NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nikolay Arefiev

GRAPHICAL INTERPRETATIONS OF RANK CONDITIONS FOR IDENTIFICATION OF LINEAR GAUSSIAN MODELS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 124/EC/2016

GRAPHICAL INTERPRETATIONS OF RANK CONDITIONS FOR IDENTIFICATION OF LINEAR GAUSSIAN MODELS

NIKOLAY AREFIEV

Abstract. The literature on graphical models and the literature on identification pursue similar goals, but do not use entirely each other's results, because represent them in different languages. To ease the communication between these fields, I translate the most important theorems on identification of linear Gaussian Simultaneous Equations Models (SEMs) and Structural Vector Autoregressions (SVARs) into the language of graphical models. I propose graphical interpretations of the rank conditions for identification of SEMs, of the rank condition of Rubio-Ramirez et al (2010) for identification of SVARs with linear and nonlinear restrictions, and of the theory of partial identification for SVARs.

Keywords: graphical models, identification, rank condition.

JEL codes: C30

1. Introduction

The probabilistic graphical approach has a lot of successful applications in the computer science, medicine and biology (Koller (2009); Pearl (2009)), and it gains popularity in econometrics (see Ahelegbey et al., 2014; Bryant and Bessler, 2011; Demiralp et al., 2014; Fragetta and Melina, 2013; Hoover, 2005; Kwon and Bessler, 2011; Oxley et al., 2009; Phiromswad, 2014; Reale and Wilson, 2001; Richardson and Spirtes, 1999; Wilson and Reale, 2008, and many others). However, the literature on graphical models and the econometric literature on identification use different languages to represent the results: the literature on graphical model usually formulates the theorems in terms of causal diagrams (Brito and Pearl (2002b); Tian (2005); Chen and Pearl (2014)), and the econometric literature represents the results in terms of matrix algebra (for example, see Greene, 2012; Rubio-Ramírez et al., 2010; Christiano et al., 1999). These branches of research, however,

*February 14, 2016.
The 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). The paper is a part of a larger projects on data-oriented identification, and I thank Alina Arefeva, Antoine d'Autume, Svetlana Bryzgalova, Jean-Bernard Chatelain, Boris Demeshev, Jean-Pierre Drugeon, Jean-Marie Dufour, Bulat Gafarov, James Hamilton, Oleg Itskhoki, Maarten Janssen, Ramis Khabibulin, Sergey Kusnetzov, Jessie Li, Judea Pearl, Anatoly Peresetsky, Christopher Sims, Alain Trognon, Ilya Voskoboynikov, and Mark Watson for helpful comments and suggestions on the project. 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.

do not substitute for each other. The econometric literature is better developed for cyclical models. The literature on graphical identification has more results for identification of models with intricate assumptions about the independence of structural shocks (Chen and Pearl (2014)), and provides tools for testing the identification assumptions. It would be desirable, therefore, to represent the results from these branches using a common language, and in this way to ease the communication between the researchers.

This paper fills this gap and translates the most important results known about identification in the econometrics into the language of graphical models. I propose graphical identification of the rank condition for identification of simultaneous equations models (SEMs) (for example, see Greene (2012)), the condition of Rubio-Ramírez, Waggoner and Zha for identification of SVARs (Rubio-Ramírez et al. (2010)), and of the theory of partial identification (as reviewed in Christiano et al. (1999)).

2. Graphical Interpretation of the Rank Condition for SEMs

Consider the following Simultaneous Equations Model (SEM):

$$\mathbf{A}Y = \mathbf{B}Z + \mathcal{E} \tag{1}$$

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 Proposition 1 below, where $\mathbf{\Sigma}$ is assumed to be a symmetric positive definite matrix without any identifying assumptions imposed. The constant term is omitted in (1) 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 $i = 1, 2, \dots, n : a_{i,i} > 0$ 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 (1).

If no identification constraints are imposed on (1), 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, in this section I consider only those identification constraints, which restrict particular parameters to zero. All identification constraints are summarized by the conditional causal diagram[1] defined as follows:

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 $\mathbf{P}$.

The literature on causality (Pearl (2009)) presumably works with unconditional causal diagrams, but in this paper I consider only conditional diagram. Hereafter, I use 'conditional causal diagram" and "causal diagram" as synonyms.

Throughout the paper I consider the following example. Assume that the simultaneous equation model is:

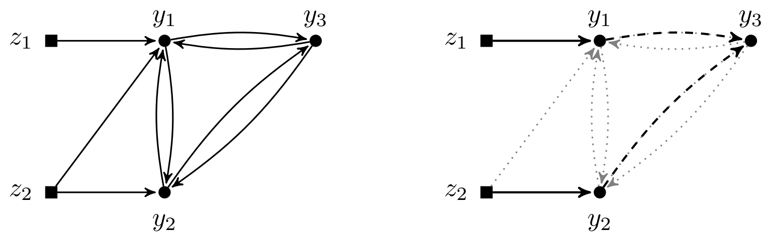
$$(2)\qquad \begin{pmatrix} a_{12} & a_{12} & a_{13} \\ a_{22} & a_{22} & a_{23} \\ a_{32} & a_{32} & a_{13} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ 0 & b_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix},$$

where coefficients b_{12}, b_{21} and b_{32} are constrained to zero, so they are substituted by zeros in (2). The conditional causal diagram for model (2) is depicted in Figure 1a, and it have been drawn in the following way. I have five random variables: y_1, y_2, y_3, z_1, and z_2, so I have drawn five respective vertices. The first equation in (2) is associated with node y_1 in Figure 1a. Since no coefficients in this line are constrained to zero, each other node is parent of y_1 in Figure 1a. The second line is associated with node y_2, and since coefficient b_{21} is constrained to zero, there is no edge $z_1 \to y_2$ in Figure 1a, but the other nodes are parents of y_2. Finally, coefficients b_{31} and b_{32} are constrained to zero, z_1 and z_2 are not parents of y_3.

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_1} is ancestor of y_{j_N}, and y_{j_N} is descendant of y_{j_1}. If there is a path which starts and ends with the same node, this path is called a *cycle*. If there is no cycles on the conditional causal diagram, the model is *recursive*, otherwise it is *cyclical.* Two paths are independent if they do not intersect on any node. Each node is interpreted as a path of length 1.

[1]The conditional causal diagram can be interpreted as a C-component of the full causal diagram, see Tian (2005).

FIGURE 1. Conditional causal diagram for model (2) and identification of y_3.



a) Conditional causal diagram b) Identification of y_3

Model (2) is cyclical, because there are many cycles in the causal diagram in Figure 1a, for example, $y_1y_2y_1$. Cyclical models have multiple causal representation. For example, if I change the order of equations in (2), I get a different causal diagram. Hopefully, the results presented bellow do not depend on the particular causal representation that I use.

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 the constraints summarized by the conditional causal diagram suffice for the identification of all parameters in the i^{th} rows of $\mathbf{A}$ and $\mathbf{B}$ in almost all parameter points.

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 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 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. □

Consider the causal diagram depicted in Figure 1a, and see whether node y_3 is identified. This node has two parents: y_1 and y_2, so I need 2 independent identifying paths for the identification of y_3. These paths do exist, see Figure 1b. The identifying path for y_1 is $z_1 \to y_1$. By Definition 2, this path starts with instrument z_1 and reaches y_1, which is parent of y_3. Similarly, the identifying path for y_2 is $z_2 \to y_2$, which starts with instrument z_2 and reaches parent y_2. These paths do not intersect on any node, so they are independent. Therefore, node y_3 is identified, which means that all parameters in the third line of (2) and structural shock ε_3 are identified in almost all parameter points.

The causal diagram in Figure 1, however, does not suffice for identification of y_1 or y_2. Indeed, node y_1 has 4 parents, and node y_2 has three parents, but only two primary instruments in the whole model are available. Since this is not possible to draw 3 or 4 independent paths starting with only 2 instruments, nodes y_1 and y_2 are not identified.

3. Graphical Identification of Models with Orthogonal Shocks

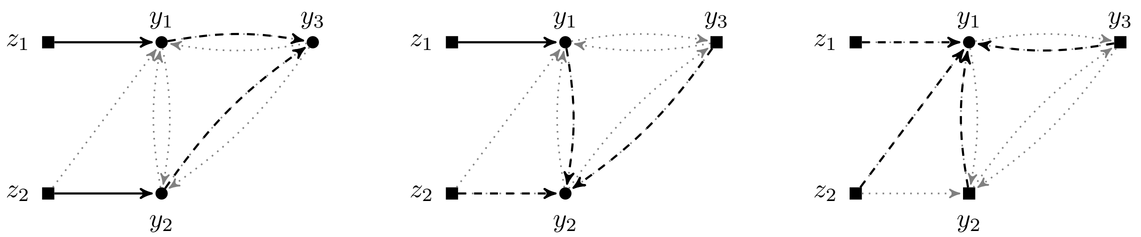
Assume that the structural shocks are orthogonal, so $\mathbf{\Sigma}$ is diagonal. When the independence assumption is made, some endogenous variables may possess the same properties as the primary instruments, so they can produce additional identifying paths and identify additional parameters. I introduce 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 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).

Figure 2. Identification under the shock orthogonality assumption.



a) Identification of y_3 b) Identification of y_2 c) Identification of y_1

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. Use in Proposition 3 below, and consider the case $\mathbf{F}(\mathbf{A}, \mathbf{B}) = (\mathbf{A} \ \ \mathbf{B})$. □

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.

Consider again model (2), and see which nodes are identified by causal diagram depicted in Figure 1 under assumption of independent structural shocks. As previously, node y_3 has two parents, y_1 and y_2, with identifying paths $z_1 \to y_1$ and $z_2 \to y_2$, so y_3 is identified, see Figure 2a. But because of the shock independence assumption, I can now use y_3 as a recursive instrument for identification of other nodes.

Consider node y_2, see Figure 2b. 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 $z_1 y_1$, so node y_2 is also identified. In the same way, it is possible to show that y_1 is also identified, see Figure 2c. Therefore, under the shock orthogonality assumption, the identification assumptions summarized by the causal diagram depicted in Figure 1a suffice for the full identification of the structural model.

Using recursive and respective instruments together, I can partially identify new models, which could not be identified using the results from the theory on partial identification and from the Rubio-Ramírez et al.'s

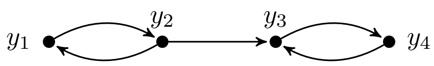


Figure 3. Causal diagram for example (3).

(2010) theorem when they are applied separately from each other. Consider a model with the following identification assumptions:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} \end{pmatrix}_{4\times 0} \tag{3}$$

The theory of partial identification does not suffice to identify any equation in this model, because each node of the causal diagram for this model is part of a causal cycle. In Appendix B.1 I demonstrate why the Rubio-Ramírez et al.'s (2010) Theorem 1 does not suffice to identify any node. Consider, however, the causal diagram for this model depicted in Figure 3. Node y_2 acts as a respective instrument for identification of y_4, and y_4 is recursive instrument for identification of y_3, so nodes y_3 and y_4 are identified.

4. Nonlinear Identifying Restrictions for SVAR models

A special case of (1) is the following structural vector autoregression (SVAR) model:

$$\mathbf{A}_0 Y_t = \sum_{i=1}^{l} \mathbf{A}_i Y_{t-i} + \mathcal{E}_t, \tag{4}$$

where l is the number of lags. This model reduces to (1) using variable substitution $\mathbf{A}_0 \equiv \mathbf{A}$, $[\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_l] \equiv \mathbf{B}$, $Y_t \equiv Y$, and $\left(Y_{t-1}^T \ Y_{t-2}^T \ \dots \ Y_{t-l}^T\right)^T \equiv Z$.

The literature on SVARs uses not only inclusions and exclusions summarized by the causal diagram, but also various nonlinear restrictions, such as restrictions imposed on the matrix of long-run impulse-responses defined by:

$$\mathbf{IR}_\infty = \left(A_0 - \sum_{i=1}^{l} A_i \right)^{-T}, \tag{5}$$

where $[\mathbf{IR}_\infty]_{ji} \equiv \mathrm{ir}_{ji}$ is the long-run response of y_i to ε_j. Identifying restrictions may require that particular entries of $\mathbf{IR}_\infty$ be or be not constrained zero.

To deal with long-run restrictions, I follow Rubio-Ramírez et al. (2010), and use parameter transformation function $\mathbf{F}(\mathbf{A}, \mathbf{B})$. This transformation takes as inputs the parameters of the structural model $\mathbf{A}$ and $\mathbf{B}$, and produces a matrix that has n rows and an arbitrary number of columns. This transformation must satisfy the admissibility condition, so for each orthonormal matrix $\mathbf{R} : \mathbf{F}(\mathbf{RA}, \mathbf{RB}) = \mathbf{RF}(\mathbf{A}, \mathbf{B})$, and the strong regularity condition, so the transformation must be dense. The generalized inclusion and exclusion restrictions are imposed directly on the transformation, and require that a particular entry of $\mathbf{F}(\mathbf{A},\mathbf{B})$ be or be not constrained to zero. Without loss of generality, I assume that $\mathbf{A}$ is always included into the transformation, and placed in the front of other parameters, so $\mathbf{F}(\mathbf{A},\mathbf{B}) = (\mathbf{A}\ \ldots)$. Examples of such transformations are the transformation of the parameters $\mathbf{A}$ and $\mathbf{B}$ into matrix $(\mathbf{A}\ \mathbf{B})$, so $\mathbf{F}(\mathbf{A},\mathbf{B}) = (\mathbf{A}\ \mathbf{B})$, the transformation into the long-run impulse-responses, $\mathbf{F}(\mathbf{A},\mathbf{B}) = (\mathbf{A}\ \mathbf{IR}_\infty)$, or a combination of the previous two transformations, $\mathbf{F}(\mathbf{A},\mathbf{B}) = (\mathbf{A}\ \mathbf{B}\ \mathbf{IR}_\infty)$.

All inclusions and exclusions in this section are summarized by the identification diagram:

Definition 6 (Identification diagram)**.** The identification diagram for transformation $\mathbf{F}(\mathbf{A},\mathbf{B})$ is a directed graph, where each column f_i of $\mathbf{F}(\mathbf{A},\mathbf{B})$ produces node f_i, and there is edge $f_i \to f_j$ if and only if entry $[\mathbf{F}(\mathbf{A},\mathbf{B})]_{ji}$ is not restricted to zero by the identifying restrictions.

- The nodes associated with columns of $\mathbf{A}$ are denoted as the endogenous variables $y_1, y_2, \ldots, y_n$.
- If $\mathbf{B}$ is included into the transformation, the nodes associated with columns of $\mathbf{B}$ are denoted as the primary instruments $z_1, z_2, \ldots, z_m$.
- The nodes associated with the other columns are denoted as $\zeta_1, \zeta_2, \ldots$, and are treated in the same way as the primary instruments.

The conditional causal diagram is a special case of the identification diagram, where $\mathbf{F}(\mathbf{A},\mathbf{B}) = (\mathbf{A}\ \mathbf{B})$. However, the identification diagram may have more nodes, and it does not necessarily have a causal interpretation.

Proposition 3 generalizes Proposition 2 for identification diagrams.

Proposition 3 (Recursive condition for identification with nonlinear restrictions)**.** *Assume that the structural shocks are independent, so $\mathbf{\Sigma}$ is a positive diagonal matrix. If for each parent of y_i in the identification diagram there is an independent primary, recursive or respective identifying path, then y_i is globally identified by the identification diagram in almost all parameter points.*

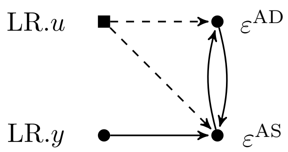
Proof. See Appendix B. □

The literature on long run identification restrictions starts with Blanchard and Quah's (1993) paper. As an application of Proposition 3, in this section I demonstrate how to draw the identification diagram for their identifying restrictions, and how to apply Proposition 3 to verify that their identifying assumptions suffice for the full identification. Let $Y = (u\ y)^T$, where u is the unemployment rate, and y is the log GDP. The structural shocks $\mathcal{E} = \left(\varepsilon^{\mathrm{AD}}\ \varepsilon^{\mathrm{AS}}\right)^T$ are interpreted as the aggregate demand and the aggregate

TABLE 1. Association between entries ir_{ij} of transformation $\mathbf{F}(\mathbf{A},\mathbf{B})$ in (6) and long run restrictions in the Blanchard and Quah's (1993) example.

Structural shock	Aggregate Demand shock	Aggregate Supply shock
Long run response of unemployment	ir_{11}	ir_{12}
Long run response of log GDP	ir_{21}	ir_{22}

FIGURE 4. The identification diagram for the Blanchard and Quah's (1993) model. There is reverse causal interpretations of the edges associated with long-run restrictions: edge $\mathrm{LR}.y \to \varepsilon^{\mathrm{AS}}$ indicates that the long-run response of y to $\varepsilon^{\mathrm{AS}}$ is not zero. For the chosen association between the structural shocks and structural equations, node labels ($\varepsilon^{\mathrm{AD}}$, $\varepsilon^{\mathrm{AS}}$) are synonyms for labels (y, u).



supply shocks. I assume that the aggregate demand shock is associated with the equation defining the unemployment, and the aggregate supply shock is associated with the equation defining the output. This association is arbitrary, but the conclusions do not depend on a particular assumption about the association between the shocks and equations that I use.

Consider the transformation that includes matrix $\mathbf{A}$ and the long-run impulse-response functions $\mathbf{IR}_\infty$:

$$\mathbf{F}(\mathbf{A},\mathbf{B}) = \begin{pmatrix} \mathbf{A} & \mathbf{IR}_\infty \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \mathrm{ir}_{11} & \mathrm{ir}_{12} \\ a_{21} & a_{22} & \mathrm{ir}_{21} & \mathrm{ir}_{22} \end{pmatrix}, \tag{6}$$

The association between entries ir_{ij} of this transformation and the long run restrictions is given in Table 1. Blanchard and Quah (1993) use the identifying assumption that only the aggregate supply shock affects the output in the long run ($\mathrm{ir}_{21} = 0$ and $\mathrm{ir}_{22} \neq 0$). The macroeconomic theory also predicts that the long-run response of u to $\varepsilon^{\mathrm{AD}}$ is zero ($\mathrm{ir}_{11} = 0$), and the response of u to $\varepsilon^{\mathrm{AS}}$ may be negative or zero ($\mathrm{ir}_{12} \leq 0$). No entries of $\mathbf{A}$ are restricted to zero. These identification assumptions are summarized by:

$$\mathbf{F}(\mathbf{A},\mathbf{B}) = \begin{pmatrix} a_{11} & a_{12} & ? & ? \\ a_{21} & a_{22} & 0 & \mathrm{ir}_{22} \end{pmatrix} \tag{7}$$

where the question marks indicate that the respective entries are not constrained to zero in the Blanchard and Quah's paper, but the macroeconomic theory predicts that they may be zero.

These identification constraints are depicted in the identification diagram in Figure 4. Since the aggregate demand shock is associated with the equation for u, and the aggregate supply shock is associated with the equation for y, instead of notation u and y for the respective nodes I use $\varepsilon^{\mathrm{AD}}$ and $\varepsilon^{\mathrm{AS}}$, which simplifies the interpretation of the identification diagram. There is inverse causal interpretation of edges associated with long-run impulse responses: edge LR.$y_i \to \varepsilon_j$ is present in the causal diagram if the long-run response of y_i to structural shock ε_j is not zero. Therefore, edge LR.$y \to \varepsilon^{\mathrm{AS}}$ indicates that there is long-run response of y to $\varepsilon^{\mathrm{AS}}$, and anti-edge LR.$y \to \varepsilon^{\mathrm{AD}}$ indicates that the long-run response of y to the aggregate demand shock is zero. The dashed edges indicate that no particular assumptions have been made on how the unemployment responds to the structural shocks in the long run.

By Proposition 3, this model is fully identified whether or not the dashed edges are present in the identification diagram. For example, if the dashed edges are absent, then the only parent of $\varepsilon^{\mathrm{AD}}$ is $\varepsilon^{\mathrm{A}S}$, for which there exists identifying path LR.$y \to \varepsilon^{\mathrm{A}S}$. Since $\varepsilon^{\mathrm{A}D}$ is identified, it becomes recursive instrument for identification of $\varepsilon^{\mathrm{A}S}$, so the model is fully identified.

5. A note on estimation technique

If the structural model is fully identified using only primary and recursive instruments and the identification diagram is the causal diagram, the structural equations can be estimated one at a time, using, for example, two-stage or three-stage least squares estimator. Consider model (2), and assume that the structural shocks are orthogonal, so Σ is diagonal. By Proposition 2, this model is fully identified. To estimate the parameters of this structural model, start with node y_3, and estimate the parameters in the third structural equation and structural shock ε_3 using the two- or three-stage least square procedure. The estimated structural shock ε_3 processes all properties required for the instruments, so consider this shock as a new primary instrument. Using this instrument, I can identify y_2 and estimate the second structural equation, and ε_2. Finally, use z_1, z_2, ε_2 and ε_3 to identify and estimate the first structural equation and ε_1.

References

Ahelegbey, D. F., M. Billio, and R. Casarin. (2014): "Bayesian Graphical Models for Structural Vector Autoregressive Processes," *Journal of applied Econometrics*, forthcoming.

Blanchard, O. J. and D. Quah (1993): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 83, 655–673.

Brito, C. and J. Pearl (2002b): "A new identification condition for recursive models with correlated errors," *Journal Structural Equation Modeling.*

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.

Demiralp, S., K. Hoover, and S. Perez (2014): "Still puzzling: Evaluating the price puzzle in an empirically identified structural vector autoregression," *Empirical Economics*, 46, 701–731, cited By 0.

Fragetta, M. and G. Melina (2013): "Identification of monetary policy in SVAR models: a data-oriented perspective," *Empirical Economics*, 45, 831–844.

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.

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.

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.

Phiromswad, P. (2014): "Measuring monetary policy with empirically grounded identifying restrictions," *Empirical Economics*, 46, 681–699, cited By 0.

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.

Richardson, T. and P. Spirtes (1999): "Automated discovery of linear feedback models," in *Computation, causation and discovery*, ed. by C. Glymour and G. F. Cooper, Menlo Park: MIT Press.

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.

Tian, J. (2005): "Identifying direct causal effects in linear models," in *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, vol. 20, 346.

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 $\mathcal{Y} = \{y_1, y_2, \dots, y_n\}$ be the set of nodes in the identification diagram associated with the endogenous variables, $\mathcal{Z} = \{z_1, z_2, \dots, z_m\}$ be the set of nodes associated with the exogenous or predetermined variables or nodes $\zeta_1, \zeta_2, \dots$ from Definition 6, 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 identification diagram induced by nodes $\mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{X}_1$, 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 have path $x_1 x_2 x_1 x_4$, I consider instead the path, where cycle $x_1 x_2 x_1$ is removed, so I consider $x_1 x_4$.

Consider matrix $\mathbf{M}$ obtained from transformation $\mathbf{F}$ in the following way: take the rows of matrix $\mathbf{F}$ having the indices of elements of $\mathcal{Y}_1 \cup \mathcal{Y}_2$, and take the columns of $\mathbf{F}$ having the indices of $\mathcal{Y}_2 \cup \mathcal{X}_1$.

If there is path $x_{j_1} x_{j_2} \dots x_{j_s}$ in the identification diagram, the set of parameters associated with this path consists of the following elements of matrix $\mathbf{F}$: $\left\{f_{j_2 j_1}, f_{j_3 j_2}, \dots, f_{j_s j_{s-1}}\right\}$. Therefore, the diagonal elements of $\mathbf{A}$ are not considered as parameters associated with any path. By definition of the identification 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 $\mathcal{L}$ 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 gain intuition, consider the following example. Assume that the structural model is:

$$\begin{pmatrix} 1 & a_{12} & 0 & 0 & a_{15} & a_{16} \\ 0 & 1 & a_{23} & 0 & 0 & 0 \\ 0 & 0 & 1 & a_{34} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & a_{54} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b_{43} & 0 \\ 0 & 0 & 0 & b_{54} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}, \tag{8}$$

which causal diagram is depicted in Figure 5a. 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 5b. Assume the transformation is: $\mathbf{F}(\mathbf{A}, \mathbf{B}) = (\mathbf{A} \ \ \mathbf{B})$. Matrix $\mathbf{F}$ is:

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

a) Causal diagram

b) Subgraph $\mathbb{G}$ induced by $\{y_2, y_3, y_4, y_5, y_6, z_2, z_3, z_4\}$

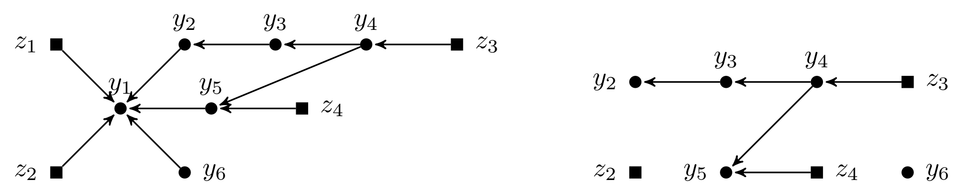


FIGURE 5. Example of causal diagram and subgraph $\mathbb{G}$

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

$$\mathbf{M} = \begin{pmatrix} \underline{a_{23}} & 0 & 0 & 0 & 0 \\ 1 & \underline{a_{34}} & 0 & 0 & 0 \\ 0 & 1 & 0 & \underline{-b_{43}} & 0 \\ 0 & a_{54} & 0 & 0 & \underline{-b_{54}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9}$$

There are two independent paths in $\mathbb{G}$ starting with nodes in $\mathcal{X}_1$ and reaching $\mathcal{Y}_1$, see Figure 5b, 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 (9), and is $[a_{23} \cdot a_{34} \cdot (-b_{43}) \cdot (-b_{54})]$. This permutation has four elements, so $\mathcal{L} = 4$. Finally, there are 2 nodes in set $\mathcal{Y}_2$, so $|\mathcal{Y}_2| = 2$.

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* $\mathcal{Y}_2$*:*

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

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 $\mathcal{L} = |\mathcal{Y}_2|$.

If $\mathbb{G}$ is empty, the only non-zero parameters of $\mathbf{F}$ 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 (8), 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 $|\mathcal{Y}_2|$.

<u>Step 3.</u> Prove that $\mathcal{L} \geq N + |\mathcal{Y}_2|$.

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 is proven in Step 2. Add independent paths from $\mathbb{G}$ into this graph one-by-one. When a new path $x_{j_0} x_{j_1} \ldots x_{j_s}$ is added to the graph, modify the permutation in the following manner:

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

(2) For $k = 1, 2, \ldots, s-1$, remove $f_{j_k j_k}$, and add $f_{j_k j_{k+1}}$. Since $x_{j_k} \in \mathcal{Y}_2$ and $x_{j_{k+1}} \in \mathcal{Y}_1 \cup \mathcal{Y}_2$, parameters $f_{j_k j_k}$ and $f_{j_k j_{k+1}}$ are in $\mathbf{M}$. Since the new path is independent of the previously added paths, $f_{j_k j_{k+1}}$ is located in a different row and in a different column than the permutations associated with the previously added paths, so it can be included into the permutation. Each parameter $f_{j_0 j_1}, f_{j_1 j_2}, \ldots, f_{j_{s-1} j_s}$ 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 $\mathcal{L} \geq N + |\mathcal{Y}_2|$.

In example (8), 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 produces a permutation of length 4.

<u>Step 4.</u> Prove that $N \geq \mathcal{L} - |\mathcal{Y}_2|$

Consider a permutation of length $\mathcal{L}$. Since all parameters associated with one permutation are located in different columns of matrix $\mathbf{F}$, at least $\mathcal{L} - |\mathcal{Y}_2|$ parameters must be located in the columns associated with the indices of $\mathcal{X}_1$. I prove that each such parameter 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 $f_{j_1 j_0}$, 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 $f_{j_1 j_0}$ have been included into the permutation, parameter $f_{j_1 j_1}$, which is normalized to be positive, cannot be included into this permutation, because it is in the same row as $f_{j_1 j_0}$. Therefore, column j_1 either is not included into permutation, or there exists parameter $f_{j_2 j_1}$, 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 $\mathcal{L}$, so consider that parameter instead of $f_{j_1 j_0}$. In the second case, see where the edge associated with $f_{j_2 j_1}$ 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 $\mathcal{L} - |\mathcal{Y}_2|$ 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, $N \geq \mathcal{L} - |\mathcal{Y}_2|$

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

A.2. **Review of the 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:

$$\mathbb{E}(Y|Z) = \mathbf{A}^{-1}\mathbf{B} \cdot Z \equiv \mathbf{\Lambda} \cdot Z \tag{10a}$$

$$\mathrm{Var}(Y - \mathbb{E}(Y|Z)) = \left(\mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A}\right)^{-1} \equiv \mathbf{\Omega} \tag{10b}$$

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 (1) unless $n = 1$. The reason is that there exist many different structural models observationally equivalent to model (1), 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 (10). 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 $\underset{n\times(n+m)}{\hat{\mathbf{P}}} = \begin{pmatrix} \underset{n\times n}{\mathbf{A}} & \underset{n\times m}{-\mathbf{B}} \end{pmatrix}$ are written as:

$$e_i^T \hat{\mathbf{P}} \mathbf{\Psi}_i = 0 \tag{11}$$

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 $\hat{\mathbf{P}}$.

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

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

The constraints on parameters are summarized by:

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

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 $\hat{\mathbf{P}}$ are identified if and only if rank $\left(\hat{\mathbf{P}}\mathbf{\Psi}_i\right) = n-1$, see, for example, Greene (2012). In the considered example (??), all parameters are identified in almost all parameter points, because in almost all parameter points I have:

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

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 $\hat{\mathbf{P}}\mathbf{\Psi}_i$ by deleting the i^{th} row. Since each element in the i^{th} row of $\bar{\mathbf{P}}\mathbf{\Psi}_i$ is constrained to zero by definition of $\mathbf{\Psi}_i$, I have: $\text{rank}\,(\mathbf{M}_i) = \text{rank}\left(\hat{\mathbf{P}}\mathbf{\Psi}_i\right)$.

By definition of $\mathbf{\Psi}_i$, each column of $\hat{\mathbf{P}}\mathbf{\Psi}_i$, 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 $\text{rank}\,(\mathbf{M}_i) = n-1$, 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, and of the theory of partial identificatin.** 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 (10). 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 orthonormal matrix $\mathbf{R}$ such that $\tilde{\mathbf{A}} = \mathbf{R}\mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{R}\mathbf{B}$, where orthonormal matrix $\mathbf{R}$ by definition must satisfy $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. Since the orthonormal 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 the full identification.

Rubio-Ramírez et al. (2010) in Theorem 1 propose the following condition for identification. To verify the identification of parameters located in the i^{th} row of $\hat{\mathbf{P}}$, calculate the rank of matrices $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_i$ composed in the following way:

(12)
$$\mathbf{M}_j = \left(\begin{bmatrix} \\ \mathbf{F}(\mathbf{A}, \mathbf{B})\,\mathbf{\Psi}_j \\ \\ \end{bmatrix} \quad \begin{matrix} \begin{bmatrix} \mathbf{I}_{j\times j} \end{bmatrix} \\ \\ \begin{bmatrix} \mathbf{0}_{(n-j)\times j} \end{bmatrix} \end{matrix} \right)$$

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 $\hat{\mathbf{P}}$ is globally identified in almost all parameter points.

In example (2), to verify the identification of parameters under the assumption of shocks independence, assume $\mathbf{F}(\mathbf{A}, \mathbf{B}) = (\mathbf{A} \;\; \mathbf{B})$, reorder the variables in the reverse order, and calculate the rank of the following matrices:

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -b_{22} & 0 \\ -b_{11} & -b_{12} & 0 \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -b_{11} & 0 & 0 \end{pmatrix} \quad \mathbf{M}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13}$$

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, model (3). 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 (3) are identified.

B.2. **Proof of Propositions 2 and 3.** 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 "$\setminus$" 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 Propositions 2 and 3. 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 $n_1 = |\bar{\mathcal{D}}_i|$. 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 orthonormal matrix $\mathbf{R}$, such that $\tilde{\mathbf{A}} = \mathbf{RA}$, $\tilde{\mathbf{B}} = \mathbf{RB}$, and $\mathbf{R}$ has the following block structure:

$$\mathbf{R} = \begin{pmatrix} \underset{n_1 \times n_1}{\mathbf{R}_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix} \tag{14}$$

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 $\bar{\Phi} \cap \mathcal{D}_i$.

This paragraph of the proof follows the lines of the proof of Theorem 1 in Rubio-Ramírez et al. (2010). Consider matrix $\hat{\mathbf{M}}_i$ obtained from $\mathbf{F\Psi}_i$ 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 (12) is also not full. Indeed, if y_i is not identified then there must exist orthonormal matrix $\mathbf{R}$, having the following properties. First, because of its special structure given by (14), and because nodes $y_{n_1+1}, y_{n_1+2}, \ldots, y_{i-1}$ are identified, $\mathbf{R}$ has the following structure:

$$\mathbf{R} = \begin{pmatrix} \underset{(i-1) \times (i-1)}{\mathbf{I}} & \mathbf{0} \\ \mathbf{0} & \underset{(n-i) \times (n-i)}{\mathbf{R}_{33}} \end{pmatrix} \tag{15}$$

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 $e_i \mathbf{F\Psi}_i = 0$ and

$e_i \mathbf{RF\Psi}_i = \mathbf{0}$, so $e_i (\mathbf{R} - \mathbf{I}) \mathbf{F\Psi}_i = 0$. 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.

□

NIKOLAY AREFIEV
The National Research University Higher School of Economics. E-mail: n.are_ev@gmail.com.
Address: o_ce 708A, Shabolovka
31, 119049, Moscow, Russia.

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