of the probability distribution function that $\hat{c}_{34} = 0$.

Examples demonstrating how to formulate testable exclusion restrictions using the results presented in this section can be found in Section 2.

4.2. **Testing Inclusion Restrictions.** Let matrix $\mathbf{\Pi}_i(\mathcal{P}_i|Z)$ be defined by the following operator in the true model:

$$\mathbb{E}\left(\mathcal{P}_i|Z\right) = \mathbf{\Pi}_i(\mathcal{P}_i|Z)Z \tag{13}$$

where $\mathcal{P}_i$ are the parents of y_i on the map of exclusion restrictions. The constant term is omitted in (13), because all variables have been centralized, so the term is zero.

Proposition 4 (Reduced Rank Condition)**.** *Assume that* $\mathbf{\Sigma}$ *is a symmetric positive definite matrix, and no identification constraints are imposed on* $\mathbf{\Sigma}$*. Node* y_i *is identified in the given parameter point if and only if* $\mathbf{\Pi}_i(\mathcal{P}_i|Z)$ *has full row rank.*

Proof. See Appendix C. □

Consider an example, which demonstrates the intuition behind Proposition 4, and shows that if a node is not identified, then the condition formulated in Proposition 4 is not satisfied, so the row rank of $\mathbf{\Pi}_i(\mathcal{P}_i|Z)$ is not full. Consider the map of exclusion restrictions depicted in Figure 2b. Assume edge z_2y_2 is absent in the causal diagram, but the other edges depicted in Figure 2b are present in the causal diagram. By any Proposition, 1 or 2, y_3 is not identified. I shall intuitively demonstrate that the row rank of $\mathbf{\Pi}_3(\mathcal{P}_3|Z)$ is not full in this case. Matrix $\mathbf{\Pi}_3(\mathcal{P}_3|Z)$ includes the coefficients of the regressions of the parents of y_3 in the identification map, which are y_1 and y_2, on the instruments z_1 and z_2 (see equations (2) and (3)). Each row of $\mathbf{\Pi}_3(\mathcal{P}_3|Z)$ corresponds to a parent, and each column corresponds to an instrument. In the causal diagram, I observe that if edge z_2y_2 is absent, the expected values of y_1 and y_2 can be expressed as functions of y_1 alone, so the rows of matrix $\mathbf{\Pi}_3(\mathcal{P}_3|Z)$ are linearly dependent, in which case the row rank of $\mathbf{\Pi}_3(\mathcal{P}_3|Z)$ is not full, and the condition formulated in Proposition 4 is, in fact, not satisfied. Section 2.1 demonstrates how to apply this result to achieve testable identification and how to test the rank of $\mathbf{\Pi}_i(\mathcal{P}_i|Z)$.

5. Existence of testable identification

In this section, I prove the thesis that if appropriate instruments are available, then testable identification is possible. To prove this, I analyze *cliques* in the partial moral graph. A *clique* in an undirected graph is a set of nodes such that every two nodes are adjacent. In Figure 1b, sets of nodes $\{z_2, y_1, y_2\}$ and $\{z_2, y_1\}$ are examples of cliques, but set $\{z_2, y_1, y_2, y_3\}$, is not a clique, because z_2 and y_3 are not adjacent. Clique $\{z_2, y_1, y_2\}$ is a *maximal clique*, because it cannot be extended by including any other nodes. The *clique cover* problem is to find as few cliques as possible that includes all nodes of the graph. In Figure 1b, cliques $\{z_1, y_1\}$, $\{z_2, y_1, y_2\}$, and $\{y_1, y_2, y_3\}$ solve the clique cover problem.

An analysis of the cliques covering the partial moral graph is useful for testable identification because of the following property:

Proposition 5 (Structural equations and cliques)**.** *For each structural equation in the simultaneous equations model, there is a clique in the partial moral graph, such that all variables entering into the structural equation are in the clique.*

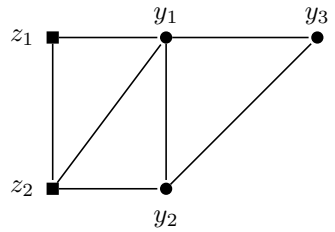


FIGURE 6. partial moral graph for model (14), where the solution to the clique cover problem is not unique, but testable identification is possible.

Proof. This result is evident from Definitions 7 and 8. □

Proposition 5 produces the following result:

Corollary 3. *If the clique cover problem has a unique solution with n cliques, there is a one-to-one association between the cliques and the structural equations, such that each clique consists of the variables present in the associated structural equation.*

Consider the example in Figure 1b. The clique cover problem has the following unique solution: $\{z_1, y_1\}$, $\{z_2, y_1, y_2\}$, and $\{y_1, y_2, y_3\}$. Therefore, this partial moral graph can be associated with the only structural model, where the first equation uses variables z_1 and y_1, the second uses z_2, y_1, and y_2, and the third uses y_1, y_2, and y_3. The partial moral graph does not indicate, however, which variables should be put on the left-hand side, and which on the right-hand side of each structural equation, but knowing that z_1 and z_2 are exogenous, the only model consistent with the partial moral graph in Figure 1b is (1). Similarly, the unique solution to the clique cover problem for the partial moral graph in Figure 5c is $\{z^d, p, q\}$, and $\{z^s, p, q\}$, which is consistent with the only structural model (7), but it does not indicate whether the representation in panel a or b of Figure 5 is correct.

The condition formulated in Corollary 3 is a sufficient, but not a necessary condition for the unique association between the variables and the structural equations using testable restrictions. Consider, for example, a modification of (1), where z_2 is also included in the structural equation for y_1:

$$y_1 = c_1 + b_{11}z_1 + b_{12}z_2 + \varepsilon_1 \tag{14a}$$

$$y_2 = c_2 + a_{21}y_1 + b_{22}z_2 + \varepsilon_2 \tag{14b}$$

$$y_3 = c_3 + a_{31}y_1 + a_{32}y_2 + \varepsilon_3 \tag{14c}$$

The partial moral graph for model (14) is drawn in Figure 6. The clique cover problem has two solutions. Both solutions include cliques $\{z_1, z_2, y_1\}$ and $\{y_1, y_2, y_3\}$, the first solution also includes $\{z_2, y_1, y_2\}$, and the

second includes instead $\{z_2, y_2\}$. Therefore, the partial moral graph does not suffice to conclude whether y_1 is present or not into the second structural equation. The structural model, nevertheless, can be identified using only testable restrictions. A heuristic argument is that the constraints implied by the moral graph suffice for the full identification, and once the structural model has been identified, the hypothesis of whether y_1 is included or not into y_2 can be tested.

Definition 9 (Marker)**.** An exogenous or endogenous variable is a *marker* for a given structural equation, if the variable is present in this and only this structural equation.

In model (1), for example, z_1, z_2, and y_3 are markers respectively for the first, second, and third structural equations. In (7), the markers are z^d for the demand equation, and z^s for the supply equation.

Proposition 6 (Deciphering the partial moral graph)**.** *Consider a simultaneous equations model, where each structural equation has a marker. The partial moral graph suffices to identify all markers, and to decide which variables are included into the structural equation associated with each marker.*

Proof. To demonstrate this result, I prove that the solution to the clique cover problem is unique and consists exactly of n cliques. Since two markers never appear in the same equation, they are not adjacent in the moral graph, and so any two markers never pertain to the same clique. This guarantees that the clique cover problem cannot be solved in less than n cliques. The solution has no more than n cliques, because the whole moral graph has been produced by n cliques, each associated with one structural equation, so it is possible to cover the partial moral graph with n cliques.

It remains to be proven that the solution is unique. Each marker is adjacent to each variable from the associated structural equation, and all these variables are adjacent to each other. This defines the clique associated with each marker in a unique way, so the solution to the problem is unique.

Then, from Corollary 3 to Proposition 5, it is possible to identify the markers, and to say which variables are included into the structural equation associated with any given marker. □

In the abstract and introduction I claim that, for Gaussian simultaneous equations models with orthogonal structural shocks, if appropriate instruments are available then there is a set of inclusion and exclusion restrictions sufficient for full identification, such each restriction from this set is testable. Given Proposition 6, this result is evident. Indeed, if for each endogenous variable there is an instrument, which influences the considered endogenous variable, but does not influence any other endogenous variable, then the true structure can be specified using only testable restrictions. From Proposition 2, these restrictions suffice for

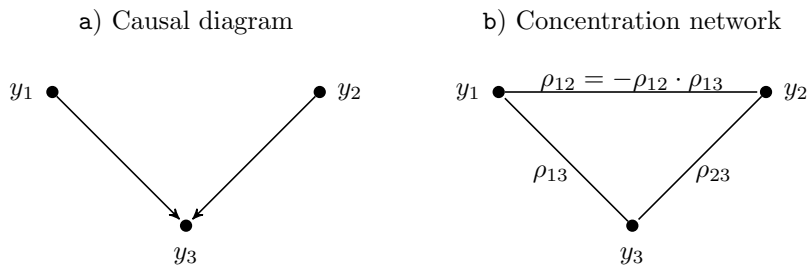


FIGURE 7. Partial correlations associated with moralization edges are typically smaller than partial correlations associated with true edges in the causal diagram. Notation: $\rho_{ij} \equiv \text{corr}(y_i, y_j | Y_{\{-i,-j\}})$.

full identification. The SEM representation may be not unique, as in the example in Section 2.2, nevertheless, it is possible identify all parameters in these equations.

6. Strategy of testable identification and test power

Although the existence of testable identification depends only on the shock orthogonality assumption and on the availability of appropriate instruments, indirect tests associated with exclusion restrictions require stronger instruments than direct tests, which should be taken into consideration when a strategy of testable identification is designed. There are two reasons why indirect tests are less powerful and require stronger instruments than direct tests. The first reason is that the partial correlation induced by the moralization of two causal effects is of order of the product of partial correlations, induced by edges in the causal diagram, which are associated with these causal effects. Consider, for example, Figure 7. In this example, the partial correlation associated with the moralization edge is equal to the minus product of the partial correlations induced by the true causal effects. Since each partial correlation in its absolute value is strictly less than one, the partial correlation associated with the moralization edge in absolute value in this example is smaller than each partial correlation associated with a true edge.

To see how strong the power of the test may be, I use Fisher's (1924) approximation of the distribution function for partial correlations. Fisher's z-transform of partial correlation coefficient ρ is defined as:

$$\zeta(\rho) = 0.5 \cdot (\ln(1+\rho) - \ln(1-\rho)), \tag{15}$$

and under the null hypothesis that $\rho = 0$, the value of $\zeta(\rho) \cdot \sqrt{T - (n+m) + 3}$ is approximately normally distributed with zero mean and standardized variance, where T is the number of observations.

In the example depicted in Figure 7, if $\rho_{13} = \rho_{23} = 0.2$ and I have 100 observations, for ρ_{13} and ρ_{23} I have $\zeta(0.2) \approx 0.202$, and $\zeta(0.2) \cdot \sqrt{T - (m+n) + 3} \approx 2.02$, so I expect to correctly reject the null hypothesis that

$\rho_{13} = 0$ or that $\rho_{23} = 0$ at the significance level of 10% approximately in 50% of experiments. The partial correlation associated with the moralization edge, however, is relatively small: $\rho_{12} = -0.2 \cdot 0.2 = -0.04$, and $\zeta(0.04) \cdot \sqrt{T - (m+n) + 3} \approx 0.4$, so in most experiments the null hypothesis is not rejected. Calculations suggest that to achieve a 50% rejection rate for the moralization edge, I must either increase the number of observations from 100 to approximately 2500, or if I keep the number of observations constant, then I need partial correlations, associated with the true causal edges, be greater or equal to approximately 0.45 in its absolute value.

The second reason why the indirect tests discussed in Section 2.2 are less powerful than the direct tests discussed in 2.1 is that in the indirect tests I subtract the estimated value of $\mathbb{E}(ZZ^T)^{-1}$ from the right-bottom block of the concentration matrix before testing that some elements of this matrix are zero (see equation (11)). Obviously, this procedure adds some random components to the estimated parameters, so it decreases the power of the test. Therefore, direct tests for exclusion restrictions in applications should be preferred over indirect tests.

References

Bryant, H. L. and D. A. Bessler (2011): "Proving causal relationships using observational data," Tech. rep.

Chen, B. and J. Pearl (2014): "Graphical Tools for Linear Structural Equation Modeling," *Psychometrica*, forthcoming.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What Extend?" in *Handbook of Macroeconomics, vol. 1A*, North Holland: Amsterdam.

Fisher, R. A. (1924): "The distribution of the partial correlation coefficient," *Metron*, 3-4, 329–332.

Greene, W. H. (2012): *Econometric Analysis*, Prentice Hall.

Hoover, K. D. (2005): "Automatic Inference of the Contemporaneous Causal Order of a System of Equations," *Econometric Theory*, 69–77.

Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Klein, R. and F. Vella (2010): "Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions," *Journal of Econometrics*, 154, 154–164.

Koller, D. (2009): *Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)*, The MIT Press.

Kwon, D.-H. and D. Bessler (2011): "Graphical Methods, Inductive Causal Inference, and Econometrics: A Literature Review," *Computational Economics*, 38, 85–106.

Li, H. and U. K. Müller (2009): "Valid Inference in Partially Unstable GMM Models," *Review of Economic Studies*, 76, 343.

Lowbel, A. (2012): "Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business and Economic Statistics*, 1, 67–80.

Magnusson, L. M. and S. Mavroeidis (2014): "Identification using Stability Restrictions," *Econometrica*, 82, 1799–1851.

Oxley, L., M. Reale, and G. Wilson (2009): "Constructing Structural VAR models with conditional independence graphs," *Mathematics and Computers in Simulation*, 79, 2910–2916.

Pearl, J. (2009): *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2nd ed.

Reale, M. and G. Wilson (2001): "Identification of vector AR models with recursive structural errors using conditional independence graphs," *Statistical Methods and Applications*, 10, 49–65.

Rigobon, R. (2013): "Identification Through Heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

Rossi, B. (2005): "Optimal Tests for Nested Model Selection With Underlying Parameter Instability," *Econometric Theory*, 5, 962–990.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77, 665–696.

Wilson, G. T. and M. Reale (2008): "The sampling properties of conditional independence graphs for I(1) structural VAR models," *Journal of Time Series Analysis*, 29, 802–810.

Appendix A. Proof of Proposition 1

A.1. **A Lemma.** Let $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$ be the set of the nodes of the causal diagram associated with the endogenous variables, $\mathcal{Z} = \{z_1, z_2, \ldots, z_m\}$ be the set of the nodes associated with the exogenous or predetermined variables, and $\mathcal{X} = \mathcal{Y} \cup \mathcal{Z}$ be the set of all nodes. Let $\mathcal{Y}_1$, $\mathcal{Y}_2$, and $\mathcal{X}_1$ be independent subsets of $\mathcal{X}$ satisfying: $\mathcal{Y}_1 \subset \mathcal{Y}$, $\mathcal{Y}_2 \subset \mathcal{Y}$, $\mathcal{X}_1 \subset X$, $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$, $\mathcal{Y}_1 \cap \mathcal{X}_1 = \emptyset$, and $\mathcal{Y}_2 \cap \mathcal{X}_1 = \emptyset$. Let $\mathbb{G}$ be the subgraph of the causal diagram induced by $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{X}_1$, and N be the number of independent paths in $\mathbb{G}$ starting with nodes in $\mathcal{X}_1$ and reaching nodes in $\mathcal{Y}_1$. Without loss of generality, I consider only paths without cycles. For example, if I given with a path $x_1x_2x_1x_4$, I consider instead the path, where cycle $x_1x_2x_1$ has been cut out, so I consider x_1x_4.

Let $\bar{\mathbf{P}}$ be the first n lines of matrix $\mathbf{P}$, so $\bar{\mathbf{P}} = (\mathbf{A} \quad -\mathbf{B})$.Consider matrix $\mathbf{M}$ obtained from $\bar{\mathbf{P}}$ in the following way. Take the rows of matrix $\bar{\mathbf{P}}$ having the indices of elements of $\mathcal{Y}_1 \cup \mathcal{Y}_2$, and take the columns of $\bar{\mathbf{P}}$ having the indices of $\mathcal{Y}_2 \cup \mathcal{X}_1$.

If there is a path $x_{j_1} x_{j_2} \dots x_{j_s}$ in the causal diagram, the set of parameters associated with this path consists of the following elements of matrix $\bar{\mathbf{P}}$: $\left\{p_{j_2 j_1}, p_{j_3 j_2}, \dots, p_{j_s j_{s-1}}\right\}$. Therefore, the diagonal elements of $\mathbf{A}$ are not considered as parameters associated with any path. By definition of the conditional causal diagram, the parameters associated with different paths are not constrained to zero by the identification restrictions.

In the proof of Proposition 1 below, I use Leibniz formula for determinant, which expresses the determinant as a sum over all permutations. Since matrix $\mathbf{M}$ may be not square, I consider partial permutations, which do not necessarily take all rows and all columns of $\mathbf{M}$. Let l be the length of the lengthiest partial permutation in $\mathbf{M}$ such that each element of the permutation is not restricted to zero by the identification constraints.

To make the lemma below clearer, consider the following example. Assume that the structural model is:

$$\begin{pmatrix} 1 & a_{12} & 0 & 0 & a_{15} & a_{16} \\ 0 & 1 & a_{23} & 0 & 0 & 0 \\ 0 & 0 & 1 & a_{34} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & a_{54} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b_{43} & 0 \\ 0 & 0 & 0 & b_{54} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}, \tag{16}$$

which causal diagram is depicted in Figure 8a. Consider the following sets of nodes: $\mathcal{Y}_1 = \{y_2, y_5, y_6\}$, $\mathcal{Y}_2 = \{y_3, y_4\}$, $\mathcal{X}_1 = \{z_2, z_3, z_4\}$. Subgraph $\mathbb{G}$, which by the definition is induced by $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{X}_1$, is drawn in Figure 8b. Matrix $\bar{\mathbf{P}}$ is:

$$\bar{\mathbf{P}} = \begin{pmatrix} 1 & a_{12} & 0 & 0 & a_{15} & a_{16} & -b_{11} & -b_{21} & 0 & 0 \\ 0 & 1 & a_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & a_{34} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -b_{43} & 0 \\ 0 & 0 & 0 & a_{54} & 1 & 0 & 0 & 0 & 0 & -b_{54} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Matrix $\mathbf{M}$ takes rows 2, 3, 4, 5, 6, and columns 3, 4, 8, 9, and 10 of matrix $\bar{\mathbf{P}}$, so I get:

$$\mathbf{M} = \begin{pmatrix} \underline{a_{23}} & 0 & 0 & 0 & 0 \\ 1 & \underline{a_{34}} & 0 & 0 & 0 \\ 0 & 1 & 0 & \underline{-b_{43}} & 0 \\ 0 & a_{54} & 0 & 0 & \underline{-b_{54}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{17}$$

There are two independent paths in $\mathbb{G}$ starting with nodes in $\mathcal{X}_1$ and reaching $\mathcal{Y}_1$, see Figure 8b, they are $z_3 \to y_4 \to y_3 \to y_2$ and $z_4 \to y_5$, so $N = 2$. The sets of parameters associated with these paths are $\{-b_{43}, a_{34}, a_{23}\}$ and $\{-b_{54}\}$. The lengthiest unconstrained partial permutation in $\mathbf{M}$ is underlined in equation (17), and is $[a_{23} \cdot a_{34} \cdot (-b_{43}) \cdot (-b_{54})]$. This permutation has four elements, so $l = 4$. Finally, there are 2 nodes in set $\mathcal{Y}_2$, so $|\mathcal{Y}_2| = 2$.

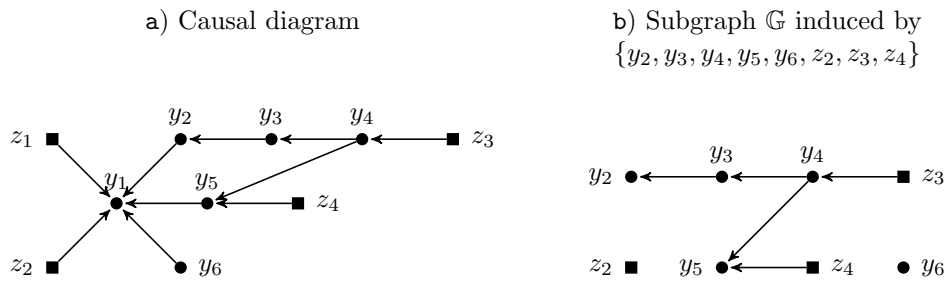


FIGURE 8. Example of causal diagram and subgraph $\mathbb{G}$

Lemma 1. *The length of the lengthiest unconstrained partial permutation in* $\mathbf{M}$ *is equal to the number of independent paths in* $\mathbb{G}$ *starting with nodes in* $\mathcal{X}_1$ *and reaching* $\mathcal{Y}_1$ *plus the number of nodes in* $\mathcal{Y}_2$*:*

$$l = N + |\mathcal{Y}_2|$$

Proof. Step 1. Prove that two paths intersect in $\mathbb{G}$ if and only if the parameters associated with these paths do not pertain to the same partial permutation in $\mathbf{M}$.

Indeed, two paths intersect in $\mathbb{G}$ if and only if there exists a node $x_j \in \mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{X}_1$ such that at least one of the following conditions hold:

(1) There are two incoming edges to node x_j associated with two different paths, in which case the parameters associated with these edges are located in the same row of $\mathbf{M}$.
(2) There are two outgoing edges from x_j associated with two different paths, in which case the parameters associated with the outgoing edges are located in the same column of $\mathbf{M}$.

Two parameters pertain to the same row or to the same column of $\mathbf{M}$ if and only if they do not pertain to the same permutations.

Step 2. Prove that if graph $\mathbb{G}$ is empty then $l = |\mathcal{Y}_2|$.

If $\mathbb{G}$ is empty, the only non-zero parameters of $\bar{\mathbf{P}}$ included into $\mathbf{M}$ are the on-diagonal elements of $\mathbf{A}$, which are normalized to be strictly positive. There are $|\mathcal{Y}_2|$ such parameters in $\mathbf{M}$, and all of them are located in different columns and different rows, which gives a permutation of length $|\mathcal{Y}_2|$.

In example (16), matrix $\mathbf{M}$ associated with the empty graph is:

$$\mathbf{M}_{\text{empty}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the length of the lengthiest unconstrained partial permutation is 2, which equals $|\mathcal{Y}_2|$.

Step 3. Prove that $l \geq N + |\mathcal{Y}_2|$.

Start with the empty graph spanning $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{X}_1$, which gives the permutation of length $|\mathcal{Y}_2|$, as it was described in Step 2. Add independent paths from $\mathbb{G}$ into this graph one-by-one. When a new path $x_{j_0}x_{j_1}\dots x_{j_s}$ is added to the graph, modify the permutation in the following manner:

(1) Add element $p_{j_1j_0}$ from matrix $\bar{\mathbf{P}}$ to the permutation. Since $x_{j_0} \in \mathcal{X}_1$ and $x_{j_1} \in \mathcal{Y}_1 \cup \mathcal{Y}_2$, parameter $p_{j_1j_0}$ is in $\mathbf{M}$.

(2) For $k = 1, 2, \dots, s-1$, remove $p_{j_kj_k}$, and add $p_{j_kj_{k+1}}$. Since $x_{j_k} \in \mathcal{Y}_2$ and $x_{j_{k+1}} \in \mathcal{Y}_1 \cup \mathcal{Y}_2$, parameters $p_{j_kj_k}$ and $p_{j_kj_{k+1}}$ are in $\mathbf{M}$. Since the new path is independent of the previously added paths, $p_{j_kj_{k+1}}$ is located in a different row and in a different column than the permutations associated with the previously added paths, so it was included into the permutation. Each parameter $p_{j_0j_1}, p_{j_1j_2}, \dots, p_{j_{s-1}j_s}$ and the parameters kept from the previous paths pertain to the same permutation by the result demonstrated in Step 1.

Therefore, adding a new independent path increases the number of parameters included into the permutation by 1. When other parameters, which are not associated with the considered independent paths, are added to matrix $\mathbf{M}$, the length of the permutation does not decrease, so $l \geq N + |\mathcal{Y}_2|$.

In example (16), adding path $z_3 \to y_4 \to y_3 \to y_2$ gives:

$$\begin{pmatrix} 0&0&0&0&0\\ 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix} \Rightarrow \begin{pmatrix} 0&0&0&0&0\\ 1&0&0&0&0\\ 0&1&0&-b_{43}&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix} \Rightarrow \begin{pmatrix} 0&0&0&0&0\\ 1&a_{34}&0&0&0\\ 0&0&0&-b_{43}&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix} \Rightarrow \begin{pmatrix} a_{23}&0&0&0&0\\ 0&a_{34}&0&0&0\\ 0&0&0&-b_{43}&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix}$$

and adding $z_4 \to y_5$ produces:

$$\begin{pmatrix} a_{23}&0&0&0&0\\ 0&a_{34}&0&0&0\\ 0&0&0&-b_{43}&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{pmatrix} \Rightarrow \begin{pmatrix} a_{23}&0&0&0&0\\ 0&a_{34}&0&0&0\\ 0&0&0&-b_{43}&0\\ 0&0&0&0&-b_{54}\\ 0&0&0&0&0 \end{pmatrix}$$

which gives a permutation of length 4.

Step 4. Prove that $N \geq l - |\mathcal{Y}_2|$

Consider a permutation of length l. Since all parameters associated with one permutation are located in different columns of matrix $\bar{\mathbf{P}}$, at least $l - |\mathcal{Y}_2|$ parameters must be located in the columns associated with the indices of $\mathcal{X}_1$. Let me prove that each such parameters guarantees the existence of one path from $\mathcal{X}_1$ to $\mathcal{Y}_1$, and from Step 1 I know that all these paths must be independent.

Consider one such parameter, say $p_{j_1j_0}$, where $x_{j_0} \in \mathcal{X}_1$. If $x_{j_1} \in \mathcal{Y}_1$, then the path is found. Assume that $x_{j_1} \notin \mathcal{Y}_1$, so $x_{j_1} \in \mathcal{Y}_2$. Since $p_{j_1j_0}$ have been included into the permutation, parameter $p_{j_1j_1}$, which is normalized to be positive, cannot be included into this permutation, because it is in the sam row as $p_{j_1j_0}$. Therefore, column j_1 either is not included into permutation, or there exists parameter $p_{j_2j_1}$, which is

included. In the first case there must be at least one more parameter included into the permutation from the columns associated with the indices of $\mathcal{X}_1$, because otherwise the total length of the permutation would be less that l, so consider that parameter instead of $p_{j_1 j_0}$. In the second case, see where the edge associated with $p_{j_2 j_1}$ leads to. If $x_{j_2} \in \mathcal{Y}_1$, then a path have been found. If $x_{j_2} \in \mathcal{Y}_2$, keep going through the permutation until $\mathcal{Y}_1$ is reached or this is determined that there exists another parameter in this permutation in a column associated with $\mathcal{X}_1$.

Therefore, there is at least $l - |\mathcal{Y}_2|$ independent paths starting with a node in $\mathcal{X}_1$ and reaching nodes in $\mathcal{Y}_1$. Because adding new edges does not decrease the number of the existing independent paths, $N \geq l - |\mathcal{Y}_2|$

From Steps 3 and Step 4 I conclude that $l = N + |\mathcal{Y}_2|$ □

A.2. **Review of Rank Condition.** Because of the normality assumption, $f(Y|Z)$ can be uniquely specified by matrices $\mathbf{\Lambda}$ and $\mathbf{\Omega}$, which are defined by:

$$\mathbb{E}(Y|Z) = \mathbf{A}^{-1}\mathbf{B} \cdot Z \equiv \mathbf{\Lambda} \cdot Z \tag{18a}$$

$$\text{Var}(Y - \mathbb{E}(Y|Z)) = \left(\mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A}\right)^{-1} \equiv \mathbf{\Omega} \tag{18b}$$

Knowing matrices $\mathbf{\Lambda}$ and $\mathbf{\Omega}$, however, does not suffice for estimation of parameters $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{\Sigma}$ of the structural model (8) unless $n = 1$. The reason is that there exist many different structural models observationally equivalent to model (8), and all observationally equivalent models by definition produce the same values of $\mathbf{\Lambda}$ and $\mathbf{\Omega}$. Indeed, two models with different parameter values $(\mathbf{A}, \mathbf{B}, \mathbf{\Sigma})$ and $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}})$ are observationally equivalent if and only if there exists nonsingular $n \times n$ matrix $\mathbf{D}$ such that $\tilde{\mathbf{A}} = \mathbf{DA}$, $\tilde{\mathbf{B}} = \mathbf{DB}$, and $\tilde{\mathbf{\Sigma}} = \mathbf{D\Sigma D}^T$, which result can be verified directly using (18). To estimate the structural model, therefore, additional restrictions need to be imposed on the matrices of parameters, which are known as the identification constraints.

The identification constraints on row i of parameters $\underset{n\times(n+m)}{\bar{\mathbf{P}}} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ {\scriptstyle n\times n} & {\scriptstyle n\times m} \end{pmatrix}$ are written as:

$$e_i^T \bar{\mathbf{P}} \mathbf{\Psi}_i = 0 \tag{19}$$

where e_i is the i^{th} row of the identity matrix, and $\mathbf{\Psi}_i$ is the matrix summarizing the constraints imposed on row i of $\bar{\mathbf{P}}$.

Consider example (5). Matrix $\bar{\mathbf{P}}$ for this model is given by:

$$\bar{\mathbf{P}} = \begin{pmatrix} 1 & 0 & 0 & -b_{11} & 0 \\ -a_{21} & 1 & 0 & 0 & -b_{22} \\ -a_{31} & -a_{32} & 1 & 0 & 0 \end{pmatrix}$$

The constraints on parameters are summarized by:

$$\boldsymbol{\Psi}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{\Psi}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \boldsymbol{\Psi}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The identification of a given parameter is usually verified in the literature using the rank condition. The rank condition says that the parameters in row i of matrix $\bar{\mathbf{P}}$ are identified if and only if $\operatorname{rank}\left(\bar{\mathbf{P}}\boldsymbol{\Psi}_i\right) = n-1$, see, for example, Greene (2012). In the considered example (5), all parameters are identified in almost all parameter points, because in almost all parameter points I have:

$$\operatorname{rank}\left(\bar{\mathbf{P}}\boldsymbol{\Psi}_1\right) = \operatorname{rank}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -b_{22} \\ -a_{32} & 1 & 0 \end{pmatrix} = 2;\ \operatorname{rank}\left(\bar{\mathbf{P}}\boldsymbol{\Psi}_2\right) = \operatorname{rank}\begin{pmatrix} 0 & -b_{11} \\ 0 & 0 \\ 1 & 0 \end{pmatrix} = 2;\ \operatorname{rank}\left(\bar{\mathbf{P}}\boldsymbol{\Psi}_3\right) = \operatorname{rank}\begin{pmatrix} -b_{11} & 0 \\ 0 & -b_{22} \\ 0 & 0 \end{pmatrix} = 2.$$

A.3. **Proof of Proposition 1.** Let $\mathcal{P}_i$ be the set of parents of y_i, and $\bar{\mathcal{P}}_i = \mathcal{P}_i^c \setminus y_i$, where $\mathcal{P}_i^c$ is the complement of $\mathcal{P}_i$ in $\mathcal{X}$, and "\" is the set difference operator. Let $\mathcal{Y}_{-i} = \mathcal{Y} \setminus y_i$.

Proof of Proposition 1. Consider matrix $\mathbf{M}_i$ obtained from $\bar{\mathbf{P}}\boldsymbol{\Psi}_i$ by deleting the i^{th} row. Since each element in the i^{th} row of $\bar{\mathbf{P}}\boldsymbol{\Psi}_i$ is constrained to zero by definition of $\boldsymbol{\Psi}_i$, I have: $\operatorname{rank}\left(\mathbf{M}_i\right) = \operatorname{rank}\left(\bar{\mathbf{P}}\boldsymbol{\Psi}_i\right)$.

By definition of $\boldsymbol{\Psi}_i$, each column of $\bar{\mathbf{P}}\boldsymbol{\Psi}_i$, as well as each column of $\mathbf{M}_i$, has the index of a variable from $\bar{\mathcal{P}}_i$, and each node from $\bar{\mathcal{P}}_i$ has the index of a column of $\mathbf{M}_i$. Therefore, using notation from Lemma 1, I can write: $\mathcal{Y}_2 \cup \mathcal{X}_1 = \bar{\mathcal{P}}_i$. Each row of $\mathbf{M}_i$ has the index of an endogenous variable, and each endogenous variable except y_i has the index of a column of $\mathbf{M}_i$, so I can use: $\mathcal{Y}_1 \cup \mathcal{Y}_2 = \mathcal{Y}_{-i}$. This gives: $\mathcal{Y}_1 = \mathcal{Y}_{-i} \cap \mathcal{P}_i$, $\mathcal{Y}_2 = \mathcal{Y}_{-i} \cap \bar{\mathcal{P}}_i$, and $\mathcal{X}_1 = \mathcal{Z} \cap \bar{\mathcal{P}}_i$.

Let me prove the necessity of the graphical rank condition. If y_i is identified then the rank condition is satisfied, so $\operatorname{rank}\left(\mathbf{M}_i\right) = n-1$, and there exists $n-1$ independent columns in $\mathbf{M}_i$; consider any set of $n-1$ independent columns. The determinant of the matrix obtained from the independent columns of $\mathbf{M}_i$ must be not zero, therefore, in Leibniz formula for determinant of $\mathbf{M}_i$, there exists at least one unconstrained permutation of length $n-1$. Then, from Lemma 1, there exists $n-1-|\mathcal{Y}_2| = |\mathcal{Y}_1|$ independent paths starting in $\mathcal{X}_1$ and reaching $\mathcal{P}_i$. Therefore, for each $y_j \in \mathcal{Y}_{-i} \cap \mathcal{P}_i$ there exists an independent path starting in $\mathcal{Z} \cap \bar{\mathcal{P}}_i$ and reaching y_j. Proposition 1 also says that for each node $z_j \in \mathcal{P}_i \cap \mathcal{Z}$ there exists an independent path starting in $\mathcal{Z}$ and reaching z_j; however, the latter condition is always satisfied.

Now let me prove the sufficiency. If for each parent of y_i there exists and independent identifying path, then for each $y_j \in \mathcal{Y}_1$ there exists an independent path starting with a node in $\mathcal{X}_1$ and reaching y_j. By Lemma 1, there exists a partial permutation of length $(n-1)$ in $\mathbf{M}_i$ such that each parameter of this permutation is not constrained to zero. I take the columns of $\mathbf{M}_i$ associated with this permutation, and calculate the determinant of the obtained square matrix. Since the determinant can be calculated using Leibniz formula

as a sum over all permutations, and since one permutation is not constrained to zero, the determinant is zero only if this non-zero permutation is exactly offset by other non-zero permutations, which does not happen in almost all parameter points. Therefore, in almost all parameter points $\operatorname{rank}(\mathbf{M}_i) = (n-1)$, so the rank condition is satisfied. □

Appendix B. Proof of Proposition 2

B.1. **Review of Rubio-Ramírez et al. (2010) condition for identification.** Unlike the literature on simultaneous equations models, the literature on structural vector autoregression models usually assumes that the structural shocks are independent, so matrix $\mathbf{\Sigma}$ is diagonal. In the Gaussian case, two SVAR models are said to be observationally equivalent if they produce the same values of $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ defined by 18. This is well-known that two SVAR models defined by parameter points $(\mathbf{A}, \mathbf{B})$ and $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are observationally equivalent if and only if there exists rotation matrix $\mathbf{R}$ such that $\tilde{\mathbf{A}} = \mathbf{R}\mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{R}\mathbf{B}$, where rotation matrix $\mathbf{R}$ by definition must satisfy $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. Since the rotation matrix has $n(n-1)/2$ degrees of freedom, a necessary condition for identification formulated by Rothenberg (1971) requires at least $n(n-1)/2$ constraints imposed on matrix $\bar{\mathbf{P}} = (\mathbf{A} \quad -\mathbf{B})$ for full identification.

Rubio-Ramírez et al. (2010) propose a sufficient condition for identification, which is applicable to a much larger class of identification constraints than I consider in this paper. However, I concise the analysis to the case, where the identification constraints are formulated as (19). To verify the identification of parameters located in the i^{th} row of $\bar{\mathbf{P}}$, calculate the rank of matrices $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_i$ composed in the following way:

$$\mathbf{M}_j = \begin{pmatrix} \begin{bmatrix} \\ \bar{\mathbf{P}}\mathbf{\Psi}_j \\ \\ \end{bmatrix} & \begin{bmatrix} \mathbf{I}_{j\times j} \end{bmatrix} \\ & \\ & \begin{bmatrix} \mathbf{0}_{(n-j)\times j} \end{bmatrix} \end{pmatrix} \tag{20}$$

The rank of matrices $\mathbf{M}_j$ for $j = 1, 2, \ldots, i$ may depend on the order of variables in vector Y. Rubio-Ramírez et al. (2010) prove that if there exists such order that for $j = 1, 2, \ldots, i$ the rank of $\mathbf{M}_j$ is n, then the i^{th} row of $\bar{\mathbf{P}}$ is globally identified in almost all parameter points.

In example (6), to verify the identification of parameters under the assumption of shocks independence, reorder variables in the reverse order, and calculate the rank of the following matrices:

$$\text{(21)} \qquad \mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -b_{22} & 0 \\ -b_{11} & -b_{12} & 0 \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -b_{11} & 0 & 0 \end{pmatrix} \quad \mathbf{M}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Matrices $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ have rank 3 in almost all parameter points, therefore, this model is fully identified in almost all parameter points.

Theory of partial identification, reviewed in Christiano et al. (1999), proposes another sufficient condition for identification. If all variables in Y can be divided into three groups, such that the first group has the only variable y_i, the second group includes the variables, which influence y_i but not influenced by y_i, and the third group includes the variables influenced by y_i, but which do not influence y_i, then y_i is identified. I combine the sufficient condition of Rubio-Ramírez et al. (2010) with the theory of partial identification, and in this way I can prove partial identification of a new class of models. Consider, for example, the following identification restrictions:

$$\text{(22)} \qquad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \end{pmatrix}_{4\times 0}$$

The theory of partial identification does not prove identification of any parameter in this model, because each variable of Y pertain to one of causal cycles. Rubio-Ramírez et al. (2010) condition for identification is not satisfied for any parameters, because whichever the order of variables in Y, $\text{rank}\,(\mathbf{M}_1) < 4$. However, I can use Proposition 2 to show that a combination of these approaches suffices to prove that the third and forth lines of $\mathbf{A}$ in (22) are identified.

B.2. **Proof of Proposition 2.** Use the notation that was introduced in Appendix A, and add the following one. Let $\Phi \subset \mathcal{Y}$ be the set of nodes, which have been identified, and Φ^c be the complement of Φ in $\mathcal{Y}$, so $\Phi^c = \mathcal{Y} \setminus \Phi$, where "\" is the set difference operator. Let $\mathcal{D}_i$ be the set of descendants of y_i, $\mathcal{D}_i^c = \mathcal{Y} \setminus \mathcal{D}_i$, and $\bar{\mathcal{D}}_i = \mathcal{D}_i^c \setminus y_i$. By definition in Proposition 2, a path in the causal diagram is identifying path for parent $y_j \in \mathcal{P}_i$ of node y_i if it starts with a node in $\mathcal{Z} \cup \Phi \cup \bar{\mathcal{D}}_i$ and reaches y_j. Proposition 2 says that if for each node from $\mathcal{P}_i$ there exists an independent identifying path, node y_i is globally identified in almost all parameter points.

Proof of proposition 2. Since the order of variables is arbitrary, reorder the variables in such way that the variables from $\bar{\mathcal{D}}_i$ have indices $1, 2, \ldots, n_1$, where $n_1 = \left|\bar{\mathcal{D}}_i\right|$. Divide $\mathbf{A}$ into four matrices in a similar manner:

$$\mathbf{A} = \begin{pmatrix} \underset{n_1 \times n_1}{\mathbf{A}_{11}} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

Observe that matrix $\mathbf{A}_{12}$ must be zero, because in the opposite case there would exist a path from a descendant of y_i to a non-descendant, but then the latter vertex would also be descendant of y_i, which produces a contradiction.

Apply the argumentation from the literature on partial identification, reviewed, for example, in Christiano et al. (1999), which proves that if block $\mathbf{A}_{12}$ is constrained to $\mathbf{0}$, then two models defined by parameter points $(\mathbf{A}, \mathbf{B})$ and $\left(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}\right)$ satisfying this restriction are observationally equivalent if and only if there exists rotation matrix $\mathbf{R}$, such that $\tilde{\mathbf{A}} = \mathbf{R}\mathbf{A}$, $\tilde{\mathbf{B}} = \mathbf{R}\mathbf{B}$, and $\mathbf{R}$ has the following block structure:

$$\mathbf{R} = \begin{pmatrix} \underset{n_1 \times n_1}{\mathbf{R}_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix} \tag{23}$$

Now use the approach developed in Rubio-Ramírez et al. (2010). Reorder the variables in $\mathcal{Y}$ in such way that the variables with indices $1, 2, \ldots, n_1$ be the non-descendants of y_i, variables with indices $n_1 + 1, n_1 + 2, \ldots, i - 1$ be the variables associated with $\Phi \cap \mathcal{D}_i$, y_i be the node which identification is being examined, and variables with indices $i + 1, i + 2, \ldots, n$ be the variables associated with $\bar{\Phi} \cap \mathcal{D}_i$.

Consider matrix $\hat{\mathbf{M}}_i$ obtained from $\mathbf{P}\mathbf{\Psi}_i$ by deleting rows $1, 2, \ldots, i$, and prove that if y_i is not identified then the row rank of $\hat{\mathbf{M}}_i$ is not full, in which case the rank of $\mathbf{M}_i$ defined by (20) is also not full. Indeed, if y_i is not identified then there must exist rotation matrix $\mathbf{R}$, having the following properties. First, because of its special structure given by (23), and because nodes $y_{n_1+1}, y_{n_1+2}, \ldots, y_{i-1}$ are identified, $\mathbf{R}$ has the following structure:

$$\mathbf{R} = \begin{pmatrix} \underset{(i-1)\times(i-1)}{\mathbf{I}} & \mathbf{0} \\ \mathbf{0} & \underset{(n-i)\times(n-i)}{\mathbf{R}_{33}} \end{pmatrix} \tag{24}$$

Second, since y_i is not identified, at least one non-diagonal element in the first row of $\mathbf{R}_{33}$ must be different from zero. Let v_i^T be the vector obtained from the first row of $\mathbf{R}_{33}$ by removing the first element, so I have $v_i \neq 0$. Finally, since the two models must satisfy the identification constraints, I have $e_i \mathbf{P}\mathbf{\Psi}_i = 0$ and

$e_i\mathbf{R}\mathbf{P}\mathbf{\Psi}_i = \mathbf{0}$, so $e_i\left(\mathbf{R}-\mathbf{I}\right)\mathbf{P}\mathbf{\Psi}_i = 0$. Taking into account the properties of $\mathbf{R}$, I get $v_i^T\hat{\mathbf{M}}_i = 0$, so the row rank of $\hat{\mathbf{M}}_i$ is not full. This proves that if the row rank of $\hat{\mathbf{M}}_i$ is full then node y_i is identified.

The final step is to apply Lemma 1. By construction of $\hat{\mathbf{M}}_i$, $\mathcal{Y}_2 \cup \mathcal{X}_1 = \bar{\mathcal{P}}_i$, and $\mathcal{Y}_1 \cup \mathcal{Y}_2 = \Phi^c \cap \mathcal{D}_i$. Therefore, $\mathcal{Y}_1 = \Phi^c \cap \mathcal{D}_i \cap \mathcal{P}_i$, $\mathcal{Y}_2 = \Phi^c \cap \mathcal{D}_i \cap \bar{\mathcal{P}}_i$, and $\mathcal{X}_1 = \bar{\mathcal{P}}_i \cap \left(\Phi \cup \bar{\mathcal{D}}_i \cup \mathcal{Z}\right)$. Lemma 1 proves that if for each $y_j \in \mathcal{Y}_1$ there exists an independent path starting in $\mathcal{X}_1$ and reaching y_j, then the row rank of $\hat{\mathbf{M}}_i$ is full in almost all parameter points, so y_i is identified in almost all parameter points. Proposition 2 also requires an independent identifying path for each variable in $\mathcal{P}_i \cap \left(\Phi \cup \bar{\mathcal{D}}_i \cup \mathcal{Z}\right)$, but this condition is always satisfied.

□

Appendix C. Proof of Proposition 4

Proof of Proposition 4. Since the order of variables is arbitrary, assume that $i = 1$, so the first row in $\bar{\mathbf{P}}\mathbf{\Psi}_1$ is constrained to zero. In this appendix I prove the sufficiency of the reduced form rank condition for identification. That is, I assume that the rank condition is not satisfied, and prove that in this case $\operatorname{rank}\left(\mathbf{\Pi}\left(\mathcal{P}_1|Z_1\right)\right) < |\mathcal{P}_1|$. To prove the necessity of the reduced form rank condition, I need to assume that $\operatorname{rank}\left(\mathbf{\Pi}\left(\mathcal{P}_1|Z_1\right)\right) < |\mathcal{P}_1|$, and make all steps in the reverse order to show that the full form rank condition is not satisfied.

If the rank condition is not satisfied, there exists vector $V = \begin{pmatrix} 0 & v_2 & v_3 & \dots & v_n \end{pmatrix}^T \neq 0$ such that $V^T\mathbf{P}\mathbf{\Psi}_i = 0^T$. Rewrite (8) in terms of expectations $Y^{\mathbb{E}} = \mathbb{E}(Y|Z)$, and multiply it from the left by V^T:

$$V^T\mathbf{A}Y^{\mathbb{E}} = V^T\mathbf{B}Z \tag{25}$$

Make the following observations about (25). First, the parameters from the first rows of matrices $\mathbf{A}$ and $\mathbf{B}$ are not present in this equation, because $v_1 = 0$. Second, $V^T\bar{\mathbf{P}}$ cannot be proportional to the first line of $\bar{\mathbf{P}}$, because otherwise the first line of $\mathbf{P}$ could be expressed as a linear combination of the other lines, in which case matrix $\mathbf{A}$ would be singular, but I have assumed that this is not true. Third, for the same reason, $V^T\mathbf{A}$ is not zero. Finally, by construction of V, all columns in $V^T\mathbf{A}$ and $V^T\mathbf{B}$ associated with variables in $\bar{\mathcal{P}}_i$ are zero, so variable from $\bar{\mathcal{P}}_i$ are ignored in (25). Therefore, I can rewrite (25) in the following form:

$$V^T\tilde{\mathbf{A}}\tilde{Y}^{\mathbb{E}} = V^T\tilde{\mathbf{B}}\tilde{Z}, \tag{26}$$

where matrix $\tilde{\mathbf{A}}$ is obtained by deleting the columns associated with the indices of nodes $\bar{\mathcal{P}}_1 \cap \mathcal{Y}$ from matrix $\mathbf{A}$, matrix $\tilde{\mathbf{B}}$ is obtained from $\mathbf{B}$ by deleting the columns associated with $\bar{\mathcal{P}}_1 \cap \mathcal{Z}$, finally, $\tilde{Y}^{\mathbb{E}}$ and $\tilde{Z}$ are obtained from $Y^{\mathbb{E}}$ and Z by removing the variables associated with nodes in $\bar{\mathcal{P}}_1$.

Now I have two linear combinations of $\tilde{Y}^{\mathbb{E}}$ and $\tilde{Z}$, which are zero in the equilibrium: the first linear combination is given by the first line in (8), and the second is given by (26). Both this combination are not zero and they are linearly independent, because, as I discuss above, matrix $\mathbf{A}$ would be singular. Define these combinations as:

$$\Lambda_1^T \begin{pmatrix} \tilde{Y}^{\mathbb{E}} \\ \tilde{Z} \end{pmatrix} = 0 \tag{27a}$$

$$\Lambda_2^T \begin{pmatrix} \tilde{Y}^{\mathbb{E}} \\ \tilde{Z} \end{pmatrix} = 0 \tag{27b}$$

where Λ_1 is obtained from the first line of (8), and Λ_2 is just another way to write(26); as I discuss above, Λ_1 and Λ_2 a linearly independent.

Consider matrix $\hat{\mathbf{\Pi}}$ defined by:

$$\mathbb{E}\left(\left(\begin{pmatrix} \tilde{Y}^{\mathbb{E}} \\ \tilde{Z} \end{pmatrix} \middle| Z \right) \right) = \hat{\mathbf{\Pi}} Z \tag{28}$$

By this definition, matrix $\mathbf{\Pi}(\mathcal{P}_i|\mathcal{Z}_i)$ from Proposition 4 can be obtained by deleting the first row from $\hat{\mathbf{\Pi}}$.

For $j = 1, 2$ I have $\Lambda_j^T \hat{\mathbf{\Pi}} = 0$. Since Λ_1 and Λ_2 are linearly independent,

$$\operatorname{rank}\left(\mathbf{\Pi}(\mathcal{P}_i|\mathcal{Z}_i)\right) \leq \operatorname{rank}\left(\hat{\mathbf{\Pi}}\right) \leq \operatorname{nrow}\left(\hat{\mathbf{\Pi}}\right) - 2 = \operatorname{nrow}\left(\mathbf{\Pi}(\mathcal{P}_{\mathbf{i}}|\mathcal{Z}_{\mathbf{i}})\right) - 1$$

So the row rank of $\mathbf{\Pi}(\mathcal{P}_i|\mathcal{Z}_i)$ is not full. □

NIKOLAY AREFIEV
National Research University Higher School of Economics. E-mail: n.are_ev@gmail.com. Address: office 708A, Shabolovka 31,
119049, Moscow, Russia. Cell phone: +7 926 825 7147

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.