NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nikolay Arefiev

A THEORY OF DATA-ORIENTED IDENTIFICATION WITH A SVAR APPLICATION

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 79/EC/2014

A THEORY OF DATA-ORIENTED IDENTIFICATION WITH A SVAR APPLICATION

NIKOLAY AREFIEV

ABSTRACT. I propose a method identification of structural vector autoregressions (SVARs) and simultaneous equations models (SEMs) with orthogonal structural shocks using testable identification restrictions. If some sparsity conditions are satisfied, the method produces a set of testable inclusions and exclusions, sufficient for the full identification. The method stems from the theory of probabilistic graphical models and from the theory of identification of SVARs and SEMs, merging them into a unified approach. In the application example, I estimate a SVAR monetary model of the US economy with 6 variables, where all but one identifying restrictions are testable. The method produces relatively narrow confidence intervals for the impulse-response functions, does not generate any anomalies such as the price puzzle, and reveals importance of informational channels through which news about structural shocks spread throughout the economy.

Keywords: Identification, data-oriented identification, sparse structural models, structural vector autoregression, SVAR, simultaneous equations model, SEM, probabilistic graphical model, PGM, price puzzle, information theory in macroeconomics.

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,

*February 10, 2015.
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). I thank Alina Arefeva, Antoine d'Autume, Svetlana Bryzgalova, Jean-Bernard Chatelain, Boris Demishev, 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. 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.

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, when it is possible to measure some causal effects using only testable identification restrictions?

The paper proposes a method of testable identification of Gaussian linear structural vector autoregressions (SVARs) and simultaneous equations models (SEMs) with orthogonal structural shocks. If the matrix multiplying the vector of exogenous or predetermined variables in the structural model is sufficiently sparse but not degenerate, the method produces a set of inclusion and exclusion restrictions, satisfying the following two properties. First, each restriction from this set can be tested either as the null or as the alternative hypothesis. Second, taken together, these restrictions suffice for the full identification of the structural model.

The developed approach can be applied to a large variety of econometric problems. The existence of a set of testable restrictions sufficient for the full identification is not guaranteed, but would not be surprising in many macroeconomic applications. In Section 7 I briefly discuss how this method can be applied to dynamic stochastic general equilibrium models (Galí, 1999; Smets and Wouters, 2003, 2007; Christiano et al., 2005, and many others). In Section 8 I consider a detailed application example, where I estimate a SVAR monetary model of the US economy (following Sims, 1980, 1986, 1992; Blanchard and Quah, 1993; Christiano et al., 1999; Zha, 1999; Hanson, 2004; Uhlig, 2005; Sims and Zha, 2006, and many others). The scope of applications is not limited to these areas.

The method is based on the following results, derived in this paper. First, I propose graphical interpretations of the rank condition for identification of SEMs, of the sufficient condition for identification of structural vector autoregression (SVAR) models (Rubio-Ramírez et al. (2010)), and of the theory of partial identification (reviewed in Christiano et al. (1999)). An example illustrating 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, the proposition is formulated in Section 4.2 and proven in Appendix C. Third, I introduce *conditional partial correlations.* For a given pair of variables, the conditional partial correlation is zero in almost all parameter points if and only if there is no equation in the structural model such that both variables are present in this equation. The difference between conditional and conventional partial correlations is that for conventional partial correlations this property holds only for pairs, where at least one variable is endogenous, and for conditional partial correlations this property also holds for pairs of exogenous or predetermined variables. Finally, in Section 5 I show how the problem of identification in econometrics is related to the *clique cover problem* in computer science. This relation turns out to be helpful for finding identification restrictions, which are both consisted with the data and supported by the theory. In Section 6 I assess the power of the proposed tests.

In Section 8 I propose a detailed application example, where I estimate a SVAR monetary model of the US economy with 6 variables. I include the federal interest rate, the unemployment rate, the capacity utilization rate, the GDP growth rate, the GDP deflator inflation rate, and the commodity price inflation rate. I use these 6 variables, because I have not found a smaller model consistent with the assumption that the structural shocks are orthogonal to each other. This is possible to estimate this model using only testable identification restrictions, however, testable restrictions cannot distinguish between the aggregate demand (AD) and aggregate supply (AS) shocks, so some structural shocks in the model identified using only testable restrictions are difficult to interpret. For this reason, I introduce the only non-testable assumption that the commodity price inflation directly influences the AS but not the AD curve. This and the testable assumptions suffice for the full identification. I obtain impulse response functions, which are better than those usually obtained in the literature on SVARs in the following two respects. First, I get narrower confidence intervals for the estimated impulse-response functions than in classical papers. Second, all 36 impulse response functions that I get are consistent with the theory, producing no anomalies such as the price puzzle. In addition, the method reveals the importance of informational channels, by the mean of which news about structural shocks spread throughout the economy (Lucas, 1972; Sims, 1998; Woodford, 2001; Mankiw and Reis, 2002; Sims, 2003; Maćkowiak and Wiederholt, 2009; Veldkamp, 2011).

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, also known as acyclic or triangular models. The graphical interpretations of the various conditions for identifying 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)). I make the following two contributions to this literature. First, 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. Using only this advancement, however, I would be able to achieve testable identification only for triangular models, but not for cyclical models. The second contribution, which makes it possible to achieve testable identification for cyclical models, is the introduction of the conditional partial correlations.

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 hereafter 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 regressions, 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

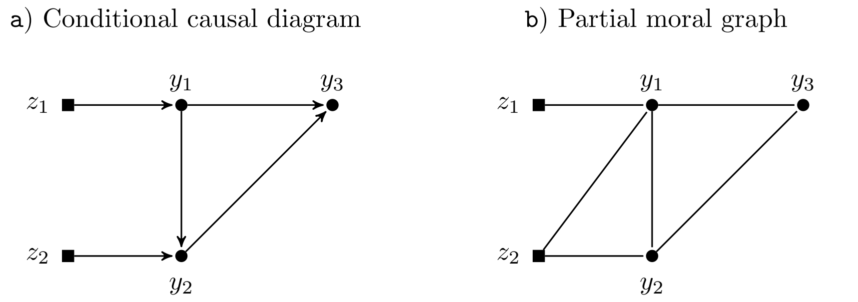

FIGURE 1. Conditional causal diagram and partial moral graph for model (1).

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, 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 the instruments, so the random process generating (z_1, z_2) 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.

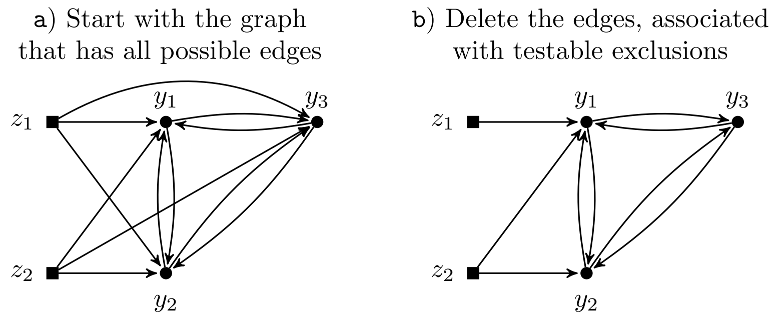

FIGURE 2. Drawing the map of testable exclusion restrictions produced by the partial moral graph.

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 full moral graphs (Koller (2009); Pearl (2009)). However, I show that this result also holds for some pairs of nodes in partial moral graphs, 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

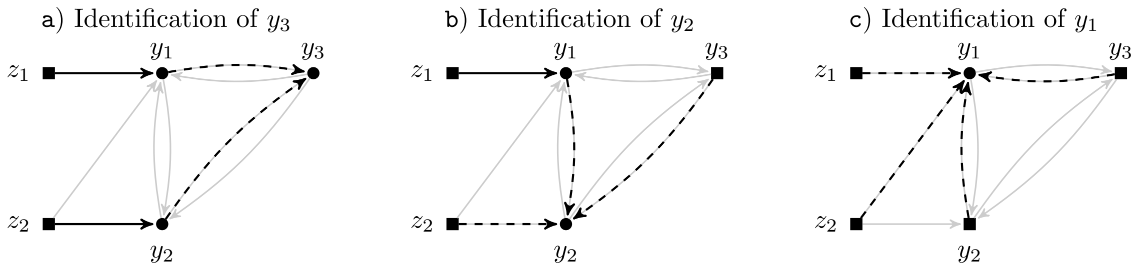

FIGURE 3. Graphical sufficient condition for identification

the respective edge from the map of exclusions. I therefore produced the map of exclusions depicted in Figure 2b.

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 path 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

TABLE 1. Testable identification restrictions for model (1)

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

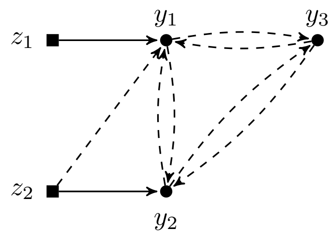

FIGURE 4. Summary of testable inclusion and exclusion restrictions.

minus the number of rows of $\mathbf{\Pi}\left(y_1, y_2|z_1, z_2\right)$ plus 1, which in the considered example is 1. Finally, I calculate the df smallest eigenvalues $\lambda_1, \ldots, \lambda_{\text{df}}$ of product $\mathbf{\Pi}\left(y_1, y_2|z_1, z_2\right) \times \mathbf{\Pi}\left(z_1, z_2|y_1, y_2\right)$ and calculate the 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 $\text{rank}\left(\mathbf{\Pi}\left(y_1, y_2|z_1, z_2\right)\right) < 2$, 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:

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

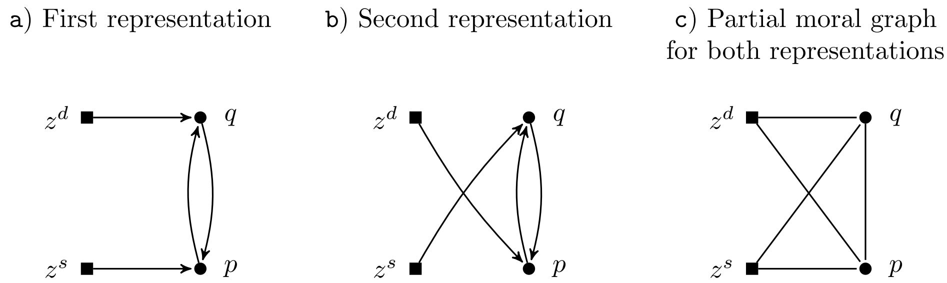

FIGURE 5. Multiple causal representations of cyclical model (7) and the unique partial moral graph.

with the estimated model:

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

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.** The second example presented in this section demonstrates how a testable identification can be achieved for cyclical models. 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, 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 parameters to be estimated.

One difficulty with cyclical models is that the same model has several SEM and causal representations. Model (7) has two representations, where the demand and supply equations are identified. The first representation 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. In this example, it is not possible to justify that one representation is better than the other, neither from the economic theory, nor from any empirical tests.

To draw the partial moral graph, moralize and disorient any causal representation of model (7) in Figure 5a or 5b. Alternatively, draw the vertices, which are the random variables of the structural model, and for each equation add the *clique* of the variables included into this equation, where a clique is defined as a set of pairwise adjacent nodes. In model (7), for example, equation (7a) includes variables p, q, and z^d, so it produces clique $\{p, q, z^d\}$ in the partial moral graph in Figure 5c, and equation (7b) includes p, q, and z^s, producing clique $\{p, q, z^s\}$.

The algorithm of testable identification proposed in the previous example does not produce any testable exclusion restriction here, so I use a different approach. Consider the *clique cover problem* for the partial moral graph. The *clique cover problem* for an undirected graph is to find as few cliques as possible to cover the entire graph. For a model with n equations, if the clique cover problem has a unique solution with n cliques, then each clique is associated with a structural equation in such way that the variables included into the clique are included into this structural equation, and the variables included into the equation are included into the clique. For the moral graph depicted in 5c, the unique solution for the clique cover problem is $\{z^d, q, p\}$ and $\{z^s, q, p\}$. I can learn asymptotically the true moral graph from data in almost all parameter points, and having solved the clique cover problem for the estimated moral graph, I can conclude which variables are present in each equation. In this example, this suffices for the full testable identification.

To estimate the moral graph in Figure 5c, I need to test the hypothesis that two instruments, z^d and z^s, are not adjacent in this graph. Unlike in the previous example, I cannot use conventional partial correlations to test this hypothesis, because they can be used only for pairs of variables, where at least one variable is endogenous. In Section 4.1, I introduce a new kind of partial correlations, referred to hereafter as conditional partial correlations, and these correlations can also be applied to pairs of instruments. This is how I can test the hypothesis that z^d and z^s are not adjacent in the conditional moral graph.

In contrast to 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 tests of null hypotheses that there is no moralization effect between instruments. As I discuss in Section 6, indirect tests may 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 5 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 Gaussian 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

$$\mathbf{P} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \qquad X = \begin{pmatrix} Y \\ Z \end{pmatrix} \qquad \mathcal{E}_X = \begin{pmatrix} \mathcal{E} \\ \mathcal{E}_Z \end{pmatrix} \tag{10}$$

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} rows 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 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. □

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 $\boldsymbol{\Sigma}$ 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).

Proposition 2 (Recursive condition for identification)**.** *Assume that the structural shocks are independent, so $\boldsymbol{\Sigma}$ 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.** Assume that the structural model is orthogonal, which means that the covariance matrix for the residuals is diagonal. 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:

$$\begin{aligned}\mathbf{C} &= \mathbf{P}^T\mathbf{P} \\ &= \begin{pmatrix}\mathbf{A}^T\mathbf{A} & -\mathbf{A}^T\mathbf{B} \\ -\mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B}+\mathbf{S}^T\mathbf{S}\end{pmatrix}\end{aligned} \tag{11}$$

The concentration matrix is useful for testable identification because it can be estimated from the data without any prior identification assumptions, and it gives estimators for $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{B}$.

The right-bottom block of $\mathbf{C}$ is not very helpful for testable identification, because it includes term $\mathbf{S}^T\mathbf{S}$, which is not of interest of the analysis. It would be more useful to have an estimator of $\mathbf{B}^T\mathbf{B}$ instead of $\mathbf{B}^T\mathbf{B}+\mathbf{S}^T\mathbf{S}$. However, the value of $\mathbf{S}^T\mathbf{S} = \left(\mathbb{E}\left(ZZ^T\right)\right)^{-1}$ can be estimated separately and subtracted from this block. I then obtain matrix $\hat{\mathbf{C}}$ referred to hereafter as the *partial concentration matrix*:

$$\begin{aligned}\hat{\mathbf{C}} &= \mathbf{C} - \begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times m} \\ \mathbf{0}_{m\times n} & \left(\mathbb{E}\left(ZZ^T\right)\right)^{-1}\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{A}^T\mathbf{A} & -\mathbf{A}^T\mathbf{B} \\ -\mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B}\end{pmatrix}\end{aligned} \tag{12}$$

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 $\hat{c}_{ij}$ is not zero, where $\hat{c}_{ij}$ is the respective element of the partial concentration matrix.

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 its relationship to the partial concentration network, I define relatives and strangers:

Definition 7 (Relatives and strangers)**.** Vertices x_i and x_j are *relatives* in the conditional causal diagram 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 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.

To explain the relationship between the partial moral graph and the partial concentration network, I make a generic 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 the* ***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. 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.

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.*

- *Assume Assumption 1 is satisfied. If an edge is absent in the partial concentration network, this edge is also absent in the partial moral graph.*

Proof. 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. For Corollary 1, consider two nodes in the partial moral graph, node x_i is a variable in Y or Z, and node y_i is 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 in almost all parameter points.*

For Corollary 2, 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 almost all parameter points.*

Testable exclusion restrictions can be tested using the concentration matrix, or, alternatively, using 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 against $X_{(-i,-j)}$. Knowing the matrix of concentration, the partial correlations can be calculated using:

$$\rho_{ij} \equiv \mathrm{corr}(x_i, x_j | X_{(-i,-j)}) = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}}, \tag{13}$$

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 $X_{(-i,-j)}$. One of implications of (13) is that for endogenous variables $\text{corr}(y_i, y_i|X_{-i}) = -1$. Partial correlations are helpful in formulating directly testable exclusion restrictions, like I did in Section 2.1.

To test indirectly testable exclusions, I introduce conditional partial correlations in the following way. For $i \neq j$ and $i, j > n$, $\hat{\rho}_{ij}$ for the true model is defined as:

$$\hat{\rho}_{ij} = \frac{-\hat{c}_{ij}}{\sqrt{c_{ii}c_{jj}}} \tag{14}$$

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 $\hat{\rho}_{34} = 0$.

The estimation of $\hat{\rho}$, however, is not straightforward, because a direct use of (12) and (14) would produce a biased estimator for $\hat{\rho}$. In Section 8 I use bootstrap to get an unbiased estimator for conditional partial correllations.

Proposition 4 (Relevance of an instrument). *Let j be the column in $\hat{\mathbf{P}}$ associated with instrument z_i. Condition $\hat{c}_{jj} = 0$ holds if and only if z_i is irrelevant, so for each $k = 1, 2, \ldots, n : b_{ki} = 0$, where $\hat{c}_{jj}$ is the respective elements of the partial concentration matrix.*

Proof. By definition of $\hat{\mathbf{C}}$, I have: $\hat{c}_{jj} = \sum_{k=1}^{n} b_{ki}^2$. Therefore, $\hat{c}_{jj}$ is zero if and only if for each k: $b_{ki} = 0$, in which case the instrument is irrelevant. □

The relevance of an instrument can be tested using the partial concentration matrix, or the conditional partial correlation $\hat{\rho}_{ii}$ for $i > n$ defined by (14).

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

$$\mathbb{E}\left(\mathcal{P}_i|Z\right) = \mathbf{\Pi}_i(\mathcal{P}_i|Z)Z \tag{15}$$

where $\mathcal{P}_i$ are the parents of y_i on the map of exclusion restrictions. The constant term is omitted in (15), because all variables have been centralized, so the term is zero.

Proposition 5 (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 5, and shows that if a node is not identified, then the condition formulated in Proposition 5 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 5 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. Identification and the Clique Cover Problem

In this section, I show how to reduce the problem of formulating testable exclusion and inclusion restrictions to the *clique cover problem* explored in the computer science literature. A *clique* in an undirected graph is a set of nodes such that every two nodes are adjacent. This implies that the subgraph spanned by the nodes of a clique is the full graph. In Figure 1b, for example, sets $\{z_2, y_1, y_2\}$ and $\{z_2, y_1\}$ are examples of cliques, but $\{z_2, y_1, y_2, y_3\}$, is not a clique, because z_2 and y_3 are not adjacent. The *clique cover problem* is to find as few cliques as possible that include all nodes and cover all edges 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.

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

Proposition 6 (Structural equations and cliques)**.** *For each structural equation, consider the clique composed of the variables included into this structural equation. The partial moral graph is the graph sum over all cliques defined in this way.*

Proof. By Definition 7, two variables are relatives in the partial moral graph if and only if there exists an equation in the structural model, where both variables are included. Therefore, each equation produces a clique in the partial moral graph, such that all variables present in the structural equation are included into the clique. By Definition 8, the partial moral graph is the graph sum over all cliques defined in this way. □

Proposition 6 produces the following three results:

Proposition 7 (Existence of a solution with no more than n cliques)**.** *The clique cover problem for the partial moral graph can be solved with* n *or less cliques.*

Proof. By Proposition 6, the partial moral graph is a graph sum over n cliques. Therefore, there exist n cliques covering the entire graph. These cliques either form a solution to the clique cover problem, or the solution consists of less than n cliques. □

Proposition 8 (A sufficient condition for testable identification)**.** *If the clique cover problem has a unique solution with* n *cliques, there is a one-to-one association between the cliques solving the clique cover problem and the structural equations, such that each clique consists of the variables present in the associated structural equation. In this case, if the true structural model is identified using its inclusion and exclusion restrictions, then the estimated model is identified using only testable restrictions in almost all parameter points.*

Proof. The first part is straightforward from Proposition 6. The sufficiency for identification follows from the fact that if the clique cover problem for the true moral graph has a unique solution with n cliques, all inclusion and exclusion restrictions by Proposition 6 are testable in almost all parameter points, so if they suffice for the identification of the true model, they also suffice for the identification of the estimated model. □

Proposition 9 (A necessary condition for testable identification)**.** *If the clique cover problem can be solved with less than* n *cliques, then the partial moral graph does not suffice for the full or partial identification of the structural model.*

Proof. Since the clique cover problem can be solved with less than n cliques, at least one clique is associated with two or more structural equations. There is no testable restrictions, which distinct one equation from another in this clique, so neither equation is identified. The partial moral graph does not suffice to conclude which cliques are associated with multiple equations, so no equations in the entire model are identified. □

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 includes variables z_1 and y_1, the second includes z_2, y_1, and y_2, and the third includes y_1, y_2, and y_3. The partial moral graph does not indicate 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 $\{z^s, p, q\}$,

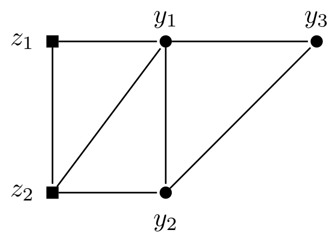

FIGURE 6. Partial moral graph for model (16), where the solution to the clique cover problem is not unique, but testable identification is possible.

which is consistent with the only structural model (7), but it does not indicate whether the representation in part `a` or `b` of Figure 5 is correct.

The condition formulated in Proposition 8 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 into the structural equation for y_1:

(16a) $$y_1 = c_1 + b_{11}z_1 + b_{12}z_2 + \varepsilon_1$$

(16b) $$y_2 = c_2 + a_{21}y_1 + b_{22}z_2 + \varepsilon_2$$

(16c) $$y_3 = c_3 + a_{31}y_1 + a_{32}y_2 + \varepsilon_3$$

The partial moral graph for model (16) is drawn in Figure 6. The clique cover problem has two solutions with 3 cliques. 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 in the second structural equation. This restriction, nevertheless, is testable. 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.

To apply Proposition 8 in practical problems, I may need to solve the NP-hard clique cover problem. Proposition 10 below provides another criteria for the existence of testable identification. On the one hand, this criteria is less general than those in Proposition 8. On the other hand, it does not require drawing the moral graph and solving NP-hard problems. It uses the following definition:

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 10 (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 conclude 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. If two given markers are associated with different equations, they never appear in the same equation, so they are not adjacent in the moral graph. Therefore, markets associated with different equations never appear in the same clique, and this guarantees that the clique cover problem cannot be solved in less than n cliques. By Proposition 7, the solution has no more than 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 this defines the clique associated with each marker in a unique way. Therefore, the solution to the problem is unique. By Proposition 8, it is possible to identify the markers, and to say which variables are included into the structural equation associated with any given marker. □

6. Tests power and multiple hypotheses testing problem

This section answers two questions. First, how to predict the power of each individual test having some prior information about the true structural model? Second, how to aggregate efficiently the results of individual tests, assessing the multiple hypothesis testing problem?

6.1. Power of individual tests.

I use *Pythagorean weights* of variables in equations to predict the power of individual tests. These weights are defined by:

Definition 10 (Pythagorean weight of x_j in equation i)**.** The Pythagorean weight of x_j in the i^{th} structural equation is:

$$w_i(x_j) = p_{ij} \left(\sum_{i=1}^{n+m} p_{ij}^2 \right)^{-\frac{1}{2}}$$

where p_{ij} is the respective element of matrix $\mathbf{P}$, see (10).

By Definition 10, for each i and j: $-1 \leq w_i(x_j) \leq 1$ and $\sum_{k=1}^{n+m} \rho_k(x_j)^2 = 1$.

The power of each individual test depends on the respective partial correlation, which can be obtained from Pythagorean weights using:

Proposition 11 (Pythagorean weights and partial correlations)**.** *Partial correlation* $\rho_{ij} = \text{corr}\left(x_i, x_j | X_{(-i,-j)}\right)$ *equals to the negative sum of products of the Pythagorean weights for* x_i *and* x_j *over all equations:*

$$\rho_{ij} = -\sum_{k=1}^{n+m} w_k(x_i) \cdot w_k(x_j)$$

Proof. This result follows from (11), (13), and Definition 10. □

I can predict the power of each test using Proposition 11 and the result that partial correlations follow Wishart distribution. Alternatively, I can use Fisher's (1924) approximation of the distribution function for partial correlations, which uses z-transform of partial correlations defined by:

$$\zeta(\rho) = \text{artanh}(\rho) = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \tag{17}$$

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. If, for example, $\rho = 0.17$, $n + m = 5$, and $T = 100$, then then $\zeta(0.17) \approx 0.17$, and $\zeta \cdot \sqrt{T - (n+m) + 3} \approx 1.7$, so I expect to correctly reject the null hypothesis at the significance level of 10% in about 50% of experiments.

Proposition 11 produces the following five observations. First, the weight of any endogenous marker is either 1 or -1 in the marked equation and 0 elsewhere. The absolute values of the partial correlations of the other variables with the marker equal to the absolute values of their weights in the marked equation. Therefore, the partial correlations between the endogenous marker and the other variables in the structural equation may be strong. Second, the weights of instruments are proportional to their relevance. Indeed, the weight of z_i in the k^{th} equation can be developed as:

$$w_k(z_i) = \hat{w}_k(z_i) \cdot \hat{\rho}_{ii} \tag{18}$$

where $\hat{\rho}_{ii}$ is defined by (14), and $\hat{w}_k(z_i)$ is the Pythagorean weight of z_i in the structural equations:

$$\hat{w}_k(z_i) = \frac{b_{ki}}{\sum_{\hat{k}=1}^{n} b_{\hat{k}i}^2} \tag{19}$$

The relevance of each instrument, $\hat{\rho}_{ii}$, can be estimated separately before identification. If the relevance is high, the produced partial correlations may be high, but instruments with low relevance produce weak partial correlations. Third, the weight of exogenous marker z_i is $\hat{\rho}_{ii}$ in the marked equation and zero elsewhere. The partial correlations of the other variables with the marker equal to the product of their weights by $\hat{\rho}_{ii}$. These partial correlations may be strong when the relevance is high, but they are low, when the relevance is low. Fourth, variables present in many equations have low average weights, so produce low partial correlations on average. Finally, if two variables are present in many equations, effects produced by two different equations onto the partial correlation may complement or partially offset each other depending on the signs of the weights (see Assumption 1).

Consider example (1). Variable y_3 is an endogenous marker for the third equation, so its relative weight in the third equation is 1 or -1, and $|\rho(y_1, y_3)| = |w_3(y_1)|$, and $|\rho(y_2, y_3)| = |w_3(y_2)|$. These partial correlations may be strong. Variables z_1 and z_2 are exogenous markers for the first and second equations, so their relative weights are $\hat{\rho}_{44}$ and $\hat{\rho}_{55}$, which can be estimated before identification. If the relevance in high, partial correlations $\rho(z_1, y_1)$, $\rho(z_2, y_1)$, and $\rho(z_2, y_2)$ may be strong, but if the relevance is low, these partial correlations are weak. Finally, $\rho(y_1, y_2) = -w_1(y_1)w_1(y_2) - w_2(y_1)w_2(y_2)$. If these products have opposite signs, they partially offset each other in the equilibrium, and if they have the same sign, they complement each other.

6.2. **Multiple hypothesis testing.** There are $N = \left((n+m)^2 + m - n\right)/2$ individual hypotheses, so a multiple hypotheses testing procedure is required. A natural benchmark is the control of the *expected false discovery rate.* Let *total discoveries* be the number of rejected null hypotheses, and *false discoveries* be the number of wrongly rejected null hypotheses. The false discovery rate is defined as the ratio of the false discoveries to the total discoveries when the number of total discoveries is positive, and defined to be zero when the number of total discoveries is zero. Benjamini and Hochberg (1995) prove that the following procedure controls the expected false discovery rate below or at level q^*: estimate individual p-values $p_1, p_2, \ldots, p_N$, sort p-values in increasing order $p_1 \leq p_2, \leq \ldots, \leq p_N$, find the largest k for which $p_k < \frac{k}{N} q^*$, reject the null hypotheses associated with $p_1, \ldots, p_k$, and accept the null hypotheses associated with $p_{k+1}, \ldots, p_N$. q-value associated with i^{th} null hypothesis is defined as the minimal value of q^* for which i^{th} null hypothesis is rejected.

Benjamini and Hochberg (1995) assume independence of individual null hypotheses, but this assumption may be not satisfied when a concentration network is estimated. Assume the model has been generated by T independent realizations of $\mathcal{E}_X$, written into $(n+m) \times T$ matrix $\hat{\mathcal{E}}$, and $\mathbf{E}$ is defined by $\mathbf{E} = \left(\hat{\mathcal{E}}_X \hat{\mathcal{E}}_X'\right)^{-1}$. Each element of the empirical partial concentration matrix can be expressed as $\hat{c}_{ij} = \sum_{k,l} p_{ik} \cdot p_{jl} \cdot e_{kl}$, where e_{kl} is the respective element of $\mathbf{E}$. If the null hypothesis is correct for some $\hat{c}_{i_1 j_1}$ and $\hat{c}_{i_2 j_2}$, then $\mathbb{E}\left(\hat{c}_{i_1 j_1}\right) = \mathbb{E}\left(\hat{c}_{i_2 j_2}\right) = 0$. However, if for a particular realization the linear combination of e_{kl} defining $\hat{c}_{i_1 j_1}$ is outside the confidence interval, this may be more likely that another linear combination of e_{kl}, defining $\hat{c}_{i_2 j_2}$, is also outside of the confidence interval. Therefore, the test statistics may be positively dependent.

Benjamini and Yekutieli (2001) prove that the procedure described above correctly controls the expected false discovery rate at the level below or equal to q^* in the case of positive dependency.

7. Testable identification of policy functions of DSGE models

A typical dynamic stochastic general equilibrium (DSGE) model can be reduced to the following system of equations:

$$\mathbf{A}Y_t = \mathbf{B}_L Y_{t-1} + \mathbf{B}_F \mathbb{E}_t \tilde{Y}_{t+1} + \mathbf{W}\mathcal{E}_t + \mathbf{W}_L \mathcal{E}_{t-1} \tag{20}$$

where Y is vector of the state variables of the DSGE model, $\tilde{Y}$ is vector of all variables, including the state and the forward-looking variables, $\mathbf{A}$, $\mathbf{B}_L$, $\mathbf{B}_F$, $\mathbf{W}$, and $\mathbf{W}_L$ are matrices of parameters. Matrices $\mathbf{W}$ and $\mathbf{W}_L$ describe the moving average component for the structural shocks. I assume that $\mathbf{W}$ and $\mathbf{W}_L$ are diagonal, which is a conventional assumption for DSGE models. In this section I use normalization rule $\mathbf{W}_L = \mathbf{I}$ and for each $i : a_{ii} > 0$, where a_{ii} is the respective element of matrix $\mathbf{A}$.

Model (20) has a different structure than (8), so I need a new strategy of finding testable identification restrictions. I can use the following algorithm. First, use the result that the value of $\mathbb{E}_t \tilde{Y}_{t+1}$ can be represented as a function of the state, and exclude the expected values from (20). Assume that matrices $\mathbf{B}_0$ and $\mathbf{B}_1$ are such that:

$$\mathbb{E}_t \tilde{Y}_{t+1} = \mathbf{B}_0 Y_t + \mathbf{B}_1 \mathcal{E}_t, \tag{21}$$

so (20) can be represented in the following structural ARMA form:

$$\bar{\mathbf{A}} Y_t = \mathbf{B}_L Y_{t-1} + \bar{\mathbf{W}}\mathcal{E}_t + \mathcal{E}_{t-1} \tag{22}$$

where $\bar{\mathbf{A}} = (\mathbf{A} - \mathbf{B}_0)$, and $\bar{\mathbf{B}} = (\mathbf{W} + \mathbf{B}_1)$.

Second, estimate VARMA(1, 1) in the following reduced form:

$$Y_t = \mathbf{M}_1 Y_{t-1} + \mathbf{M}_2 u_t + u_{t-1} \tag{23}$$

Let $\mathbf{\Omega} = \mathbb{E}(u_{t-1} u'_{t-1})$ be the covariance matrix for the lagged residuals. Third, use Cholesky decomposition of $\mathbf{\Omega}^{-1}$ to estimate an observationally equivalent structural model. If $\tilde{\mathbf{A}} = \text{Chol}(\mathbf{\Omega}^{-1})$, then the estimated observationally equivalent model is:

$$\tilde{\mathbf{A}}Y_t = \tilde{\mathbf{B}}_L Y_{t-1} + \tilde{\mathbf{W}}\tilde{\mathcal{E}}_t + \tilde{\mathcal{E}}_{t-1} \tag{24}$$

where $\tilde{\mathbf{B}}_L = \tilde{\mathbf{A}}\mathbf{M}_1$, and $\tilde{\mathbf{W}} = \tilde{\mathbf{A}}\mathbf{M}_2$. Finally, use the result that for all observationally equivalent models, $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} = \bar{\mathbf{A}}^T\bar{\mathbf{A}}$, $\tilde{\mathbf{A}}^T\tilde{\mathbf{B}}_L = \bar{\mathbf{A}}^T\bar{\mathbf{B}}_L$, and so on. Therefor, I can estimate $\bar{\mathbf{A}}^T\bar{\mathbf{A}}$, $\bar{\mathbf{A}}^T\bar{\mathbf{B}}_L$, $\bar{\mathbf{A}}^T\bar{\mathbf{W}}$, $\bar{\mathbf{B}}_L^T\bar{\mathbf{B}}_L$, $\bar{\mathbf{B}}_L^T\bar{\mathbf{W}}$, and $\bar{\mathbf{W}}^T\bar{\mathbf{W}}$ before identification, and use the obtained estimators for verifying testable inclusion and exclusion restrictions.

Is it likely that a typical DSGE model produces a set of testable restrictions sufficient for the full identification? To answer this question, observe that matrix $\mathbf{B}_L$ in many DSGE models is diagonal or close to diagonal. Indeed, contemporaneous values of state variables in DSGE models usually depend on their own lagged values, but not on the lagged values of the other variables. For example, habit formation assumption produces a dynamic equation, where the contemporaneous consumption depends on the lagged consumption, but not on the lagged capital or other lagged variables. If $\bar{\mathbf{B}}_L$ is exactly diagonal, then it gives exactly one marker for each structural equation, so by Proposition 10, model (22) can be identified using only testable restrictions. If there are a few off-diagonal elements in matrix $\bar{\mathbf{B}}_L$, then I need some zeros in $\bar{\mathbf{A}}$ or $\bar{\mathbf{W}}$ to achieve testable identification.

8. Application Example

8.1. **Framework: a SVAR model.** In this section, I use US quarterly data from 1967:Q1 to 2007:Q4 to estimate a SVAR model. I do not include later data, because the monetary policy rule has been modified during the crisis of 2008. There are 6 variables in the estimated model, the federal interest rate r, the inflation rate π measured as the GDP deflator growth rate, the commodity price inflation rate π^c, the GDP growth rate g, the capacity utilization rate c, and the unemployment rate u. The data on π, π^c, g, c, and u are seasonally adjusted, and the data on r, π, π^c and g has been annualized.

Variables r, π, and g are included because they are of the primary interest of the analysis. Including the other variables increases the complexity of the model and the identification procedure, however, I have not found it possible to identify a structural model with less than 6 variables. The reason to include the capacity utilization rate is that it helps to distinguish between the aggregate demand and the aggregate supply shocks. Roughly speaking, an increase in the GDP growth rate keeping constant the capacity utilization rate is interpreted as a positive aggregate supply shock, and an increase in the capacity utilization rate keeping constant the GDP growth rate is interpreted as a simultaneous positive aggregate demand shock and a negative aggregate supply shock; the robust identification procedure in Section 8.4 uses a more sophisticated assumption to distinguish between the AD and AS shocks. If I do not distinguish the AD and AS shocks, I mix them up in the estimated model, which may produce the price puzzle, because they may have opposite effects on the inflation. This is also important to include the unemployment rate, and, following Sims (1992), the commodity pricy inflation rate, because they are confounders to other variables of the structural model, and omitting them would produce biased impulse-response functions, including the price puzzle (see Sims, 1992; Clarida et al., 1998; Christiano et al., 1999; Galí et al., 2003; Hanson, 2004), where a restrictive monetary policy shock significantly increases the inflation rate in the short run.

The model is estimated in the following structural form:

$$\mathbf{A}Y_t = \mathbf{B}_1 Y_{t-1} + \mathbf{B}_2 Y_{t-2} + \mathcal{E}_t \tag{25}$$

where $Y = (r\ u\ c\ g\ \pi\ \pi^c)^T$, and the constant term is zero because all variables have been centralized. I keep the notation of model (8), using $Y_t \equiv Y$, $(Y_{t-1}^T\ Y_{t-2}^T)^T \equiv Z$, and $(Y^T\ Z^T)^T \equiv X$. The structural shocks are assumably independent, so $\mathbb{E}\left(\mathcal{E}\mathcal{E}^T\right)$ is diagonal. In this section the main diagonal of $\mathbf{A}$ is normalized to ones.

I include 2 lags for the following reasons. If I use only one lag, the estimated model produces the price puzzle, suggesting that the state of the economy is not well described with this specification, in which case I cannot use the assumption that the structural shocks are independent. It is helpless to include more than 4 lags for testable identification, because I have not found any significant partial correlation between the contemporaneous variables and the variables lagged for more than 4 quarters. Models with 2, 3, and 4 lags produce similar results, so I choose 2 lags to ease the identification procedure.

8.2. **Estimation of the partial concentration network.** To estimate the partial concentration network and to conclude which partial correlations are significant, I use block bootstrap to test individual hypotheses, and I control the expected false discovery rate for multiple hypotheses testing. Each individual null hypothesis is that $\hat{c}_{ij} = 0$, where $\hat{c}_{ij}$ is the respective element of matrix $\hat{\mathbf{C}}$ defined by (12). The alternative is that $\hat{c}_{ij} \neq 0$ for $i \neq j$, and $\hat{c}_{ij} > 0$ for $i = j > n$. The hypotheses are not tested for $i = j \leq n$.

At each iteration of the bootstrap procedure I repeat the following steps:

(1) Construct a resample Y_1 by shuffling overlapping blocks of Y of length 4. Y_1 approximates a random vector with the same marginal distribution as Y, but independent of Z.
(2) Construct resamples X^R and X_1^R by shuffling overlapping blocks of $X = (Y\ Z)$ and $X_1 = (Y_1\ Z)$ of length 4. The order of shuffling is the same for X^R and X_1^R.

(3) Calculate empirical concentration matrices $\mathbf{C} = \left(X^T X\right)^{-1} \cdot (T-1)$ and $\mathbf{C}_1 = \left(X_1^T X_1\right)^{-1} \cdot (T-1)$, where $T = 162$ is the number of observations.

(4) Estimate $\hat{\mathbf{C}}$ in the following way. The first n rows and the first n columns copy from matrix $\mathbf{C}$. In the remaining block, on the right of column n and below row n, the value of each entry is calculated as the difference between the values of the respective entries of $\mathbf{C}$ and $\mathbf{C}_1$.

Let freq_{ij} be calculated as the number of iteration in the bootstrap procedure where $\hat{c}_{ij}$ is positive minus the number of iterations where $\hat{c}_{ij}$ is negative divided by the total number of iterations where $(X^T X)$ is not singular. The p-value associated with the null hypothesis that $\hat{c}_{ij} = 0$ is obtained using:

$$\text{pvalue}_{ij} = \begin{cases} 1 - \left|\text{freq}_{ij}\right|, & \text{if } i \neq j \\ \frac{1}{2} \times \left(1 - \text{freq}_{ij}\right), & \text{if } i = j > n \end{cases} \tag{26}$$

The p-value is not calculated for $i = j \leq n$. Then I use the procedure of Benjamini and Hochberg (1995) described in Section 6.2 for calculating q-values to handle the multiple hypotheses testing problem.

The conditional partial correlations are estimated in the following way. Estimate $\mathbf{C}_1$ in the same way as in the bootstrap procedure above, estimate $\mathbf{C}$ as the inverse covariance matrix of X, calculate $\hat{\mathbf{C}}$ as in the fourth step of the bootstrap procedure, and calculate the partial correlations using:

$$\hat{\rho}_{ij} = \begin{cases} \frac{-\hat{c}_{ij}}{\sqrt{c_{ii} c_{jj}}}, & \text{if } i \neq j \text{ or } i = j > n \\ -1, & \text{if } i = j \leq n \end{cases} \tag{27}$$

The estimated conditional partial correlations and their significance based on the expected false discovery rate are reported in Table 2 and interpreted in the following way. Consider, for example, the partial correlation between the contemporaneous values of u and r, which is equal to -0.24, and is significant at 5% q-value level. This discovery suggests that there exists at least one equation in the estimated model, where both r and u are included, and the coefficients before r and u have the same sign. This may be the Taylor monetary policy rule, where the unemployment negatively affects the interest rate, or some other equation, where r and u influence the dependent variable in the same direction. Another example, the partial correlation between the lagged value of π and the contemporaneous value of r is small and insignificant, pointing at no evidence that r and $\mathbb{L}.\pi$ appear together at least in one equation. The third example, the partial correlation between the lagged variables of u and c is -0.28 and significant at 5% q-value level, suggesting that $\mathbb{L}.u$ and $\mathbb{L}.c$ have at least one common child among the endogenous variables, influencing

TABLE 2. Conditional partial correlations

	r	u	c	g	π	π^c	$\mathbb{L}.r$	$\mathbb{L}.u$	$\mathbb{L}.c$	$\mathbb{L}.g$	$\mathbb{L}.\pi$	$\mathbb{L}.\pi^c$	$\mathbb{L}^2.r$	$\mathbb{L}^2.u$	$\mathbb{L}^2.c$	$\mathbb{L}^2.g$	$\mathbb{L}^2.\pi$	$\mathbb{L}^2.\pi^c$
r	−1 N.A																	
u	-0.24 *	−1 N.A																
c	0.26 $\bullet$	-0.46 ***	−1 N.A															
g	-0.13	-0.18 $\bullet$	0.47 ***	−1 N.A														
π	-0.05	0.08	0.3 ***	-0.22 $\bullet$	−1 N.A													
π^c	0.07	0.02	0.14	-0.19 $\bullet$	0.25 ***	−1 N.A												
$\mathbb{L}.r$	0.7 ***	0.16	-0.25	0.13	0.11	-0.02	−0.48 ***											
$\mathbb{L}.u$	0.07	0.78 ***	0.38 ***	0.08	-0.11	-0.06	-0.04	−0.62 ***										
$\mathbb{L}.c$	-0.24 $\bullet$	0.34 *	0.79 ***	-0.36 $\bullet$	-0.27 *	-0.07	0.22	-0.28 *	−0.6 ***									
$\mathbb{L}.g$	0.09	-0.12	-0.13	-0.05	0	-0.04	-0.07	0.09	0.08	0.01								
$\mathbb{L}.\pi$	0.05	0.02	-0.17	0.21 $\bullet$	0.47 ***	0.06	-0.08	0.02	0.14	0.03	−0.23 *							
$\mathbb{L}.\pi^c$	0.18 *	-0.05	-0.2 $\bullet$	0	0	0.31 $\bullet$	-0.14 $\bullet$	0.07	0.14	-0.02	-0.06	-0.14						
$\mathbb{L}^2.r$	0.03	0.04	0.03	-0.08	-0.07	-0.08	-0.03	-0.05	-0.01	-0.01	0.05	0.01	0.01					
$\mathbb{L}^2.u$	0.11	-0.23 $\bullet$	-0.11	0.06	0.1	0.06	-0.08	0.2	0.08	-0.03	-0.05	-0.03	0.04	-0.09				
$\mathbb{L}^2.c$	0.15	-0.11	-0.27 $\bullet$	0.11	0.16	-0.03	-0.12	0.09	0.19	-0.01	-0.07	-0.03	-0.01	-0.02	-0.04			
$\mathbb{L}^2.g$	-0.08	-0.22 $\bullet$	-0.06	-0.04	-0.03	-0.05	0.05	0.17	0.04	0	0.04	0.02	0.02	-0.05	-0.01	-0.01		
$\mathbb{L}^2.\pi$	0.18	0.08	-0.04	0.1	0.22 $\bullet$	0.02	-0.15	-0.02	0.05	-0.01	-0.14	-0.05	0.03	-0.05	-0.06	0.04	-0.06	
$\mathbb{L}^2.\pi^c$	-0.17	-0.03	-0.01	-0.04	0.12	-0.14	0.11	-0.01	0.01	0	-0.03	0.06	0	0.05	-0.01	-0.01	0.01	-0.04

Significance code based on the expected false discovery rate: *** 0.005 ** 0.01 * 0.05 $\bullet$ 0.1

Partial correlations marked by gray color are associated with p-values greater than 0.1

the endogenous variable in the same direction. Finally, the partial correlation between $\mathbb{L}.r$ and itself is -0.48 and significant at 0.005 q-value level, so $\mathbb{L}.r$ is relevant.

8.3. **An approximate data-mining approach to identification.** The approximate model is estimated using only the empirical distribution function for identification. The advantage of this approach is that it is easy to implement, so it can be used for data mining. I take the first 6 columns of Table 2, and constraint the coefficients associated with insignificant partial correlations to zero. For example, the partial correlation between the contemporaneous values of g and r is -0.13 and insignificant at 10% q-value level, so coefficients a_{14} and a_{41} of matrix $\mathbf{A}$ are constrained to zero. I set the threshold on the q-values at 10% level, however, this choice is arbitrary.

I obtain the following identification scheme for the approximate structural model, which turns out to be sufficient for the full identification:

$$\mathbf{A} = \begin{pmatrix} 1 & a_{12} & a_{13} & 0 & 0 & 0 \\ a_{21} & 1 & a_{23} & a_{24} & 0 & 0 \\ a_{31} & a_{32} & 1 & a_{34} & a_{35} & 0 \\ 0 & a_{42} & a_{43} & 1 & a_{45} & a_{46} \\ 0 & 0 & a_{53} & a_{54} & 1 & a_{56} \\ 0 & 0 & 0 & a_{64} & a_{65} & 1 \end{pmatrix} \quad \mathbf{B}_1 = \begin{pmatrix} b^1_{11} & 0 & b^1_{13} & 0 & 0 & b^1_{16} \\ 0 & b^1_{22} & b^1_{23} & 0 & 0 & 0 \\ 0 & b^1_{32} & b^1_{33} & 0 & 0 & b^1_{36} \\ 0 & 0 & b^1_{43} & 0 & b^1_{45} & 0 \\ 0 & 0 & b^1_{53} & 0 & b^1_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & b^1_{66} \end{pmatrix} \quad \mathbf{B}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b^2_{22} & 0 & b^2_{24} & 0 & 0 \\ 0 & 0 & b^2_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b^2_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{28}$$

where the constrained parameters are substituted with zeros.

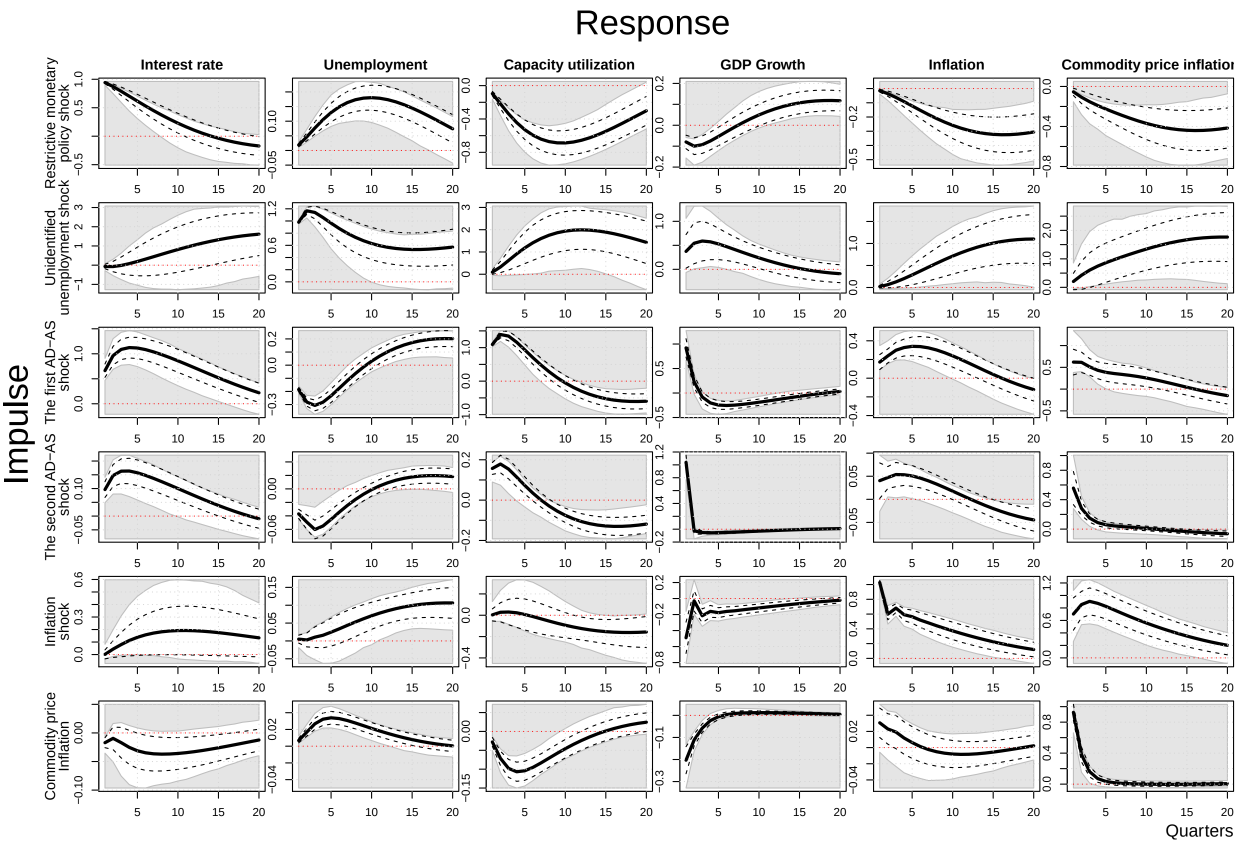

FIGURE 7. Approximate model, estimated impulse-response functions. The thick line in each panel is the impulse-response function, the dashed lines confine the ± one standard deviation zone, the gray area depicts the outside 90% bootstrap confidence intervals.

The estimated impulse response functions are depicted in Figure 7. All significant results are consistent with the theory. For example, the first row represents the response of the economy to a restrictive monetary policy shock. As the theory predicts, the unemployment significantly increases, the capacity utilization decreases, and the GDP growth rate first falls below its long-run level, but then rises above the long-run level, what indicates that the GDP level eventually recovers. The inflation rate and the commodity price inflation rate decrease significantly without pointing at any symptoms of the price puzzle.

However, the impulse response functions in Figure 7 should be taken with the following cautions. First, the procedure of estimation includes a sequential hypothesis testing problem. Some hypotheses are tested for the first time to find significant partial correlations in Table 2, and then other hypotheses are sequentially tested to construct the confidence intervals in Figure 7. Although this procedure produces asymptotically unbiased impulse response functions, the confidence intervals may behave poorly for finite sample problems. Second, I expect that the true structural model is cyclical, but the proposed method of approximate identification asymptotically identifies only recursive models. Finally, some edges present in the true model are clearly

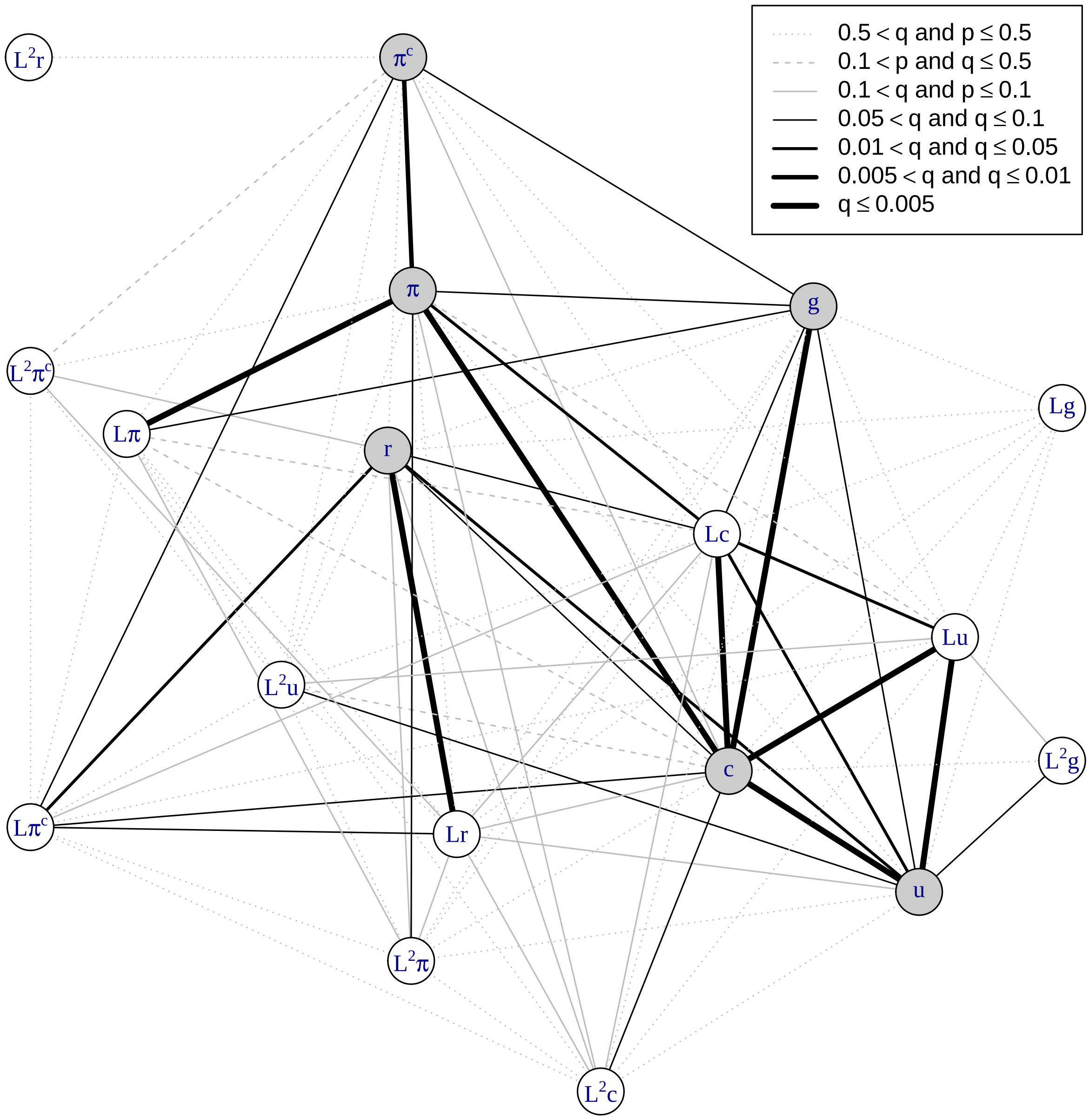

FIGURE 8. The partial concentration network.

missing in the estimated model. The estimated model has 67 constraints, so 67 - 15 = 52 constraints are over-identifying. The log likelihood ratio for this model as compared to the model identified using the Cholesky decomposition is -55, so the hypotheses that the over-identifying restrictions are not binding is rejected at 0.1% p-value significance level. Taking into account all these concerns, in the next section I propose a more sophisticated method of identification.

8.4. **A robust data-oriented identification procedure.** I assume that all variables except for g are persistent, so the lagged value of each variable is present in the structural equation for this variable. The dynamics of g may be not persistent, because the aggregate supply driven by productivity shocks may follow a random walk. In Table 2 I see that $\mathbb{L}r$, $\mathbb{L}u$, $\mathbb{L}c$, $\mathbb{L}\pi$, $\mathbb{L}\pi^c$, and $\mathbb{L}^2\pi$ are relatively strong instruments, and I expect that they possess properties similar to markers (see Definition 9), so I use presumably them to determine which variables are present in which equation.

Consider the partial concentration network is depicted in Figure 8. The main lesson from this graph is that the way information spreads throughout the economy play an important role in macroeconomic dynamics. To see this fact, first, observe that Figure 8 is not entirely consistent with the new Keynesian view. If the Phillips curve were derived from the Calvo's (1983) assumptions, the contemporaneous inflation would depend on its expected value, which, in turn, could be expressed as a function of the state of the economy represented by vector Y_t. Since $\mathbb{L}\pi$ and $\mathbb{L}\pi^c$ are assumably included into the equations respectively for π and π^c, I expect to see that all contemporaneous variables are adjacent to the contemporaneous and lagged values of π and π^c. There is no such evidence in Figure 8.

Figure 8, however, is consistent with the new Keynesian Phillips curves augmented with the Lucas's (1972) islands economy assumption, where agents act rationally, but take into account only information available on their "islands". Particularly, when a firm producing final goods sets prices or make forecasts, it takes into account its marginal costs, which depend on c and π^c, the prices set by its competitors, given by π, and its output, depending on g. Indeed, even if the variations of c, g, π and π^c are moderate at the aggregate level, they may be large at the individual level, whilst the variation of the interest rate is the same at the aggregate and at the individual level. Therefore, it may be rational for the price setters to focus their attention on the individual capacity utilization, individual demand, and on the industry-specific prices, but not on the economy-wide interest rate or unemployment. After aggregation, firm's decisions still depend on c, g, π, and π^c, but not on r or u. Similarly, a firm producing commodities takes into account the demand for its output and its marginal costs, which depend on g and π, but not information available outside the island about the interest rate or the unemployment rate. This is why the contemporaneous and lagged values of r and u are not adjacent to π, $\mathbb{L}\pi$, π^c, and $\mathbb{L}\pi^c$ in Figure 8.

The second observation about the importance of informational channels for macroeconomic dynamics concerns the unemployment equation, and is similar to the observation above about the Phillips curves. The unemployment is significantly adjacent only to the contemporaneous values of c and g, and the lagged unemployment is significantly adjacent to c. I expect to get this result in an economy, where the unemployment is mainly driven by aggregate demand and aggregate supply shocks, and where firms hiring workers consider only information available on their island.

Table 3. Identification assumptions

Structural equation	Variables, present in this equation	Variables, absent in this equation	Variables, for which no precise conclusions have been made
Taylor rule	r, u, $\mathbb{L}r$, $\mathbb{L}\pi^c$, $\mathbb{L}^2\pi$, $\mathbb{L}^2\pi^c$	π, $\mathbb{L}^2r$, $\mathbb{L}^2g$	c, g, π^c, $\mathbb{L}u$, $\mathbb{L}c$, $\mathbb{L}g$, $\mathbb{L}\pi$, $\mathbb{L}^2u$, $\mathbb{L}^2c$
Unemployment equation	u, c, $\mathbb{L}u$, $\mathbb{L}^2u$, $\mathbb{L}^2g$	r, π^c, $\mathbb{L}r$, $\mathbb{L}\pi$, $\mathbb{L}\pi^c$, $\mathbb{L}^2r$, $\mathbb{L}^2\pi$, $\mathbb{L}^2\pi^c$	g, π, $\mathbb{L}c$, $\mathbb{L}g$, $\mathbb{L}^2c$
Aggregate demand	r, c, g, π, $\mathbb{L}c$, $\mathbb{L}^2c$	π^c, $\mathbb{L}\pi^c$, $\mathbb{L}^2r$, $\mathbb{L}^2g$, $\mathbb{L}^2\pi$	u, $\mathbb{L}r$, $\mathbb{L}u$, $\mathbb{L}g$, $\mathbb{L}\pi$, $\mathbb{L}^2u$, $\mathbb{L}^2\pi^c$
Aggregate supply	c, g, π, π^c, $\mathbb{L}c$	$\mathbb{L}^2r$, $\mathbb{L}^2g$ $\mathbb{L}^2\pi$	r, u, $\mathbb{L}r$, $\mathbb{L}u$, $\mathbb{L}g$, $\mathbb{L}\pi$, $\mathbb{L}\pi^c$, $\mathbb{L}^2u$, $\mathbb{L}^2c$, $\mathbb{L}^2\pi^c$
Phillips curve for π	c, g, π, π^c, $\mathbb{L}\pi$, $\mathbb{L}^2\pi$	r, u, $\mathbb{L}r$, $\mathbb{L}u$, $\mathbb{L}^2r$, $\mathbb{L}^2u$, $\mathbb{L}^2c$, $\mathbb{L}^2g$, $\mathbb{L}^2\pi^c$	$\mathbb{L}c$, $\mathbb{L}g$, $\mathbb{L}\pi^c$
Phillips curve for π^c	g, π, π^c, $\mathbb{L}\pi^c$	r, u, $\mathbb{L}r$, $\mathbb{L}u$, $\mathbb{L}c$, $\mathbb{L}g$, $\mathbb{L}^2r$, $\mathbb{L}^2u$, $\mathbb{L}^2c$, $\mathbb{L}^2g$, $\mathbb{L}^2\pi$	c, $\mathbb{L}\pi$, $\mathbb{L}^2\pi^c$

The third observation concerns the position of the federal interest rate in the partial concentration network. The Federal Reserve is usually considered in the literature as the most informed agent. The literature on SVARs, however, recognizes that some information may be available only with a delay. In Figure 8 I see that the interest rate is significantly adjacent to the contemporaneous values of c and u, but only to the lagged value of π^c and to the second lag of π, which supports the hypothesis of information delays.

These observations about the role of information in monetary transmission helped me to formulate testable identification assumptions. All assumptions are discussed in details in Appendix D and summarized in Table 3. The assumptions for the equations for r, u, π, and π^c are based on the Lucas's (1972) assumptions about the diffusion of information discussed above. The assumptions about the AD and AS equations use the only non-testable restriction that the commodity price inflation directly influences the AS equation but not the AD equation. Variable $\mathbb{L}^2r$ is excluded from the entire model, because there is no evidence that it is adjacent to any other variable in Figure 8.

8.4.1. *Sufficiency for the identification.* All identification assumptions are represented in the conditional causal diagram depicted in Figure 9. Black solid edges are assumably present in the conditional causal diagram, and gray dashed edges may be present or not. I need to verify that the identification assumption summarized in Table 3 and depicted in Figure 9 suffice for the full identification. Since the lagged variables produce identifying path of length 1 for themselves, for each endogenous variables I need to verify that only endogenous parents have independent identifying paths.

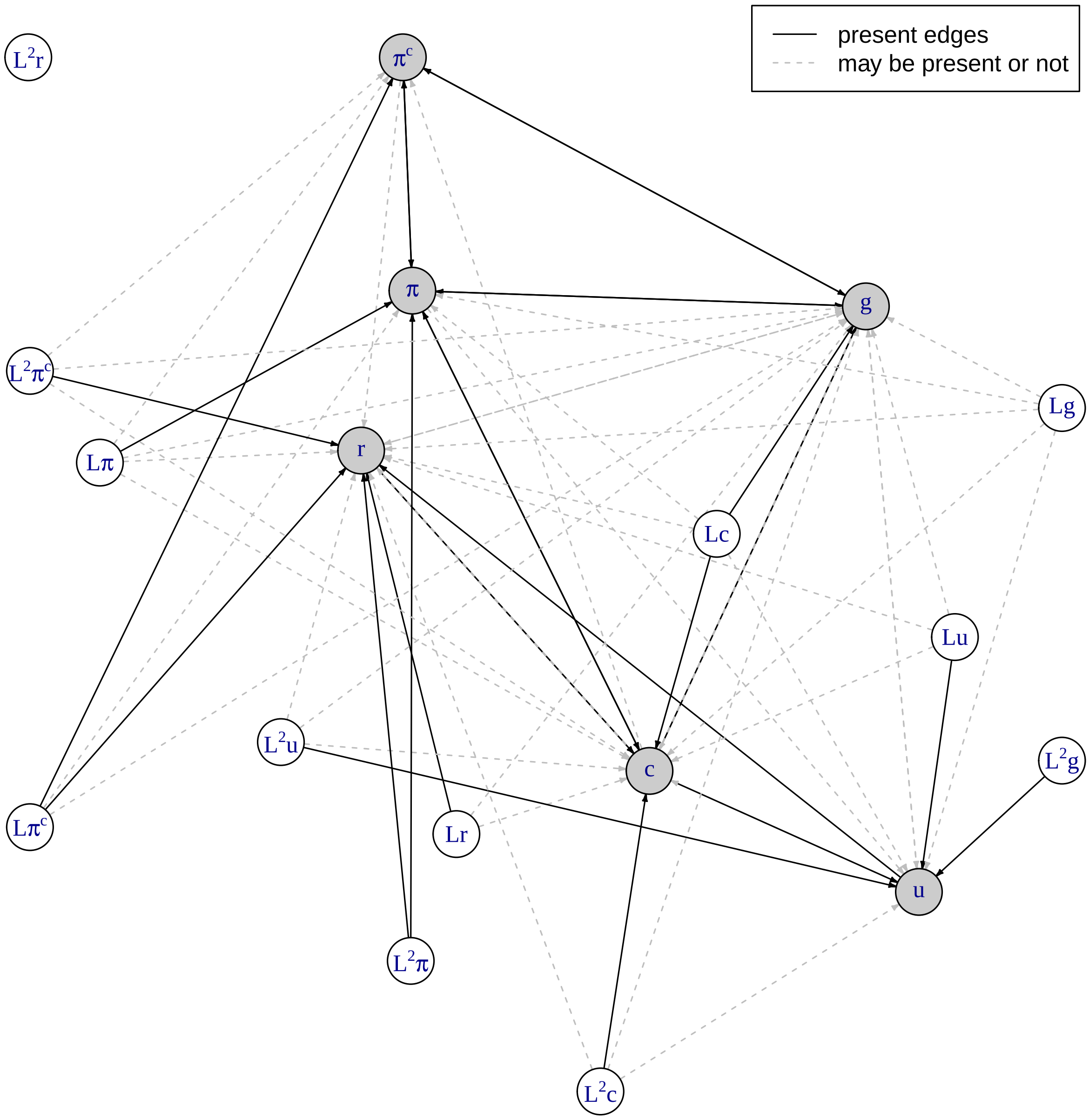

FIGURE 9. Conditional causal diagram.

The first step is to verify the identification of the structural equation for u. This node has up to three contemporaneous parents, c, and possibly g and π. Consider c. I cannot use $\mathbb{L}c$ as an instrument for the identifying path for c, because $\mathbb{L}c$ itself may be a parent of u, in which case path $\mathbb{L}c \to c$ intersects with the self-identifying path for $\mathbb{L}c$ on node $\mathbb{L}c$, so this path is not independent. Nor can I use $\mathbb{L}r$, because edge $\mathbb{L}r \to c$ may not exist. I use, therefore, $\mathbb{L}^2\pi \to r \to c$, which is a valid identifying path, because r and $\mathbb{L}^2\pi$ are not parents of c, and because the presence of edges $\mathbb{L}^2\pi \to r$ and $r \to c$ is guaranteed by the identification assumptions summarized in Table 3. Similarly, the identifying path for g is $\mathbb{L}\pi^c \to \pi^c \to g$, and the path for π is $\mathbb{L}\pi \to \pi$. These paths suffice for the identification of u.

The second step is to verify the identification of π^c, which has no more than three contemporaneous parents, g, π, and possibly c. The identifying path for g is $\mathbb{L}c \to g$, the path for c is $\mathbb{L}^2 c \to c$, and the path for π is $\mathbb{L}^2\pi \to \pi$, so π^c is identified. The third step is to identify c, which have contemporaneous parents r, π, and may be u and g. Since u have already been identified, I need only have identifying paths for r, π, and g, which are $\mathbb{L}^2\pi \to r$, $\mathbb{L}\pi^c \to \pi$, and $\pi^c \to g$. The unidentified contemporaneous parents of g have identifying path $\mathbb{L}^2\pi \to \pi$ and $\mathbb{L}^2\pi^c \to r$, and all contemporaneous parents of r and π at this point have been identified. Therefore, the model is fully identified.

8.4.2. *Checking over-identifying restrictions.* There are 38 restrictions in the estimated model, so 38 - 15 = 23 restrictions are over-identifying. Compared to a model just-identified using the Cholesky decomposition, the log likelihood test does not reject the null hypothesis that the over-identifying restrictions are not binding with the p-value equal to 0.25.

8.4.3. *Estimation results.* Taking into account the identification constraints formulated in Table 3, I estimate the model with the following constraints on parameters:

(29)

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & 0 & a_{16} \\ 0 & a_{22} & a_{23} & a_{24} & a_{25} & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ 0 & 0 & a_{53} & a_{54} & a_{55} & a_{56} \\ 0 & 0 & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix}, \quad \mathbf{B}_1 = \begin{pmatrix} b^1_{11} & b^1_{12} & b^1_{13} & b^1_{14} & b^1_{15} & b^1_{16} \\ 0 & b^1_{22} & b^1_{23} & b^1_{24} & 0 & 0 \\ b^1_{31} & b^1_{32} & b^1_{33} & b^1_{34} & b^1_{35} & 0 \\ b^1_{41} & b^1_{42} & b^1_{43} & b^1_{44} & b^1_{45} & b^1_{46} \\ 0 & 0 & b^1_{53} & b^1_{54} & b^1_{55} & b^1_{56} \\ 0 & 0 & 0 & 0 & 0 & b^1_{66} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} 0 & b^2_{12} & b^2_{13} & 0 & b^2_{15} & b^2_{16} \\ 0 & b^2_{22} & b^2_{23} & b^2_{24} & 0 & 0 \\ 0 & b^2_{32} & b^2_{33} & 0 & 0 & b^2_{36} \\ 0 & b^2_{42} & b^2_{43} & 0 & 0 & b^2_{46} \\ 0 & 0 & 0 & 0 & b^2_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & b^2_{66} \end{pmatrix}, \quad \mathcal{E}_t = \begin{pmatrix} \varepsilon_t^{\mathrm{MP}} \\ \varepsilon_t^{u} \\ \varepsilon_t^{\mathrm{AD}} \\ \varepsilon_t^{\mathrm{AS}} \\ \varepsilon_t^{\pi} \\ \varepsilon_t^{\pi^c} \end{pmatrix}$$

where $\varepsilon^{\mathrm{MP}}$ is the monetary policy shock, ε^u is the unidentified unemployment shock, $\varepsilon^{\mathrm{AD}}$ is the AD shock, $\varepsilon^{\mathrm{AS}}$ is the AS shock, ε^{π} is the inflation shock, ε^{π^c} is the stagflation shock, and the constrained parameters are substituted with zeros.

The estimated impulse response functions are depicted in Figure 10. The response of the economy to the restrictive monetary policy shock is depicted in the first row. As the theory predicts, this shock temporary raises the unemployment rate and lowers the capacity utilization rate. The GDP growth rate is below the long run level during the first and second year after the shock, and rises above afterwards, so the GDP level recovers. The inflation immediately goes down, without any symptoms of the price puzzle. The commodity price inflation also immediately goes down, and as the theory predicts, it decreases faster than the GDP inflation rate, because π^c is a leading indicator for π.

The second row depicts the response of the economy to the unidentified unemployment shock. The response of the federal interest rate to the unemployment shock is immediate and strong, an 1% increase in

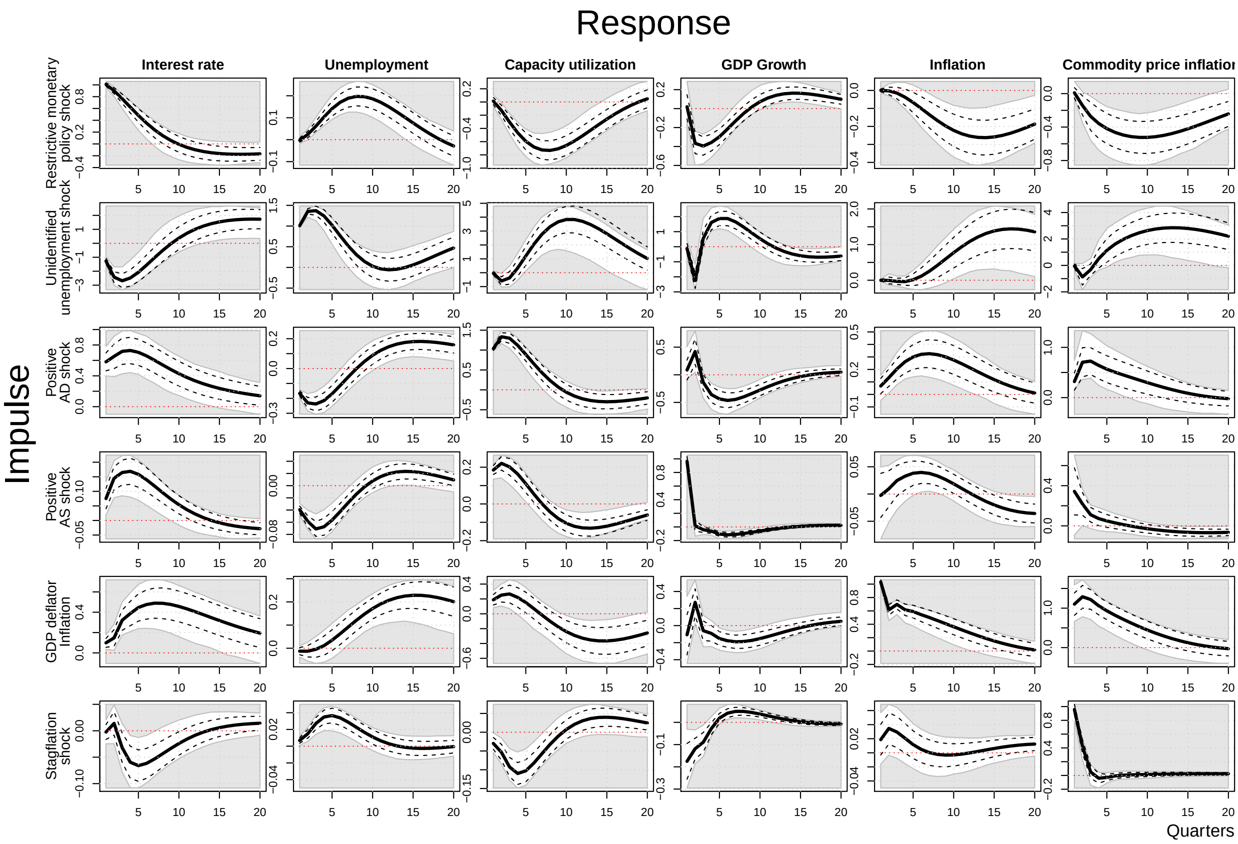

FIGURE 10. Final model, estimated impulse-response functions. The thick line in each panel is the impulse-response function, the thin dashed lines show the ± one standard deviation zone, the gray zone depicts the outside 90% confidence intervals.

the unemployment rate produces about 2.5% decrease in the interest rate. Since I do not distinguish between the labor demand and labor supply shocks, I cannot theoretically predict whether the unemployment shock increases or decreases the capacity utilization rate and the GDP growth rate. In Figure 10 I see that they temporary increase, what suggests that the labor supply effect on average dominates. The inflation rate and the commodity price inflation rate significantly rise, possibly because of the stimulating monetary policy response to the positive unemployment shock.

The third line represents the response of the economy to an aggregate demand shock. The GDP grows significantly in the second quarter after the shock, and then it reverts towards the initial trend. The response of the monetary policy to the AD shock is also strong, the AD shock increasing the output approximately by 0.5% raises the interest rate almost by 1%. In agreement with the theory, the unemployment decreases, the capacity utilization, the inflation, and the commodity price inflation go up.

The response of the economy to a positive AS shock is presented in the forth line. From the theoretical perspective, I expect to see two effects of the AS shock on the economic activity. First, the shock permanently increases the potential GDP, so I expect that g increases in the period of shock, gets back to the initial value afterwards, and the values of other variables remain unchanged. Second, this shock decreases the long-run equilibrium ratio of the GDP deflator to the commodity price index, and because of price rigidities, it temporary increases the capacity utilization, decreases the unemployment, and produces a trend-reverting dynamics of the GDP level. In Figure 10 I see that both effects are significant. The AS shock increases the capacity utilization and decreases the unemployment. The federal interest rate possibly responds to these changes and capacity utilization movements, and significantly increases. The GDP growth rate is significant and large in the period of the shock, and has a small but significant revert component after the shock. If scaled to 1% GDP increase shocks, responses of r, u, c and π to the AS shock are similar to the responses to the AD shocks, but quantitatively much smaller. In contrary, the response of the commodity price inflation quantitatively is about the same, because both the aggregate demand and the aggregate supply shocks increase the demand for the commodities.

The response of the economy to the inflation shock is depicted in line 5 of Figure 10. One percent increase in π, which is not attributed to the aggregate demand or aggregate supply shocks, increases r approximately by 0.6%, and this turns out to be sufficient to suppress the inflation. Like the theory predicts, an increase in the inflation rate is associated with a rise in the economic activity in the short run, but the following restrictive monetary policy and increase in the price produce a slowdown. This explains the decrease of u and increase of c and g in the period of the shocks, with the subsequent rising of u and falling of c and g in the medium run.

The response of the economy to a commodity price inflation shock is depicted in the sixth line. Like the theory predicts, this shock produces stagflation. In Figure 10, the unemployment after this shock rises, the capacity utilization decreases, the GDP growth rate temporary decreases, and the inflation goes up.

Therefore, all finding in Figure 10 are consistent with the macroeconomic theory.

9. Conclusions

This paper proposes a method of testable identification of SEMs and SVARs with orthogonal structural shocks. A sufficient condition for the existence of testable identification is that each structural equation has a marker, and that the true structural model is identified. This method produces narrow and theory-consistent confidence intervals in the application example, where I estimate a SVAR monetary model of the US economy.

The method can be used in a large variety of applications, including SEMs, SVARs, DSGE models, large Bayesian SVARs, and in other areas, where theoretical assumptions do not suffice for the full identification, and where the shock orthogonality assumption is appropriate. The shock orthogonality assumption applies only when the state of the economy is well-described by the variables included into the structural model. In the application example, the shock orthogonality assumption is not appropriate if model with r, g and π does not include u, c, and π^c, because of the confounding effect.

The estimated SVAR model revealed the importance of informational channels through which information about the structural shocks propagate throughout the economy. This model is a promising framework for analysis of interactions of the real sector of the US economy with the oil market, financial markets, the world economy and so on.

10. Computational details

The SVAR model of the US economy was estimated in `R` (R Core Team, 2012) using packages `igraph` (Csardi and Nepusz, 2006) and `NLopt` (Johnson, 2014).

References

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.

Benjamini, Y. and D. Yekutieli (2001): "The Control of the False Discovery Rate in Multiple Testing under Dependency," *The Annals of Statistics*, 29, 1165–1188.

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

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

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12, 383 – 398.

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

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1 – 45.

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.

Clarida, R., J. Galí, and M. Gertler (1998): "Monetary policy rules in practice: some international evidence," *European Economic Review*, 42, 1033–1067.

Csardi, G. and T. Nepusz (2006): "The igraph software package for complex network research," *InterJournal*, Complex Systems, 1695.

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

Galí, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89, 249.

Galí, J., J. D. Lopez-Salido, and J. Valles (2003): "Technology shocks and monetary policy: assessing the Fed's performance," *Journal of Monetary Economics*, 50, 723–743.

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

Hanson, M. S. (2004): "The "price puzzle" reconsidered," *Journal of Monetary Economics*, 51, 1385–1413.

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.

Johnson, S. G. (2014): "The NLopt nonlinear-optimization package," http://ab-initio.mit.edu/nlopt.

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.

Lucas, R. E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4, 103–124.

Maćkowiak, B. and M. Wiederholt (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99, 769–803.

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

Mankiw, N. G. and R. Reis (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 1295–1328.

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.

R Core Team (2012): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

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.

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.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

——— (1986): "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2 – 16.

——— (1992): "Interpreting the macroeconomic time series facts: the effects of monetary policy," *European Economic Review*, 36, 2–16.

——— (1998): "Stickiness," *Carnegie-Rochester Series on Public Policy*, 49, 317–356.

——— (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

Sims, C. A. and T. Zha (2006): "Were There Regime Switches in US Monetary Policy?" *American Economic Review*, 96, 54–81.

Smets, A. and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97, 586 – 606.

Smets, F. and F. Wouters (2003): "An estimated stochastic general equilibrium model of the Euro Area," *Journal of the European Economic Association*, 1, 1123 – 1175.

Uhlig, H. (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Veldkamp, L. L. (2011): *Information choice in macroeconomics*, Princeton University Press.

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.

Woodford, M. (2001): "Imperfect Common Knowledge and the Effects of Monetary Policy," *NBER*, Working Paper no. 8673.

Zha, T. (1999): "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics*, 90, 291–316.

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 am 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 $x_{j_1}x_{j_2}\ldots x_{j_s}$ in the causal diagram, the set of parameters associated with this path consists of the following elements of matrix $\bar{\mathbf{P}}$: $\{p_{j_2j_1}, p_{j_3j_2}, \ldots, p_{j_sj_{s-1}}\}$. 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 $\mathbb{L}$ 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.

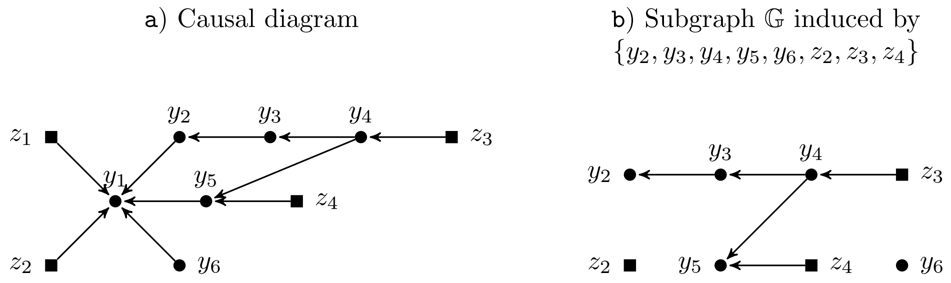

FIGURE 11. Example of causal diagram and subgraph $\mathbb{G}$

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

$$\begin{pmatrix} 1 & a_{12} & 0 & 0 & a_{15} & a_{16} \\ 0 & 1 & a_{23} & 0 & 0 & 0 \\ 0 & 0 & 1 & a_{34} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & a_{54} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b_{43} & 0 \\ 0 & 0 & 0 & b_{54} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}, \tag{30}$$

which causal diagram is depicted in Figure 11a. 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 11b. 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:

$$\mathbf{M} = \begin{pmatrix} \underline{a_{23}} & 0 & 0 & 0 & 0 \\ 1 & \underline{a_{34}} & 0 & 0 & 0 \\ 0 & 1 & 0 & \underline{-b_{43}} & 0 \\ 0 & a_{54} & 0 & 0 & \underline{-b_{54}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{31}$$

There are two independent paths in $\mathbb{G}$ starting with nodes in $\mathcal{X}_1$ and reaching $\mathcal{Y}_1$, see Figure 11b, 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 (31), and is $[a_{23} \cdot a_{34} \cdot (-b_{43}) \cdot (-b_{54})]$. This permutation has four elements, so $\mathbb{L} = 4$. 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* $\mathcal{Y}_2$*:*

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

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

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 (30), 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 $\mathbb{L} \geq N + |\mathcal{Y}_2|$.

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 $x_{j_0} x_{j_1} \ldots x_{j_s}$ is added to the graph, modify the permutation in the following manner:

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

In example (30), 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.

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

Consider a permutation of length $\mathbb{L}$. Since all parameters associated with one permutation are located in different columns of matrix $\bar{\mathbf{P}}$, at least $\mathbb{L} - |\mathcal{Y}_2|$ 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 $p_{j_1 j_0}$, 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 $p_{j_1 j_0}$ have been included into the permutation, parameter $p_{j_1 j_1}$, which is normalized to be positive, cannot be included into this permutation, because it is in the sam row as $p_{j_1 j_0}$. Therefore, column j_1 either is not included into permutation, or there exists parameter $p_{j_2 j_1}$, 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 $\mathbb{L}$, 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 $\mathbb{L} - |\mathcal{Y}_2|$ 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 $\mathbb{L} = N + |\mathcal{Y}_2|$ □

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:

$$\mathbb{E}(Y|Z) = \mathbf{A}^{-1}\mathbf{B}\cdot Z \equiv \mathbf{\Lambda}\cdot Z \tag{32a}$$

$$\mathrm{Var}(Y-\mathbb{E}(Y|Z)) = \left(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A}\right)^{-1} \equiv \mathbf{\Omega} \tag{32b}$$

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 (32). 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 $\underset{n\times(n+m)}{\bar{\mathbf{P}}} = \begin{pmatrix}\underset{n\times n}{\mathbf{A}} & \underset{n\times m}{-\mathbf{B}}\end{pmatrix}$ are written as:

$$e_i^T\bar{\mathbf{P}}\mathbf{\Psi}_i = 0 \tag{33}$$

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:

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

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 $\bar{\mathbf{P}}$ are identified if and only if $\mathrm{rank}\left(\bar{\mathbf{P}}\mathbf{\Psi}_i\right) = n-1$, 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:

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

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 "$\setminus$" 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 $\bar{\mathbf{P}}\mathbf{\Psi}_i$ 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: $\operatorname{rank}(\mathbf{M}_i) = \operatorname{rank}\left(\bar{\mathbf{P}}\mathbf{\Psi}_i\right)$.

By definition of $\mathbf{\Psi}_i$, each column of $\bar{\mathbf{P}}\mathbf{\Psi}_i$, 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 $\operatorname{rank}(\mathbf{M}_i) = n - 1$, 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.** 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 32. 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 $\tilde{\mathbf{A}} = \mathbf{RA}$ and $\tilde{\mathbf{B}} = \mathbf{RB}$, 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 (33). 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:

$$\mathbf{M}_j = \left(\begin{bmatrix} \\ \bar{\mathbf{P}}\mathbf{\Psi}_j \\ \\ \end{bmatrix} \quad \begin{matrix} \begin{bmatrix} \mathbf{I}_{j\times j} \end{bmatrix} \\ \\ \begin{bmatrix} \mathbf{0}_{(n-j)\times j} \end{bmatrix} \end{matrix} \right) \tag{34}$$

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:

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -b_{22} & 0 \\ -b_{11} & -b_{12} & 0 \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -b_{11} & 0 & 0 \end{pmatrix} \quad \mathbf{M}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{35}$$

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:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \\ \\ \\ \\ \end{pmatrix}_{4\times 0} \tag{36}$$

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 (36) 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 "$\setminus$" 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 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 $\tilde{\mathbf{A}} = \mathbf{RA}$, $\tilde{\mathbf{B}} = \mathbf{RB}$, and $\mathbf{R}$ has the following block structure:

$$\mathbf{R} = \begin{pmatrix} \underset{n_1 \times n_1}{\mathbf{R}_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix} \tag{37}$$

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 $\mathbf{P}\boldsymbol{\Psi}_i$ 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 (34) 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 (37), and because nodes $y_{n_1+1}, y_{n_1+2}, \ldots, y_{i-1}$ are identified, $\mathbf{R}$ has the following structure:

$$\mathbf{R} = \begin{pmatrix} \underset{(i-1)\times(i-1)}{\mathbf{I}} & \mathbf{0} \\ \mathbf{0} & \underset{(n-i)\times(n-i)}{\mathbf{R}_{33}} \end{pmatrix} \tag{38}$$

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 $e_i \mathbf{P}\boldsymbol{\Psi}_i = 0$ and $e_i \mathbf{R}\mathbf{P}\boldsymbol{\Psi}_i = \mathbf{0}$, so $e_i (\mathbf{R} - \mathbf{I}) \mathbf{P}\boldsymbol{\Psi}_i = 0$. 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 5

Proof of Proposition 5. Since the order of variables is arbitrary, assume that $i = 1$, so the first row in $\bar{\mathbf{P}}\boldsymbol{\Psi}_1$ 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 $\text{rank}\left(\mathbf{\Pi}\left(\mathcal{P}_1|Z_1\right)\right) < |\mathcal{P}_1|$. To prove the necessity of the reduced form rank condition, I need to assume that $\text{rank}\left(\mathbf{\Pi}\left(\mathcal{P}_1|Z_1\right)\right) < |\mathcal{P}_1|$, 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:

$$V^T\mathbf{A}Y^{\mathbb{E}} = V^T\mathbf{B}Z \tag{39}$$

Make the following observations about (39). 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 (39). Therefore, I can rewrite (39) in the following form:

$$V^T\tilde{\mathbf{A}}\tilde{Y}^{\mathbb{E}} = V^T\tilde{\mathbf{B}}\tilde{Z}, \tag{40}$$

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 (40). 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:

$$\Lambda_1^T \begin{pmatrix} \tilde{Y}^{\mathbb{E}} \\ \tilde{Z} \end{pmatrix} = 0 \tag{41a}$$

$$\Lambda_2^T \begin{pmatrix} \tilde{Y}^{\mathbb{E}} \\ \tilde{Z} \end{pmatrix} = 0 \tag{41b}$$

where Λ_1 is obtained from the first line of (8), and Λ_2 is just another way to write(40); as I discuss above, Λ_1 and Λ_2 a linearly independent.

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

$$\mathbb{E}\left(\left.\begin{pmatrix}\tilde{Y}^{\mathbb{E}}\\ \tilde{Z}\end{pmatrix}\right| Z\right) = \hat{\mathbf{\Pi}} Z \tag{42}$$

By this definition, matrix $\mathbf{\Pi}(\mathcal{P}_i|\mathcal{Z}_i)$ from Proposition 5 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,

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

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

Appendix D. Identification assumptions for the estimated SVAR model

To formulate robust identification restrictions, I analyze the concentration network depicted in Figure 8, and compare it with predictions from the macroeconomic theory. I need to solve the clique cover problem for the partial concentration network knowing that it has a solution with no more than 6 cliques, and taking into account that the evidence presented in Figure 8 is not precise, because some edges may be missing for the reason of low power of the respective tests, and other edges may represent false discoveries. Therefore, I need to make reasonable assumptions about the true moral graph, taking into account the theory and the estimated concentration network.

For each equation, I divide all variables into three groups. The first group includes the variables assumably present in the respective structural equation, the second group consists of the variables assumably absent, and the variables for which no precise conclusions have been made are put into the third group. All variables from the first and third groups are included into the structural equation of the estimated model, and all variables from the second group are excluded from this equation. The variables from the first and second groups are used to analyze the partial concentration network and to make conclusions about the presence of other variables in the respective structural equation, and the variables from the third group are not used for this purpose.

I use the following two-step procedure to formulate testable identification restrictions for each structural equation. First, I use the theory to attribute each variable to one of the three groups defined in the previous paragraph. Second, I analyze the partial concentration network depicted in Figure 8. For the variables from the first and second groups I verify whether the partial concentration network is consistent with the theoretical assumptions that have been made in the first step. For the variables from the third group I verify whether the concentration network helps to attribute them to the first or to the second group. Therefore, the theoretical assumptions about the variables in the third group may be revised after the analysis of the partial concentration network.

There are $n = 6$ endogenous variables, so I need at least $n(n-1)/2 = 15$ identification restrictions to meet the Rothenberg's (1971) necessary condition for the full identification. My objective, however, is not to find exactly 15 restrictions, but rather to find a set of restrictions sufficient for the full identification, possibly over-identifying the structural model but in such way that the over-identifying restrictions are not significantly binding. The reason is that relaxing restrictions, which are known to be true, widens confidence intervals without any gains in the consistency or efficiency. Then I use log likelihood test to verify whether the hypothesis that over-identifying restrictions are not binding is not rejected.

D.1. **Identification of the unemployment equation.** I begin with theoretical assumptions about the unemployment equation, defining the equilibrium value of u. I assume that the structural shock ε^u comprises the demand and supply shocks on the labor market. Since I do not distinguish between the labor demand and the labor supply equations, I refer to ε^u as to the unidentified unemployment shock. The structural equation for u, therefore, accounts both for the labor demand and for the labor supply effects.

I make the following theoretical assumptions. First, I assume that the aggregate demand shock affects the unemployment, so the contemporaneous value of c is present in the structural equation for u, and the lagged values of c may be present or not in this equation. Second, I do not make any assumptions about the influence of the aggregate supply shocks onto the unemployment, so the contemporaneous and lagged values of g may be present or not in this equation. Third, I assume that the unemployment is persistent, so the first lag of u is present in this equation, and the second lag may be present or not. Fourth, I assume that r and π^c affect u only with the mediation of the aggregate demand and aggregate supply, and because I have included c and g into the estimated structural equation for u, I do not have to include any contemporaneous or lagged values of r and π^c. Finally, I assume that π may be present in the unemployment equation, because the inflation may produce monetary illusions.

Consider now the evidence for the unemployment equation in Figure 8. Since c, u, and $\mathbb{L}u$ are assumably present in this equation, and because each structural equation produces a clique in the partial moral graph, all variables present in this equation are adjacent to each other and to c, u, and $\mathbb{L}.u$ in the partial moral graph. The partial concentration network depicted in Figure 8, however, gives only an estimation of the partial moral graph, where some edges may be absent because of low power of the tests, and where some present edges may represent false discoveries. Therefore, I need to make reasonable assumptions about the true moral graph taking into account the theory and the estimated concentration network.

First, I assess whether $\mathbb{L}^2u$ is present or not in the structural equation for u. As I discuss above, the theory does not provide a definite answer to this question, so consider the partial concentration network in Figure 8 and try to find an answer there. Since c, u, and $\mathbb{L}.u$ are assumed to be present in this equation, if $\mathbb{L}^2u$ is also present, then variables u, $\mathbb{L}u$, $\mathbb{L}^2u$, and c should form a clique in the partial moral graph. There is some support for this hypothesis in Figure 8. Edge $\mathbb{L}^2u - u$ is significant at 10% q-value level, and $\mathbb{L}^2u - \mathbb{L}u$ is significant at 10% p-value level. Edge $\mathbb{L}^2u - c$ is not significant, possibly because the power of the test associated with this edge is week, which is consistent with the evidence presented in Table 2 that the relevance of $\mathbb{L}^2u$ is not very strong, see the the diagonal element for $\mathbb{L}^2u$ in this table. There is no evidence that $\mathbb{L}^2u$ is adjacent to any other node in the concentration network, so the structural equation for u is probably the only one where $\mathbb{L}^2u$ is included. Taking into account all these facts, I revise the theoretical assumption about $\mathbb{L}^2u$, and assume hereafter that $\mathbb{L}^2u$ is present in the structural equation for u, see Table 3.

Consider now g and its lags. Edge $\mathbb{L}^2g - u$ is significant at 10% q-value level, and edge $\mathbb{L}^2g - Lu$ is significant at 10% p-value level. There is no other significant edges connecting $\mathbb{L}^2g$ to any node of the graph, which suggests that $\mathbb{L}^2g$ is present only in the equation for u. Hereafter I assume that $\mathbb{L}^2g$ is in the equation for u.

There is no evidence that $\mathbb{L}g$ is connected to any node in the partial concentration network, so there is no evidence that Lg enters into the equation for u. The adjacency of g to u and c in the partial concentration network can be explained by other structural equations (see Table 3 summarizing all identification assumptions), and there is no evidence that g is adjacent to $\mathbb{L}u$, $\mathbb{L}^2u$ or $\mathbb{L}^2g$. So neither for $\mathbb{L}g$ nor for g there is evidence that the variable enters into the equation for u. However, having assumed that $\mathbb{L}^2g$ enters into the equation for u, it is prudent to assume that $\mathbb{L}g$ and g may also enter there.

I have made the theoretical assumption that the contemporaneous and lagged values of r and π^c influence u only with the mediation of c and g, so they are not included into the equation for u. This assumption does not contradicts the partial concentration network in Figure 8, because there is no significant edges connecting any contemporaneous or lagged values of r and π^c to $\mathbb{L}u$, $\mathbb{L}^2u$ or $\mathbb{L}^2g$. The significant edges connecting the contemporaneous or lagged values of r and π^c to the contemporaneous values of c and u are predicted below by other structural equations. There is no evidence that π or its lags are adjacent to the members of the structural equation for u, however, it is prudent to assume that π may be in the structural equation for u, because it may create monetary illusions.

D.2. **Identification of the monetary policy rule.** I assume that the federal interest rate responds to the shocks of the GDP gap and of the inflation rate. The GDP gap is measured in the estimated model by c and u, and the inflation rate is measured by π and π^c. However, following the literature on SVAR models, I do not assume that the contemporaneous values of c, u, and π are necessarily present in the monetary policy rule equation, because the Federal Reserve may have only lagged data, or because it may intentionally not respond to the contemporary value of π assuming that the contemporaneous values of c, u, and π^c provide more information about the future inflation than π itself when the lagged values of π are taken into account. Figure 8 does not provide any empirical support for the assumption that the contemporaneous value of π^c is present in the monetary policy rule equation, nevertheless I follow the literature on SVARs, assuming that π^c may be present. The GDP growth rate and its lagged values may be included or not into the monetary policy rule equation, and the reason not to include is that the GDP gap measured by c and u has already been included. I assume that the policy rule is persistent, so $\mathbb{L}r$ is present in the policy rule, and $\mathbb{L}^2r$ may be present or not.

Consider Figure 8. There is no evidence that $\mathbb{L}^2r$ is present in any structural equation, so to narrow the confidence intervals, I exclude it from the entire model. There is no evidence that g or its lagged values are present in the structural equation for r. However, including g and $\mathbb{L}g$ into the monetary policy rule makes the confidence intervals for the response function of g to the monetary policy impulse narrower without considerably affecting its expected value. Therefore, I assume that g and $\mathbb{L}g$ may be present in the monetary policy rule, but $\mathbb{L}^2g$ is not included.

Consider the contemporaneous value of u. Edge $u - r$ is significant at 5% q-value level, and edge $u - \mathbb{L}r$ is significant at 10% p-value level, but there is no significant edges making r and $\mathbb{L}r$ adjacent to the other variables present in the equation for u. This observation is consistent with the assumption that u is present in the equation for r, but r and $\mathbb{L}r$ are not present in the equation for u. Therefore, I assume that u is in the equation for r. There is no evidence that $\mathbb{L}u$ or $\mathbb{L}^2u$ affect r, but this is prudent to assume that they may.

Observe that r, $\mathbb{L}r$, c, and $\mathbb{L}c$ form a significant clique in the partial concentration network. This clique may be explained by the assumption that c and $\mathbb{L}c$ are present in the equation for r, or by the assumption that r and $\mathbb{L}r$ are present in the equation for c, but I cannot distinguish between these two assumptions. Therefore, I assume that any assumption may be true, so c and $\mathbb{L}c$ may be present in the equation for r, and r and $\mathbb{L}r$ may be present in the equation for c. There are edges $\mathbb{L}^2c - r$ and $\mathbb{L}^2c - Lr$ significant at 10% p-value level, indicating that $\mathbb{L}^2c$ may be in the policy rule equation.

There are significant edges in the partial concentration network, making $\mathbb{L}\pi^c$, $\mathbb{L}^2\pi^c$, and $\mathbb{L}\pi$ adjacent to r and $\mathbb{L}r$, indicating that they are present in the monetary policy rule. However, there is no evidence that π, $\mathbb{L}\pi$ or π^c are present there. Taking into account this evidence, I assume that π is not present in the Taylor rule, $\mathbb{L}\pi^c$, $\mathbb{L}^2\pi^c$, and $\mathbb{L}\pi$ are present there, and $\mathbb{L}\pi$ and π^c may be present or not.

D.3. **Phillips curves.** The theoretical assumptions concerning the Phillips curves are obtained assuming new Keynesian Phillips curve augmented with the Lucas's (1972) island assumption. They predict that π depends on the contemporaneous values of c and π^c, and may depend on the contemporaneous value of g and lagged values of c, π^c and g. Similarly, π^c depends on the contemporaneous values of g and π, and may depend on the contemporaneous value of c and lagged values of c, g, and π. The Lucas's (1972) island assumption excludes the contemporaneous and lagged values of r and u from the Phillips curves. The persistency assumption is that π depends on $\mathbb{L}\pi$ and may depend on $\mathbb{L}^2\pi$. Similarly, π^c depends on $\mathbb{L}\pi^c$ and may depend on $\mathbb{L}^2\pi^c$.

Consider Figure 8, and start with the Phillips curve for π. Edge $\mathbb{L}^2\pi - \pi$ is significant at 10% q-value level, and edge $\mathbb{L}^2\pi - \mathbb{L}\pi$ is significant at 10% p-value level, indicating that $\mathbb{L}^2\pi$ is present in the structural equation for π. The edges connecting $\mathbb{L}^2\pi$ to r and $\mathbb{L}r$ have been previously explained by the presence of $\mathbb{L}^2\pi$ in the Taylor rule. Therefore, I assume that $\mathbb{L}^2\pi$ is present in the Phillips curve for π and in the Taylor rule, but there is no evidence that $\mathbb{L}^2\pi$ is present in any other structural equations.

The partial correlations between $\mathbb{L}^2\pi^c$ and π^c, and between $\mathbb{L}^2\pi^c$ and π^c are not significant, so there is no evidence that $\mathbb{L}^2\pi^c$ is present in the Phillips curve for π^c. From the theoretical perspective, however, this is prudent to assume that $\mathbb{L}^2\pi^c$ may be present there.

There is a strong evidence that π and π^c are both included at least into one structural equation, however, there is no evidence that π or its lags enter into the Phillips curve for π^c, or that π^c or its lags enter into the Phillips curve for π. Indeed, the partial correlation between π and π^c is significant at 1% q-value level. The partial correlations between $\mathbb{L}\pi$ and π^c, and between $\mathbb{L}^2\pi$ and π^c are not significant, pointing at no evidence that π^c is present in the structural equation for π. Similarly, the partial correlation between $\mathbb{L}\pi^c$ and π is not significant, so there is no evidence that π is present in the equation for π^c. However, I make the theoretically prudent assumption that π is present in the structural equation for π^c, and $\mathbb{L}\pi$ may be present there. Similarly, π^c is present in the structural equation for π, and $\mathbb{L}\pi^c$ may be present or not.

There is no evidence in Figure 8 against the Lucas' island assumption. Indeed, there is not significant edges making at least one contemporaneous or lagged value of r or u adjacent to π or π^c. Therefore, I assume that this assumption is correct, so the contemporaneous and lagged values of r and u are not included into the Phillips curves.

D.4. **Identification of the AD and AS equations.** In the beginning of Section 8 I made the rough assumption that an increase in g keeping constant c is interpreted as a positive AS shock, and an increase in c keeping constant g as is interpreted a simultaneous positive AD and negative AS shock. For identification of AD and AS shocks in this section, however, I use the following refined assumption. I assume that the state of the aggregate demand and of the aggregate supply is described by the contemporaneous values of c and g, so all variables affecting c or g may affect the aggregate demand and aggregate supply equation. The AD and AS equations can be represented as:

$$\text{AD}_t = \alpha c_t + \beta g_t \tag{43a}$$

$$\text{AS}_t = (1-\alpha)\, c_t + (1-\beta)\, g_t \tag{43b}$$

where α and β are unknown parameters. The rough assumption that an increase in g keeping constant c is interpreted as a positive AS shock corresponds to $\alpha = 1$. To identify the structural model, however, I introduce a more sophisticated identification restriction, and assume that α may be not 1. Namely, the AD and AS equations are distinguished one from the other using the only untestable identification assumption that the commodity price inflation affects only the aggregate supply, but not the aggregate demand. The other identifying assumptions are the same for the AD and AS equations.

I make the following theoretical assumptions about the AD and AS equations. I assume that the commodity price inflation is not included into the AD equation, the contemporary value of π^c is included into the AS equation, and the lags of π^c may be included or not into both equations. The GDP deflator inflation is included into the AD and AS equations, and its lags may be included or not. The AD shocks are assumed to be persistent, so the lagged value of c is present in the AD equation and may be present in the AS equation. I assume that AS and AD shocks affect the GDP growth rate, so g is present in both equations. I assume that the effect of u on the aggregate demand may be delayed due to the income channel effect, so u and its lags are likely to be present in the AD equation. The contemporaneous interest rate influences the AD equation by affecting the consumption and investment, so it is present in the AD equation. The contemporaneous values of r may be present or not in the AS equation because of the lags in the investment decisions and implementation, but the lagged values of r are likely to be present there. Finally, since g is a growth value and c is a level value, if c is present into the equation for g, $\mathbb{L}c$ should also be present there. The contemporaneous and lagged values of the other variables may be present or not in each equation.

Let us take into account the evidence depicted in Figure 8. There is no evidence that $\mathbb{L}^2 r$, $\mathbb{L}^2 g$, $\mathbb{L}^2 \pi$, or $\mathbb{L}^2 \pi^c$ is adjacent to c, $\mathbb{L}c$, or g. Therefore, I assume that these variables are absent in the AD and in the AS equations. Edge $\mathbb{L}^2 c - c$ is significant at 10% q-value level, and edge $\mathbb{L}^2 c - Lc$ is significant at 10% p-value level, so I assume that $L^2 c$ is present in the AD equation, and may be present or not in the AS equation. There is also no evidence that $\mathbb{L}g$ and $L^2 u$ are present in any equation, however, this is prudent to assume that it may be present there. Since $\mathbb{L}\pi^c$ is adjacent to c and $\mathbb{L}c$, and because I have assumed that $\mathbb{L}\pi^c$ does not affect the AD, I assume $\mathbb{L}\pi^c$ is present in the AS equation. Edges $r - c$, $r - \mathbb{L}c$, and $r - \mathbb{L}^2 c$ are significant respectively at 10% q-value level, 10% q-value level and 10% p-value level, which supports the assumption that r is included into the AD equation.

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.