Bertsche, Dominik; Brüggemann, Ralf; Kascha, Christian

Directed graphs and variable selection in large vector autoregressive models

Journal of Time Series Analysis

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Bertsche, Dominik; Brüggemann, Ralf; Kascha, Christian (2022) : Directed graphs and variable selection in large vector autoregressive models, Journal of Time Series Analysis, ISSN 1467-9892, John Wiley & Sons, Ltd, Oxford, UK, Vol. 44, Iss. 2, pp. 223-246, https://doi.org/10.1111/jtsa.12664

JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **44**: 223–246 (2023)
Published online 14 August 2022 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12664

ORIGINAL ARTICLE

DIRECTED GRAPHS AND VARIABLE SELECTION IN LARGE VECTOR AUTOREGRESSIVE MODELS

DOMINIK BERTSCHE[a] RALF BRÜGGEMANN[b*] AND CHRISTIAN KASCHA[c]

[a]*SWICA Health Insurance, Winterthur ZH, Switzerland*
[b]*Department of Economics, University of Konstanz, Konstanz, Germany*
[c]*S Rating und Risikosysteme GmbH, Berlin, Germany*

We represent the dynamic relation among variables in vector autoregressive (VAR) models as directed graphs. Based on these graphs, we identify so-called strongly connected components. Using this graphical representation, we consider the problem of variable choice. We use the relations among the strongly connected components to select variables that need to be included in a VAR if interest is in impulse response analysis of a given set of variables. Our theoretical contributions show that the set of selected variables from the graphical method coincides with the set of variables that is multi-step causal for the variables of interest by relating the paths in the graph to the coefficients of the 'direct' VAR representation. An empirical application illustrates the usefulness of the suggested approach: Including the selected variables into a small US monetary VAR is useful for impulse response analysis as it avoids the well-known 'price-puzzle'.

Received 15 June 2020; Revised 30 April 2022; Accepted 18 July 2022

Keywords: Directed graphs; impulse response analysis; multi-step causality; variable selection; vector autoregression.

JEL. C32; C51; C55; E52.

1. INTRODUCTION

Vector autoregressive (VAR) models are popular tools in the analysis of multiple time series to conduct structural analysis in form of an impulse response analysis. The popularity of the VAR model is at least partly due to the fact that it typically does not require strong economic theory assumptions. Often VAR models without any restrictions on the parameters are used to describe the joint dynamics of a set of economic time series. While the general VAR lag structure allows to uncover dynamic relations between the variables included in the system, the use of unrestricted VARs comes at a cost: The number of parameters to be estimated from the data increases with the square of the number of variables in the system. Even in moderately large VARs the degrees of freedom exhaust quickly. Thus, applied researchers have to choose the number of variables to be included in the VAR wisely. On the one hand, a researcher would like to include all relevant variables to avoid omitted variable bias and to get a complete picture of the underlying dynamics. On the other hand, including too many variables makes parameter estimates unreliable and estimation uncertainty may lead to rather uninformative results such as estimated impulse responses with very wide confidence intervals.

Given variables of interest, our article suggests to use a graphical modeling approach to select a 'minimal' VAR containing only variables that are relevant to model the dynamics of the variables of interest. This approach is helpful in selecting the relevant variables for VAR analysis in a data-driven way. We argue that this is a useful addition to the toolbox of time series econometricians as on the one hand, it exploits the information from large dimensional data sets but on the other hand eventually uses smaller VAR models for structural analysis.

* Correspondence to: Ralf Brüggemann, Department of Economics, University of Konstanz, Box 129, 78457 Konstanz, Germany.
E-mail: ralf.brueggemann@uni-konstanz.de

To fix ideas, suppose a researcher is interested in a set of variables denoted by y^{I}, including say GDP growth, the consumer price (CPI) inflation, and a key interest rate. She wants to conduct an impulse response analysis for the variables in y^{I}. For this purpose, she has to decide which variables to include from a large cross-section of available time series on, for example, output, income, consumption, the labor market, orders and inventories, money and credit, interest and exchange rates, financial market variables and various price measures.[1]

In recent years, suggestions have been made on how to include the information from a large dimensional data set into VARs. Factor-augmented VARs (FAVARs) (see e.g. Bernanke *et al.*, 2005; Stock and Watson, 2016) condense the information from a large time series data set into a few factor time series, which are then included in a VAR model. Factor-augmented models have also been widely used for structural analysis.[2] Clearly, these models are only suitable if the underlying data has a factor structure, that is, if the large number of time series are really driven by a small number of common factors (see e.g. Uhlig, 2009 on this point). An alternative are large Bayesian VARs (BVARs) as suggested by Banbura *et al.* (2010).[3] In large dimensional settings, however, these models require to use a very tight prior. Consequently, using a large BVAR might impose more structure on the model than typical VAR users feel comfortable with. Other shrinkage methods have also been used for large VAR models, including the least absolute shrinkage and selection operator (LASSO) (see e.g. Kascha and Trenkler, 2015; Barigozzi and Brownlees, 2019).[4] The LASSO approach can handle large dimensional VAR models by setting some VAR coefficients to zero and at the same time shrinking the remaining coefficients. It is well known that LASSO may lead to biased estimates and recently, inference methods based on debiased estimators have been suggested (see e.g. Javanmard and Montanari, 2014; Van de Geer *et al.*, 2014; Zhang and Zhang, 2014, and in the VAR context Basu *et al.* (2019) and Zheng and Raskutti (2019) with references therein).[5] We argue below that using an estimated LASSO-VAR directly for impulse response analysis may be daunting in large VARs because it complicates the economic interpretation of results.

Consequently, the mentioned modeling approaches may not be ideal in some situations faced by applied time series econometricians. Researchers may also prefer to use smaller VARs because they are easier to interpret and resemble more closely small scale dynamic stochastic general equilibrium (DSGE) models used in macroeconomics. At the same time, the researcher would like to include additional 'relevant' variables that affect impulse responses for the variables of interest. Against the background of the large number of time series available today, this entails a variable selection procedure.

Our article is concerned with the question of how to choose variables for the smallest ('minimal') VAR containing all variables that are 'relevant' for the impulse responses of variables of interest y^{I}. In other words, we are suggesting a procedure on how to choose a suitable information set for a VAR modeling exercise. This consists of the following steps: In the first step, a data-driven model selection technique (e.g. LASSO) can be used to obtain a subset VAR. In the second step, we then use the subset VAR structure and concepts from graphical modeling to select those variables that need to be in a smaller VAR to capture all relevant information for the variables of interest. The final step of the procedure then consists of estimating and using an unrestricted VAR in all selected variables.

The main novelty of this article is in introducing and theoretically analyzing the second step. The first contribution here is to use so-called strongly connected components (SCCs) and the relation among these components for variable selection. The concept of SCCs is well-established in the graphical modeling literature but to the best of our knowledge, this concept has not been used in econometrics. We first represent a sparse VAR structure as a directed graph with vertices and edges. From this graph we identify all SCCs by using a simple graphical modeling algorithm. We show how the SCCs and their connections to other SCCs are helpful in determining the variables

[1] Large data sets of this type have been used in various studies. See, for example, Stock and Watson (2003) or McCracken and Ng (2016) with references therein. They typically contain up to 130 variables.
[2] See, for example, Stock and Watson (2002), Ludvigson and Ng (2007), Eickmeier and Ziegler (2008), Ludvigson and Ng (2009), Stock and Watson (2012), Clements (2016) and Cheng and Hansen (2015) for applications of factor-augmented regressions and FAVARs.
[3] See, for example, Carriero *et al.* (2009, 2012, 2015), Giannone *et al.* (2014), and Koop (2013) for applications of this method.
[4] See also Stock and Watson (2012).
[5] Other theoretical properties like, for example, oracle inequalities for LASSO estimators in VARs are discussed in, for example, Basu and Michailidis (2015), Kock and Callot (2015), and Medeiros and Mendes (2016).

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

that need to be included in a VAR modeling exercise. Effectively, the set of relevant variables can be found from the graphical representation of the SCCs, known as a component graph. Thus, the first innovation of our procedure is to combine an existing method of the graph theoretic literature with the problem of variable selection for VAR analysis. For VAR practitioners this may be more useful compared to just using standard regularization techniques as a LASSO-VAR, since our method enables researchers to work with smaller VAR models, which may be easier to interpret from an economic point of view. In a second theoretical contribution, we show the relation between the SCCs and the concept of multi-step causality as in Dufour and Renault (1998). In particular, given the variables of interest y^{I}, we show that a minimal VAR chosen by SCCs is identical to the VAR that contains y^{I} and all variables that are multi-step causal for y^{I}. Here we contribute to the literature by relating two existing concepts from the econometric and the graph theoretic literature.

Methodologically, our article is related to the literature on using graphical models in econometrics. Following the work on causal analysis of multivariate data (see e.g. Lauritzen, 1996; Edwards, 2000; Pearl, 2000), graphical models have also been introduced for time series models. Brillinger (1996) and Dahlhaus (2000) are the first papers mentioning the use of graphical modeling for time series data and present concepts based on the partial correlation and partial spectral coherence.[6] Dahlhaus and Eichler (2003) introduce causality graphs based on the autoregressive representation. Our work is most closely related to the work of Eichler (2006, 2007, 2012), who shows the close relation of different causality concepts (Granger-causality and multi-step causality) to graphical representations in vector autoregressive models. We add to this work the link from causality structures to variable selection using the concept of strongly connected components.[7]

Our article is also related to the work of Jarociński and Maćkowiak (2017), who also investigate the question of variable choice for VAR analysis, albeit with a different econometric approach. Based on the concept of Granger-causal priority (see Sims, 1982, 2015; Doan and Todd, 2010), their paper evaluates in a Bayesian setup the posterior probability of Granger-causal priority. For a given set of variables of interest y^{I}, Jarociński and Maćkowiak (2017) would drop a variable y_j, say, if the variables in y^{I} are likely to be Granger-causal prior to y_j. Thus, their method may also be used to choose variables for VAR analysis.

Shrinkage in the form of LASSO is also sometimes referred to as a variable selection technique. This approach typically involves shrinking individual VAR coefficients (but not deleting variables from the VAR system) and has been used in VAR models for forecasting and to estimate networks and measures of connectivity (see e.g. Davis *et al.*, 2016; Medeiros and Mendes, 2016; Barigozzi and Hallin, 2017; Barigozzi and Brownlees, 2019). Our approach is different in the sense that we use a LASSO-VAR in the first part of our procedure only as an intermediate step to select the relevant variables. Our goal is to compute impulse responses. In contrast to the mentioned studies above, we do not base the final structural analysis on the large-dimensional VAR. Rather, in our setup the final choice of variables for the VAR analysis is based on the relation among the variables of interest and the SCCs. The structural (impulse response) analysis is then conducted within the smaller VAR containing only the selected variables. This may be preferred over using a large subset or LASSO-VAR because the smaller VARs can be better linked to theoretical macroeconomic (DSGE) models. Furthermore, in contrast to the other approaches, this allows for a detailed analysis of the transmission channels among (blocks of) variables and is therefore especially useful for understanding and interpreting the effects of economic shocks. Finally, while LASSO may be used in the first step to determine the subset VAR, it is important to note that other subset techniques can be applied as well.

We illustrate the usefulness of our suggested variable selection approach in an application to US macroeconomic data. The variables of interest in y^{I} are US output, CPI inflation and the federal funds rate, three variables often used in stylized three-variable VARs for the US. Given y^{I}, we use our variable selection method based on SCCs to select a minimal VAR from 41 US time series for a period between 1975 and 2014. Starting point is a sparse VAR structure obtained from applying the LASSO to the large VAR. Regardless of the considered estimation period, six out of the 41 variables are always selected into the model and the selection is fairly stable over different samples

[6] See also Flamm *et al.* (2012) for an overview of different approaches.
[7] Graphical modeling has also been used for identifying the instantaneous relations. The first work in this area is the paper by Swanson and Granger (1997), followed by a number of studies that use graphical modeling for identifying structural VAR models (see e.g. Demiralp and Hoover, 2003; Hoover *et al.*, 2009; Heinlein and Krolzig, 2012).

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

before the financial crisis in 2008/2009. Moreover, additional variables are selected into the model in a number of periods. Consequently, the 'minimal VAR' is still relatively large, indicating that the underlying relations are typically quite complex and may not be captured adequately in a three-variable VAR. We also find that including the selected variables into the VAR leads to more reasonable responses to a monetary policy shock, indicating that the selection is useful.

The remainder of the article is structured as follows. Section 2 shows how VARs can be represented as directed graphs. We also introduce the concept of strongly connected components and explain how this can be used for variable selection and for finding a 'minimal VAR'. Section 3 relates the graph-theoretical concepts to multi-step causality and shows how variable selection based on both concepts leads to the same set of relevant variables. In Section 4, we illustrate the usefulness of our method in an empirical application. Section 5 concludes. All proofs are deferred to the appendix.

2. VECTOR AUTOREGRESSIVE MODELS, DIRECTED GRAPHS AND STRONGLY CONNECTED COMPONENTS

We explain how VAR models can be represented by directed graphs. We then review the concept of SCCs in directed graphs. Finally, we explain how SCCs can be used for selecting relevant variables. This latter part provides a novel contribution to the econometrics literature since, to the best of our knowledge, the concept of SCCs has not yet been used in econometrics.

We denote the VAR model of order p, a VAR(p) for the K-dimensional time series vector $y_t = (y_{1,t}, y_{2,t}, \dots, y_{K,t})'$ by:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.1}$$

where $A_1, \dots, A_p$ are $K \times K$ coefficient matrices and u_t is a zero mean white noise disturbance vector with non-singular covariance matrix Σ_u. We have not included deterministic terms (e.g. intercepts) into the VAR in (2.1) to simplify the notation. Adding deterministic terms would not change the results discussed below and they can be included in empirical work.

To make use of graphical modeling concepts, we represent the VAR in (2.1) as a directed graph. Following the standards in graphical modeling, a directed graph G is described by a set of vertices V and a set of edges E that are ordered pairs of vertices. In our application, the vertices correspond to the K elements in vector y_t and the edges are determined by the elements of the autoregressive matrices $A_1, \dots, A_p$ as in the following definition (Eichler, 2007):

Definition 2.1. (Directed VAR graph)**.** Given a VAR(p) model as in (2.1), the associated directed graph is $G = (V, E)$ with $V = \{1, \dots, K\}$ and

$$(i,j) \in E \;\Leftrightarrow\; \exists s \in \{1, \dots, p\} : A_{ij,s} \neq 0, \tag{2.2}$$

where $A_{ij,s}$ denotes the element in row i and column j of A_s.

Remark 2.1. In this graph, a directed edge (i,j) leads from vertex i to vertex j. This is standard in the graphical modeling literature. In our context, using this definition $(i,j) \in E$ implies that the ith variable *depends on* the jth component in y_t, however, the arrow would point from vertex i to vertex j. Thus, the direction of the arrows is reversed compared to the type of arrows sometimes used to denote (Granger-)causality.

Given a sparse VAR structure, that is, a VAR with zero restrictions on the VAR coefficients, we may use the associated directed graph to learn about the set of relevant variables. We do so by using the notion of the SCCs in a graph (Tarjan, 1972). In order to define these, one makes use of the concept of a path or a pathway. A path is defined to be a sequence of vertices to go from one vertex to another. More formally, a path P of length k leading from vertex u to u' in graph $G = (V, E)$ is a sequence $P = (v_0, v_1, \dots, v_k)$ of vertices such that $u = v_0$ and $u' = v_k$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

and $(v_{i-1}, v_i) \in E$ for $i = 1, 2, \dots, k$. If there is a path from u to u', we say that u' is *reachable* from u, denoted as $u \overset{P}{\rightsquigarrow} u'$. We may now define the strongly connected components of a directed graph:

Definition 2.2. (Strongly connected components (Tarjan, 1972))**.** Let $G = (V, E)$ be a directed graph and let two vertices u and v in G be equivalent if $u \overset{P}{\rightsquigarrow} v$ and $v \overset{P}{\rightsquigarrow} u$, that is, u and v are mutually reachable. Call the corresponding equivalence classes of vertices V_i and let $C_i = (V_i, E_i)$ where $E_i = \{(u, v) \in E : u, v \in V_i\}$ for $i = 1, \dots, k$. The subgraphs C_i are called the strongly connected components of G.

Remark 2.2. Note that if we refer to SCC C_i in the following, we mean the corresponding set of vertices V_i that belongs to C_i in the sense of Definition 2.2. For example, if we write $j \in C_i$, we mean $j \in V_i$.

Remark 2.3. Note that each vertex of graph G belongs to exactly one strongly connected component. Consequently, the set of all equivalence classes $V_1, \dots, V_k$ belonging to the strongly connected components $C_1, \dots, C_k$ forms a partition of the set of vertices V such that $V = V_1 \cup \cdots \cup V_k$ (see e.g. Duff and Reid, 1978).

Remark 2.4. Following Tarjan (1972), a depth-first search algorithm may be used to compute the strongly connected components efficiently. We have implemented a variant of Tarjan's algorithm using Matlab according to the exposition in Cormen *et al.* (2009).

In the next step, we condense the information of the graph G by moving from graph G to a graph of the SCCs. The resulting graph is called a component graph, which is defined next.

Definition 2.3. (Component graph)**.** A component graph is defined as $G^{\text{SCC}} = (V^{\text{SCC}}, E^{\text{SCC}})$, where $V^{\text{SCC}} = \{C_1, \dots, C_k\}$ is the set of strongly connected components of graph G. There is an edge $(C_i, C_j) \in E^{\text{SCC}} \Leftrightarrow \exists x \in C_i : \exists y \in C_j : (x, y) \in E$.

Remark 2.5. Definition 2.3 implies that there is only an edge (C_i, C_j) between two strongly connected components if the original graph G has a directed edge from one member of the SCC C_i to a member of the SCC C_j.

Remark 2.6. Duff and Reid (1978) suggest to order the SCCs such that there is no path from one strongly connected component to another later in the sequence, that is, the SCCs $C_1, \dots, C_k$ may be ordered such that there is no path from C_i to any C_j for $j > i$. The associated reordered matrix of the graph is then lower block-triangular. Each block on the diagonal corresponds to one of the SCCs. In the context of economic applications, the structure of the SCCs may give additional insights on the relevance of different variables.

Remark 2.7. The component graph may be viewed as a condensed view of the original graph. Essentially, the component graph collapses all edges of the original graph whose vertices are contained in the same SCC.

Finally, for a given set of variables of interest, say $y^{\text{I}} \subseteq y$, that is, $\text{I} \subseteq \{1, \dots, K\}$, we would like to identify a 'minimal' set of variables which have to be taken into account when modeling y^{I}. We denote this set of relevant variables as $R(y^{\text{I}})$. For that purpose, we define the set $R_{G^{SCC}}(C_i)$ as the set of all variables contained in SCCs that are *reachable* from C_i in G^{SCC} (including C_i). Thereby, note that *reachability* in G^{SCC} is defined analogous to G. That is, $R_{G^{SCC}}(C_i)$ can be interpreted as the set of variables on which C_i 'depends' and which have to be taken into consideration when modeling variables in C_i. We specify the 'minimal' VAR system in Definition 2.4.

Definition 2.4. (Relevant variables)**.** Given a subset of interest $y^{\text{I}} \subseteq y$, the *minimal* VAR is a VAR composed of the series that are contained in the relevant SCCs given by:

$$R(y^{\text{I}}) := \bigcup_{\{C_i : y^{\text{I}} \cap C_i \neq \emptyset\}} R_{G^{SCC}}(C_i).$$

Thereby, the set of variables that are reachable from SCC C_i $(i = 1, \dots, k)$ is defined as:

$$R_{G^{SCC}}\left(C_i\right) := \left\{ y_j \in y : \left(j \in C_i\right) \vee \left(\exists l \in \{1, \dots, k\} : j \in C_l : C_i \overset{P}{\rightsquigarrow} C_l \right) \right\}.$$

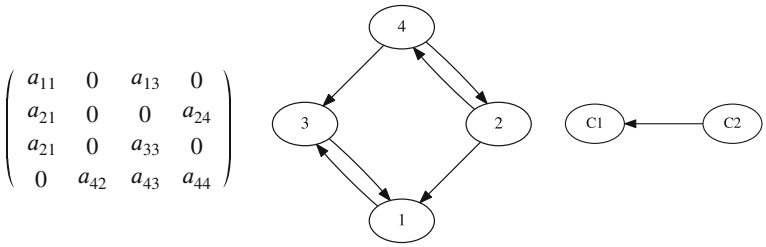


Figure 1. VAR matrix, associated directed graph, and component graph

Remark 2.8. Definition 2.4 states that the minimal VAR is the one composed of the series in the SCCs which contain elements of y^{I} and all SCCs that may be reached from these SCCs.

We illustrate the graph theoretical concepts using a simple example, starting with a four-dimensional VAR(1) with coefficient matrix as shown on the left side in Figure 1.[8] The associated directed graph indicates that this system has two strongly connected components. The set of vertices of the first SCC C_1 consists of variables 1 and 3, that is, $V_1 = \{1, 3\}$, as vertex 1 may be reached from vertex 3 and vice versa. The set of vertices of the second SCC C_2 consists of variables 2 and 4, that is, $V_2 = \{2, 4\}$. Note that the vertices within each strongly connected component are mutually reachable (see Definition 2.2) and each variable (vertex) is contained in exactly one SCC (see Remark 2.3). Moreover, SCC C_1 is reachable from SCC C_2 but not vice versa.

We may also illustrate Remark 2.6 as it is easy to see that we may reorder the variables such that a lower block-triangular matrix results. For this purpose, we order the SCC that has no leaving edges first. In our example, SCC C_1 containing the set of vertices $V_1 = \{1, 3\}$, has no leaving edges and is hence ordered first. The new ordering of variables is then (1, 3, 2, 4), which results in the following reordered VAR matrix:

$$A^* = \left(\begin{array}{cc|cc} a_{11} & a_{13} & 0 & 0 \\ a_{31} & a_{33} & 0 & 0 \\ \hline a_{21} & 0 & 0 & a_{24} \\ 0 & a_{43} & a_{42} & a_{44} \end{array}\right) = \left(\begin{array}{c|c} A_{11} & 0 \\ \hline A_{21} & A_{22} \end{array}\right),$$

where the coefficients within matrix A^* still have their original names. Obviously, this matrix has the desired block-triangular form. After having grouped the variables according to the strongly connected components, we may now draw a corresponding component graph according to the partition of the matrix A^*. The resulting component graph is shown in the right panel of Figure 1. This component graph indicates that component C_1 may be reached from component C_2 but not vice versa. Consequently, if the variable(s) of interest are included in component C_1, then the minimal VAR only includes the variables contained in C_1 but not those contained in C_2. In contrast, if the variable(s) of interest are contained in C_2, a corresponding VAR needs to include the variables from C_2 and in addition, the variables from C_1 as C_1 may be reached from C_2. For instance, if the variable of interest is, for example, variable 3, then the VAR needed in structural analysis needs to contain all variables that are in the corresponding component C_1. In our example, these are the variables 1 and 3. In contrast, if the variable of interest is variable 2, then we need to include all variables in C_2 and C_1, that is, all four variables, in the VAR model. Similarly, we may find the minimal VAR from the component graph even if we have more than one variable of interest. In our example, if variables 1 and 3 are of interest, a VAR for just these two variables suffices as both variables form a strongly connected component (C_1) and no other strongly connected component can be reached from C_1. In contrast, if variables 2 and 4 are of interest, we need a VAR with all variables from C_1 and C_2 as C_1 may be reached from C_2. In other words, we need all four variables. Now assume that variables 1 and 2 are of interest. Then again,

[8] To avoid cluttering of the graph, we exclude self-loops from the graphical representation.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

we need to consider all variables from the strongly connected components that include variable 1 and variable 2, which in our example boils down to again using all variables since there are only the two components C_1 and C_2.

According to Definition 2.4 our procedure provides us with the minimal set of variables that should be included in a VAR used, for example, for impulse response analysis. In other words, we employ the graph theoretical concept of SCCs to select the VAR variables and view our procedure as a tool to choose the VAR information set in a systematic way. In the final step of our modeling approach, we therefore estimate a VAR using the selected variables $R(y^I)$. Note that on this step, we follow the standard practice in VAR modeling used for impulse response analysis and estimate a full VAR without any restrictions on the VAR coefficients. Consequently, we estimate the VAR in this final step by standard multivariate least squares. This simple post-selection estimation approach is similar in spirit to the approach in Belloni and Chernozhukov (2013).[9] A detailed flow chart illustrates the entire procedure in Appendix C.

3. ECONOMETRIC CAUSALITY CONCEPTS AND GRAPHS

The graph theoretical concepts discussed in Section 2 have a close relation to multi-step causality concepts in time series econometrics. In this section we explain how the two concepts are related and show that the set of variables selected for a minimal VAR by the SCC method as in Definition 2.4 of Section 2 coincides with the set of variables that are multi-step causal for at least one of the variables of interest.

The simple notion of Granger-causality (see Granger, 1969) is known to neglect any indirect effects and influences of 'auxiliary' variables as it is based on 1-step ahead predictability. Consequently, the original definition of Granger non-causality is not helpful in the context of variable selection. A more general causality concept that takes into account all indirect effects of auxiliary variables is known as multi-step causality and has been formally introduced into the literature by Dufour and Renault (1998).[10] Informally, a subset y^B of the variables causes another subset y^A at a specific horizon h if the best linear forecast for y^A at horizon h can be improved by including the variables in y^B in the information set. Dufour and Renault (1998) discuss necessary and sufficient conditions for non-causality at different forecast horizons h. Dufour *et al.* (2006) focus on developing corresponding multi-step non-causality tests in the context of VAR models. For our purpose, it is convenient to note that multi-step non-causality at different horizons may be formulated as linear exclusion restrictions on the so-called direct VAR model. For $h \geq 1$, we write this direct VAR model as:

$$y_{t+h} = \Pi_1^{(h)} y_t + \cdots + \Pi_p^{(h)} y_{t-p+1} + u_{t+h}^{(h)}, \tag{3.3}$$

where this representation is obtained by successive substitution from the VAR in (2.1). Dufour and Renault (1998) show that $\Pi_1^{(0)} = I_K$, $\Pi_s^{(1)} = A_s$, $\Pi_s^{(h+1)} = A_{s+h} + \sum_{l=1}^{h} A_{h-l+1}\Pi_s^{(l)} = \Pi_{s+1}^{(h)} + \Pi_1^{(h)} A_s$ and the MA($h-1$) innovation term $u_{t+h}^{(h)} = \sum_{j=0}^{h-1} \Pi_1^{(h)} u_{t-j}$.

Given sets of indices A and B, let $\Pi_{AB,s}^{(h)}$ denote the submatrix of $\Pi_s^{(h)}$ consisting of the intersection of rows with indices in A and columns with indices in B. If A and B are singletons, say $A = \{k\}$, $B = \{l\}$, we simply write $\Pi_{kl,s}^{(h)}$. We reproduce Theorem 3.1 of Dufour and Renault (1998) tailored to the regular, finite VAR case.

Theorem 3.1. (Dufour and Renault (1998))**.** Given $y^A, y^B \subseteq y$ and y is generated by a regular, finite-order VAR(p) as in (2.1), it is:

$$y^B \not\to_h y^A \quad \Leftrightarrow \quad \forall s = 1, \ldots, p : \Pi_{AB,s}^{(h)} = 0,$$

[9] We point out, however, that this method ignores the uncertainty due to selection and standard inference confidence intervals for impulse responses may understate the true uncertainty somewhat. Accounting for selection uncertainty is left for future research.

[10] The effect of intermediate variables have also been pointed out earlier by, for example, Lütkepohl (1993), Penm and Terrell (1986), and Sims (1980) but Dufour and Renault (1998) were the first who formalized the concept of multi-step causality in a general framework.

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

where 0 indicates a zero matrix of appropriate dimension. That is, y^B does not cause y^A at horizon h if and only if all the relevant coefficients in the direct VAR model for horizon h are zero.

By definition, y^B *causes* y^A *at lag* h if at least one of the parameter matrices in the above theorem is not zero. For some indices I, we denote by $C(y^{\mathrm{I}})$ the set consisting of the variables in y^{I} itself and all variables that cause y^{I} at any horizon in the above sense. When I is a singleton, say $\mathrm{I} = \{i\}$, we write $C(y_i)$ instead. Formally, we define the causal variables in Definition 3.1.

Definition 3.1. (Causal variables)**.** Given $y^{\mathrm{I}} \subseteq y$, the set of variables that cause y^{I} is given by:

$$C\left(y^{\mathrm{I}}\right) := \left\{y_j \in y : \left(y_j \in y^{\mathrm{I}}\right) \vee \left(\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}}\right)\right\}.$$

First, we investigate the case of a VAR with $p = 1$. For this case, we show that the coefficients of the direct VAR representation $\Pi_1^{(h)}$ are related to the set of paths in the directed graph representing the VAR model. To show this, note that the direct VAR representation for $p = 1$ is:

$$y_{t+h} = \Pi_1^{(h)} y_t + u_{t+h}^{(h)}, \quad \text{with} \quad \Pi_1^{(h)} = A_1^h \text{ for all } h \in \mathbb{N}. \tag{3.4}$$

By induction, each element $\Pi_{ij,1}^{(h)}$ in $\Pi_1^{(h)}$ can be linked to the set of all paths that lead from the associated vertices i to j in h steps. We state this result formally in Theorem 3.2.

Theorem 3.2. Given two variables $y_i, y_j \in y$ following a regular VAR(1) as in (2.1), the entry at position (i,j), $\Pi_{ij}^{(h)}$, corresponds to the set of paths:

$$\mathbb{P}_{ij}^{(h)} = \left\{P : P = (e_1, \ldots, e_h) : \forall k = 1, \ldots, h : e_k = (v_{k-1}, v_k) \in E, v_0 = i, v_h = j\right\}, \tag{3.5}$$

leading from vertex i to vertex j for all $h \in \mathbb{N}$ in that:

$$\Pi_{ij}^{(h)} = \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m) \in P} a_{lm}. \tag{3.6}$$

Proof. See Appendix A.1. ∎

$\mathbb{P}_{ij}^{(h)}$ is the set of all paths of length h leading from vertex i to vertex j. Theorem 3.2 essentially states, that the coefficients of the direct VAR $\Pi_{ij}^{(h)}$ can be written in terms of sums of products of autoregressive coefficients in A_1, where the indices correspond to edges on different paths from i to j. To illustrate this, consider again our simple VAR(1) from Figure 1. In this example, there are two paths of length $h = 2$ from variable 2 to variable 3, thus the set of paths is:

$$\mathbb{P}_{23}^{(2)} = \{\langle(2,1),(1,3)\rangle, \langle(2,4),(4,3)\rangle\}.$$

Using the result from Theorem 3.2, we find for $h = 2$, $i = 2$ and $j = 3$:

$$\Pi_{23,1}^{(2)} = a_{21}a_{13} + a_{24}a_{43}.$$

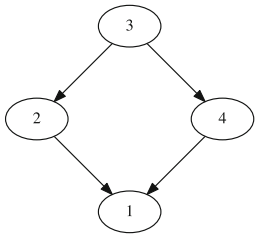
Note that the indices of the VAR coefficients match with the edges of the path set $\mathbb{P}_{23}^{(2)}$. Obviously, the same $\Pi_{23,1}^{(2)}$ would be obtained from the direct VAR coefficient definition.

At first sight, the result of Theorem 3.2 seems to imply that variable j would be multi-step causal for variable i, whenever there exists at least one path from i to j. However, $\mathbb{P}_{ij}^{(h)} \neq \emptyset$ does *not* imply that variable y_j causes y_i as

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

the following example illustrates. Consider a VAR(1) with the associated directed graph:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 \\ -0.5 & 0 & 0 & 1 \end{pmatrix}$$



In this example, we have $\mathbb{P}_{31}^{(1)} = \emptyset$ and $\mathbb{P}_{31}^{(2)} = \{\langle (3,2),(2,1)\rangle, \langle (3,4),(4,1)\rangle\}$ but variable y_3 is not caused by y_1 at neither horizon one nor horizon two since:

$$\Pi_{31}^{(2)} = a_{32}a_{21} + a_{34}a_{41} = 1/4 - 1/4 = 0.$$

Furthermore, one can easily verify $\Pi_{31}^{(h)} = 0$ and $\mathbb{P}_{31}^{(h)} \neq \emptyset$ for all $h \geq 2$. This 'canceling-out' effect of course happens very rarely with any real data set and most reasonable estimation methods. Therefore, one might exclude it by assumption.

Assumption 3.1. Given a VAR(1) system with $\Pi_{ij}^{(h_1)} = 0$ for all $h_1 > 1$, then it is:

$$\forall h_2 \in \mathbb{N} : \forall k \in \{1, \dots, K\} : a_{ik}\Pi_{kj}^{(h_2)} = 0.$$

We basically assume that if variable j is multi-step non-causal for variable i, then this is because there is no path from i to j and not because there is a path from i to j with VAR coefficients such that there is 'canceling-out'. Assumption 3.1 thus ensures the correspondence between paths and causality as in Lemma 3.1.

Lemma 3.1. Given a VAR(1) system and Assumption 3.1, for all $h \geq 1$:

$$\mathbb{P}_{ij}^{(h)} \neq \emptyset \quad \Leftrightarrow \quad \Pi_{ij}^{(h)} \neq 0. \tag{3.7}$$

Proof. See Appendix A.2. ∎

This result states that variable j is multi-step causal for variable i if and only if there is at least one path from variable i to variable j. Under Assumption 3.1, the strongly connected components can be now be interpreted very easily.

Lemma 3.2. Given a VAR(1) system and Assumption 3.1, the strongly connected components are sets of variables that are mutually causal.

Proof. See Appendix A.3. ∎

This follows immediately from the definition of a SCC as for each pair i and j, there is a path from i to j and from j to i.

Since we have proven Theorem 3.2 as well as Lemmas 3.1 and 3.2 for $p = 1$ only, we consider the companion form, that is, the VAR(1) representation of a general VAR(p) model (2.1) to generalize this result:

$$Y_t = \mathbf{A}Y_{t-1} + U_t, \tag{3.8}$$

where $Y_t = (y_t', y_{t-1}', \dots, y_{t-p+1}')'$ and $U_t = (u_t', 0, \dots, 0)'$ are $Kp \times 1$ vectors and:

$$\mathbf{A} := \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & 0 & 0 & \dots & 0 \\ 0 & I_K & 0 & \dots & 0 \\ 0 & 0 & \ddots & & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{pmatrix}. \tag{3.9}$$

The ultimate goal is to show that under mild conditions, the set of causal variables and the set of relevant variables coincide. For that purpose, we need another auxiliary result. Therefore, we first show that if Assumption 3.1 holds for the companion matrix (3.9), causality is equivalent to a non-zero entry in the very first autogressive matrix of the direct VAR(p) (3.3) for at least one horizon:

Lemma 3.3. Given a finite-order VAR(p) model (2.1), Assumption 3.1 for the companion matrix (3.9) of the corresponding VAR(1) representation (3.8) and $y^{\mathrm{I}} \subseteq y$:

$$\exists h_1 \in \mathbb{N} : y_j \to_{h_1} y^{\mathrm{I}} \quad \Leftrightarrow \quad \exists h_2 \in \mathbb{N} : \Pi_{\mathrm{I}j,1}^{(h_2)} \neq 0. \tag{3.10}$$

Proof. See Appendix A.4. ∎

The final step is to analyze multi-step causality across SCCs. Consider a strongly connected component C_i and, as in Section 2, denote the set of all SCCs that are reachable from C_i by $R_{G^{SCC}}(C_i)$. Then, due to Lemma 3.2, all variables in $R_{G^{SCC}}(C_i)$ are multi-step causal for variables in C_i as there is a path from any variable in C_i to the variables in $R_{G^{SCC}}(C_i)$.

Finally, based on the foregoing discussion and for a given set of variables of interest y^{I}, we note that the variables, which are multi-step causal for y^{I} are all the variables in the SCCs that contain elements of y^{I} and all variables in all SCCs that may be reached from these SCCs.[11] Note that under Assumption 3.1 for companion matrix (3.9), this coincides with Definition 2.4 of the minimal set of relevant variables $R(y^{\mathrm{I}})$ from Section 2. Moreover, remember Definition 3.1 of causal variables $C(y^{\mathrm{I}})$. In case of 'canceling-out', the set of relevant variables $R(y^{\mathrm{I}})$ will be larger than the set of causal variables $C(y^{\mathrm{I}})$. We summarize this result more formally in Theorem 3.3.

Theorem 3.3. Given a VAR(p) system, $C(y^{\mathrm{I}}) \subseteq R(y^{\mathrm{I}})$, that is, all variables that cause y^{I} are contained in the set of relevant variables. If Assumption 3.1 is true for companion matrix (3.9), then the set of relevant variables is identical to the set of causal variables, $C(y^{\mathrm{I}}) = R(y^{\mathrm{I}})$.

Proof. See Appendix A.5. ∎

This establishes the relation between the set of relevant variables found from the graph of strongly connected components and the variables that are multi-step causal for the variables of interest.

Remark 3.1. Given that Assumption 3.1 holds for the companion matrix of the VAR(1) representation of a general finite-order VAR(p) model, all results transfer to VAR(p) processes. We also note again that Assumption 3.1 is not restrictive at all as the 'canceling-out' effect will essentially never occur in practice.

Hence, we have shown that under very mild conditions, the set of relevant variables coincides with the set of variables that are causal for y^{I}. This is a useful result, as we note that the multi-step non-causality tests of Dufour *et al.* (2006) may be inapplicable when considering high-dimensional VARs because they require the estimation of large covariance matrices. Consequently, in large VARs they cannot be used as a variable selection tool. Our graphical approach is not affected by that shortcoming and can therefore be superior for some scenarios.

[11] Thereby, note that though we are considering system (3.8) of dimension Kp, the variables of interest y^{I} are still a subset of y, that is, $\mathrm{I} \subset \{1, \dots, K\}$.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

4. EMPIRICAL ILLUSTRATION

To illustrate the usefulness of the graph theoretical approach for selecting the relevant information set, we apply the method to a large set of US economic time series. We focus on the selection of variables and on impulse response analysis of the selected models.

We start from a set of 41 quarterly economic time series that includes a large variety of macroeconomic and financial series over a period from 1975Q1 to 2014Q4. The collection of variables is similar to related studies as, for example, Jarociński and Maćkowiak (2017) and Kascha and Trenkler (2015) and includes real GDP and its components, business cycle indicators, various price measures and interest rates, monetary aggregates and a number of labor market variables. In addition, the data includes exchange rate data together with three key variables for the Euro area (Euro area GDP, Euro area CPI and a Euro area interest rate). A detailed list with variables and data sources is provided in Table B1 in Appendix B.

To apply the approach discussed in Section 2, we first transform the data to stationarity. This involves taking logarithms and/or differences depending on the property of the respective variable.[12] We describe the details of data preparation and document the transformations by reporting the transformation codes in Appendix B.

In what follows, we apply the graph theoretical methods to a sparse VAR, that is, a VAR with a number of zero coefficients in the autoregressive matrices. In our application, these sparse VARs are selected by applying the least absolute selection and shrinkage operator (LASSO) in the context of the VAR model. There is ample evidence in the literature that LASSO is a useful device and often performs more precise than standard (unrestricted or subset) VARs (see e.g. Kascha and Trenkler, 2015 and references therein). While in principle other methods for subset selection may be employed, we only use LASSO and point out that the subset selection is not the main focus of our article. Instead, we start from a given subset structure and explore how this can be used to detect the smallest possible VAR system.

4.1. Variable Selection

To illustrate the variable selection, we choose real US GDP, the US consumer prices (CPI), and the federal funds rate as variables of interest y^{I}. This includes three key economic variables often analyzed with VARs and also corresponds to the variables chosen by Jarociński and Maćkowiak (2017). For a given sample period, we estimate a LASSO-VAR(4) in all 41 variables such that the system contains lags up to 1 year ($p = 4$). The shrinkage parameter in the LASSO approach is chosen by the Bayesian information criterion (BIC).[13] The variables that should be included in the VAR are then determined based on the SCCs according to Definition 2.4.

To investigate which of the 41 variables should be selected into a VAR and how the selection changes over time, we use our method recursively in an expanding window setup. The initial estimation period covers data from 1975Q1 to 2002Q2 (sample size $T_1 = 112$). We apply the LASSO-VAR and record the relevant variables according to the SCC structure as in Definition 2.4. We then add recursively observations to the estimation sample and re-estimate the LASSO-VAR with $T_1 + 1, T_1 + 2, \ldots, T$ observations, where the final estimation period ends in 2014Q4 (with a sample size of $T = 160$). For each estimation window considered in this recursive setup, we use the respective LASSO-VAR results to determine the relevant variables according to Definition 2.4. Thus, for each of the estimation periods ending between 2002Q4 and 2014Q4, we have a set of selected variables and show these selection results in graphical form in Figure 2. The rows in the checkerboard graph correspond to the different economic variables, whereas the columns refer to different ends of estimation samples. The filled green squares correspond to the variables of interest (here: GDP, CPI, and the Federal Funds Rate), a filled blue square in a specific row indicates that the variable in that row is selected into the minimal VAR using the sample terminating in the period corresponding to the column. Accordingly, a white square indicates that the variable has not been selected into the minimal VAR in a particular period.

[12] Transforming the variables to stationarity cancels possible common trends and cointegration relations between the variables. Extending the graphical methods to models with common trends such as cointegrated VARs and vector error correction models is left for future research.
[13] As for the implementation of the LASSO-VAR we follow the paper by Kascha and Trenkler (2015) and refer to their paper for details.

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

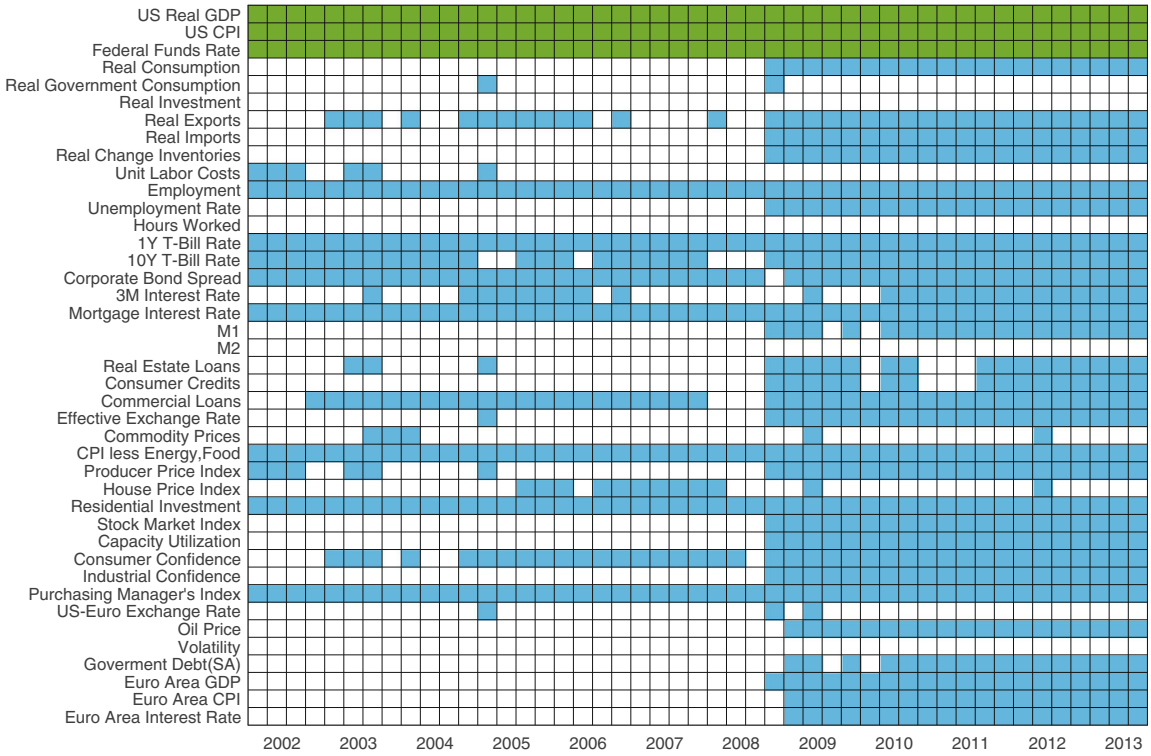


Figure 2. Variable selection results. Variables of interest y^{I}: US Real GDP, CPI and FFR (green). Relevant variables as selected by graphical method (blue) and variables not selected (white). Sample period: 1975Q1–2014Q4

We find that six variables are always selected into the minimal VAR in each of the periods considered. This includes employment, the 1-year T-Bill rate, the mortgage interest rate, the CPI less food and energy price index, residential investment and the purchasing manager's index. A tentative interpretation of this result may be that these variables form the minimal set of additional variables that should be considered if a VAR in output, inflation and interest rate is of interest. We note that some of these variables have typically not been included in related empirical studies on the effects of monetary policy shocks. We also compared our set of selected variables with those from Jarociński and Maćkowiak (2017) and note that the variables of interest y^{I} show a very low posterior probability (<0.1) of being Granger-causally prior to the variables (except residential investment) that have been always selected by our graphical methods. However, based on table 1 of Jarociński and Maćkowiak (2017), one would select 22 series since y^{I} has a posterior probability of less than 0.1 to be Granger-causal prior to them. Hence, though the variables selected by our approach are also chosen using the method of Jarociński and Maćkowiak (2017), the graphical technique applied in this article seems to be more in favor of a small VAR system of relevant variables.

In a number of periods additional variables are selected into the minimal VAR. The number of selected variables is quite large (on average 21.5 out of the 41 are selected). This provides evidence that the dynamic relationship between economic variables is more complex than small scale VARs tend to suggest. While there are some changes as we increase the estimation sample, the overall selection of variables is relatively stable for the period before the 2008/2009 economic crisis. Interestingly, when the estimation sample extends beyond the crisis period, we observe that our method tends to select more variables, possibly suggesting that the linkages between variables have become more pronounced.[14]

[14] We are grateful to one of the referees suggesting that the large number of selected variables may point to a factor structure of the underlying data. Consequently, controlling for common factors may be useful. We have done this by running our selection method after removing the

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

4.2. Impulse Response Analysis

We illustrate the effect of including the selected variables into a small VAR system on estimated impulse responses. As in Section 4.1, the small VAR consists of the variables of interest with real GDP, the CPI and the federal funds rate (FFR). These type of systems have also been used by Jarociński and Maćkowiak (2017) and Banbura *et al.* (2010).

Following standard specifications from the literature (see e.g. Christiano *et al.*, 1999), we have used VAR(4) models in the (log) levels of the variables of interest for the comparison of impulse responses. For the VAR with selected variables, we have added the six variables (again in (log) levels) that have been selected for all considered sample periods in Section 4.1. We use the standard ordering of variables and thus include employment, the CPI less food and energy price index, residential investment, and the purchasing manager's index in the group of 'slow moving' variables, that is, they are ordered above the federal funds rate variable. In contrast, the 1-year T-Bill rate and the mortgage interest rate are in the group of 'fast moving' variables and are consequently ordered below the federal funds rate. Using a Cholesky decomposition, this ordering implies that a shock in the federal funds rate (typically labeled a monetary policy shock) may have an immediate impact on the 'fast moving' variables, while the 'slow moving' variables may only react with a lag of one quarter. All VAR models are estimated by unrestricted multivariate LS (i.e. no shrinkage is applied) and the reported (pointwise) confidence intervals are asymptotic 95% intervals obtained using the 'delta method' (see e.g. Lütkepohl, 2005, section 3.7).[15]

In Figure 3, we report results for a sample that ends in 2008 to exclude the effects of the 2008/2009 financial crisis and to take into account the break in the variable selection after 2008 (see Figure 2). The left panel of the figure shows the responses of GDP, CPI and the federal funds rate to a contractionary shock in the federal funds rate within the small VAR containing three variables. We find the typical pattern with a significant and persistent drop in output. We also find a significant increase in CPI, which is known as the 'price puzzle' because economic theory suggests a decrease rather than an increase in the price level after tightening monetary policy. In other words, the response of the price level in the small VAR is counter-intuitive. Adding the six selected variables changes the response patterns substantially. First, the drop in output is now much less persistent. In fact, 2 years after the shock the response of output is no longer significantly different from zero. Moreover, the price puzzle disappears: the reaction of consumer prices to a monetary policy shock is not significantly different from zero for more than 3 years. Thereafter, it shows the expected sign since in line with conventional economic theory it starts to be significantly negative. Thus, including the variables selected by our method leads to much more reasonable impulse response patterns and changes the interpretation of the results substantially. Of course, in the context of our empirical illustration, it has already been shown that including for certain additional variables can mitigate the price puzzle. However, our method based on SCCs adds a new and more systematic way of how to expand small VAR systems. It is therefore helpful in finding the right information set, which is of obvious importance for structural analysis.

For comparison, we have also computed the responses estimated from a simple 41 variable LASSO-VAR(4) and show the results as dashed lines in Figure 3. Interestingly, the responses are fairly similar to those from the small 3-variable VAR. In particular, just using a LASSO-VAR does not help to resolve the price puzzle. This nicely illustrates the value-added of our procedure.

We report the responses for the full sample that ends after 2014 in Figure D1 in Appendix D. Again, the addition of our six selected variables clearly mitigates the price puzzle, even though it does not impose the reaction of prices to a monetary policy shock to be significantly negative.

---

first two principal components from the data. Using the remaining idiosyncratic components we find that almost no additional variables are selected. We therefore conclude that in our framework controlling for common factors is not helpful in choosing a good information set for a VAR analysis since too few variables are selected to add any value to the subsequent structural analysis.

[15] The construction of the impulse response intervals ignores the uncertainty related to the variable selection. In Appendix D.1, we suggest an empirical approach to get some indication for the effect of selection uncertainty. As expected, this leads to somewhat wider intervals, however, the main economic results are qualitatively not affected.

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

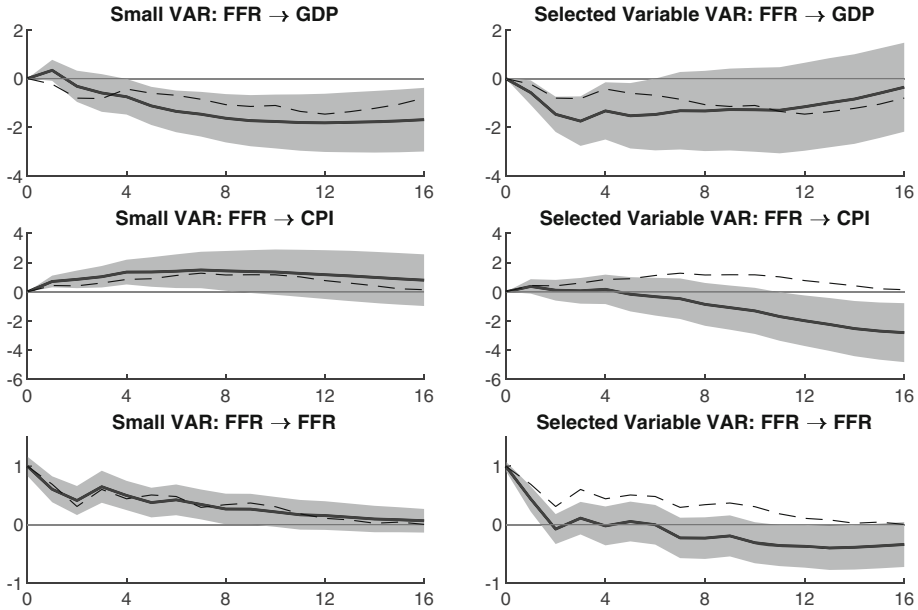


Figure 3. Impulse responses in small VAR and selected variable VAR. Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Right: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables. The dashed line shows the estimated responses from a LASSO-VAR(4). Sample period: 1975Q1–2008Q4

It is interesting to note that the changes in the response pattern obtained by just adding six variables are to some extent similar to the changes obtained by Banbura *et al.* (2010) in their medium (20 variables) and large (131 variables) system for monthly data. In other words, it seems that our methodology provides a perspective that complements the 'large VAR' idea of Banbura *et al.* (2010).

5. CONCLUSION

This article uses concepts from graph theory for variable selection in VAR models. To this end, we identify strongly connected components from the directed graph representing the dynamic relationships among the variables in a sparse VAR. We suggest to use relations among the strongly connected components in a so-called component graph to identify a minimal set of variables that we need to include in a VAR analysis for a small set of variables if impulse response analysis is of interest. The article contributes to the existing literature by introducing a graphical method, which to the best of our knowledge has not been used for variable selection in econometrics.

We also show that there is a simple relation between the graph theoretical concept and multi-step causality and relate the paths in the graph to coefficients of a direct VAR system. It follows from the results in the article that the set of relevant variables selected from the graphical approach coincides with the set of variables that are multi-step causal for the variables of interest.

We illustrate the usefulness of the variable selection method in a structural analysis of a small US monetary system (real GDP, CPI inflation and the federal funds rate) as the variables of interest. Given this set, we apply the graphical approach to select additional variables out of a large set of macroeconomic variables. The selected VAR typically includes some variables from the real sector (employment and residential investment), a forward looking indicator (purchasing manager's index), different interest rates (mortgage interest rate and 1 year T-Bill rate) and a CPI related measure (CPI less food and energy). The selection of variables seems sensible from an economic point of view. Interestingly, we find that this list includes some variables that other researchers typically have not included in small monetary systems. Moreover, we find that including the selected variables for impulse response

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

analysis is useful: In the small monetary system in output, inflation and interest rate, we find that including the selected variables avoids the so-called 'price puzzle'.

Overall, our empirical results suggest that using graphical modeling for variable selection is a useful addition to the VAR econometricians' toolbox. The method complements existing methods for large data sets and is particularly useful if a researcher prefers to work with smaller scale models, for example, for maintaining consistency with small scale theoretical models. Moreover, compared to alternative methods for large data sets, a graphical representation of the strongly connected components may give useful insights on the (causal) relationships and the transmission channels among the VAR variables.

Extensions of the current article could use graphical models for variable selection taking also the contemporaneous relationships among variables into account. Moreover, extending the approach to models with integrated and cointegrated variables would be of interest. We leave this for future research.

ACKNOWLEDGEMENTS

We thank participants of the SFB 649 final colloquium at Humboldt-University Berlin, of the University of Kiel Econometrics Seminar, the 2019 Annual Congress of the Verein für Socialpolitik and of the Conference on Computational and Financial Econometrics for useful comments on earlier versions of this article. Part of this research has been conducted while the first author was a doctoral student at the Department of Economics at the University of Konstanz, Germany and the third author was a postdoctoral research at the Department of Economics at the University of Zurich, Switzerland. Financial support by the Deutsche Forschungsgemeinschaft, Grant number BR 2941/3-1 is gratefully acknowledged. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available by the Federal Reserve Economic Data at https://fred.stlouisfed.org/.

REFERENCES

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian vector autoregressions. *Journal of Applied Econometrics* **25**:71–92.

Barigozzi M, Brownlees C. 2019. NETS: network estimation for time series. *Journal of Applied Econometrics* **34**:347–364.

Barigozzi M, Hallin M. 2017. A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society Series C – Applied Statistics* **66**:581–605.

Basu S, Michailidis G. 2015. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43**:1535–1567.

Basu S, Das S, Michailidis G, Purnanandam A. 2019. A system-wide approach to measure connectivity in the financial sector. SSRN 2816137.

Belloni A, Chernozhukov V. 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19**:521–547.

Bernanke BS, Boivin J, Eliasz P. 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* **120**:387–422.

Brillinger DR. 1996. Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16**:1–23.

Carriero A, Kapetanios G, Marcellino M. 2009. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting* **25**:400–417.

Carriero A, Kapetanios G, Marcellino M. 2012. Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking and Finance* **36**:2026–2047.

Carriero A, Clark TE, Marcellino M. 2015. Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics* **30**:46–73.

Cheng X, Hansen BE. 2015. Forecasting with factor-augmented regression: a frequentist model averaging approach. *Journal of Econometrics* **186**:280–293.

Christiano LJ, Eichenbaum M, Evans CL. 1999. *Monetary policy shocks: what have we learned and to what end?* . In *Handbook of Macroeconomics*, Vol. 1, Taylor JB, Woodford M (eds.). Elsevier, Amsterdam; 65–148, chap. 2.

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Clements MP. 2016. Real-time factor model forecasting and the effects of instability. *Computational Statistics and Data Analysis* **100**:661–675.
Cormen TH, Leiserson CE, Rivest RL, Stein C. 2009. *Introduction to Algorithms* The MIT Press, Cambridge, MA, chap. 22.
Dahlhaus R. 2000. Graphical interaction models for multivariate time series. *Metrika* **51**:157–172.
Dahlhaus R, Eichler M. 2003. *Causality and graphical models for time series*. In *Highly Structured Stochastic Systems*, Green P, Hjort N, Richardson S (eds.). Oxford University Press, Oxford; 115–137.
Davis RA, Zang PF, Zheng T. 2016. Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* **25**:1077–1096.
Demiralp S, Hoover KD. 2003. Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics* **65**:745–767.
Doan TA , Todd R. 2010. Causal ordering for multivariate linear systems. Unpublished work, Estima.
Duff IS, Reid J. 1978. An implementation of Tarjan's algorithm for the block triangularization of a matrix. *ACM Transactions on Mathematical Software* **4**:137–147.
Dufour JM, Renault E. 1998. Short run and long run causality in time series: theory. *Econometrica* **66**:1099–1125.
Dufour JM, Pelletier D, Renault E. 2006. Short run and long run causality in time series: inference. *Journal of Econometrics* **132**:337–362.
Edwards D. 2000. *Introduction to Graphical Modelling*. Springer Texts in Statistics, 2nd ed. Springer-Verlag, New York.
Eichler M. 2006. *Graphical modeling of dynamic relationships in multivariate time series*. In *Handbook of Time Series Analysis: Recent Developments and Applications*, Schelter B, Winterhalder M, Timmer J (eds.). Wiley, London.
Eichler M. 2007. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* **137**:334–353.
Eichler M. 2012. Graphical modelling of multivariate time series. *Probability Theory and Related Fields* **153**:233–268.
Eickmeier S, Ziegler C. 2008. How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting* **27**:237–265.
Flamm C, Kalliauer U, Deistler M, Waser M, Graef A. 2012. *Graphs for dependence and causality in multivariate time series*. In *System Identification, Environmental Modelling, and Control System Design*, Wang L, Garnier H (eds.). Springer London, London; 133–151.
Giannone D, Lenza M, Momferatou D, Onorante L. 2014. Short-term inflation projections: a Bayesian vector autoregressive approach. *International Journal of Forecasting* **30**:635–644.
Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**:424–438.
Hall P. 1992. *The Bootstrap and Edgeworth Expansion* Springer, New York.
Heinlein R, Krolzig HM. 2012. Effects of monetary policy on the US dollar/UK pound exchange rate. Is there a 'delayed overshooting puzzle'? *Review of International Economics* **20**:443–467.
Hoover K, Demiralp S, Perez SJ. 2009. *Empirical identification of the vector autoregression: the causes and effects of U.S. M2*. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Castle J, Shepard N (eds.). Oxford University Press, Oxford; 37–58.
Jarociński M, Maćkowiak B. 2017. Granger causal priority and choice of variables in vector autoregressions. *Review of Economics and Statistics* **99**:319–329.
Javanmard A, Montanari A. 2014. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research* **15**:2869–2909.
Kascha C , Trenkler C. 2015. Forecasting VARs, model selection, and shrinkage. Working Paper ECON.
Kilian L. 1998. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis* **19**:531–548.
Kock AB, Callot L. 2015. Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**:325–344.
Koop GM. 2013. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* **28**:177–203.
Lauritzen SL. 1996. *Graphical Models*. Oxford Statistical Science Series Oxford University Press, Oxford.
Ludvigson SC, Ng S. 2007. The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics* **83**:171–222.
Ludvigson SC, Ng S. 2009. Macro factors in bond risk premia. *Review of Financial Studies* **22**:5027–5067.
Lütkepohl H. 1993. *Testing for causation between two variables in higher dimensional VAR models*. In *Studies in Applied Econometrics*, Schneeweiß H, Zimmermann KF (eds.). Springer-Verlag, Heidelberg; 75–91.
Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis* Springer Science & Business Media, Berlin, Heidelberg.
McCracken MW, Ng S. 2016. FRED-MD: a monthly database for macroeconomic research. *Journal of Business & Economic Statistics* **34**:574–589.
Medeiros MC, Mendes EF. 2016. l(1)-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics* **191**:255–271.
Pearl J. 2000. *Causality* Cambridge University Press, New York.

Penm J, Terrell R. 1986. *The 'Derived' Moving-average Model and Its Role in Causality* Applied Probability Trust, Sheffield; 99–111.
Sims CA. 1980. Macroeconomics and reality. *Econometrica* **48**:1–48.
Sims CA. 1982. Policy analysis with econometric-models. *Brookings Papers on Economic Activity* **1**: 107–164.
Sims CA. 2015. Causal orderings and exogeneity. Lecture Notes. http://sims.princeton.edu/yftp/Times15F/GCP15.pdf.
Stock JH, Watson MW. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**:1167–1179.
Stock JH, Watson MW. 2003. Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature* **41**:788–829.
Stock JH, Watson MW. 2012. Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics* **30**:481–493.
Stock JH, Watson MW. 2016. *Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics*. In *Handbook of Macroeconomics*, Vol. 2, Taylor JB, Uhlig H (eds.). Elsevier, Amsterdam; 415–525, chap. 8.
Swanson NR, Granger CWJ. 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association* **92**:357–367.
Tarjan RE. 1972. Depth-first search and linear graph algorithms. *SIAM Journal on Computing* **1**:146–160.
Uhlig H. 2009. *Comment on 'How has the Euro changed the monetary transmission mechanism?'*. In *NBER Macroeconomics Annual 2008*, Vol. 23, Acemoglu D, Rogoff K, Woodford M (eds.). University of Chicago Press, Chicago, IL; 141–152.
Van de Geer S, Bühlmann P, Ritov Y, Dezeure R. 2014. On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics* **42**:1166–1202.
Zhang CH, Zhang SS. 2014. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B – Statistical Methodology* **76**:217–242.
Zheng LL, Raskutti G. 2019. Testing for high-dimensional network parameters in auto-regressive models. *Electronic Journal of Statistics* **13**:4977–5043.

APPENDIX A: PROOFS

A.1. Proof of Theorem 3.2

Proof. We prove Theorem 3.2 by mathematical induction.

Base Case Let $h = 1$. Then, it is $\Pi_{ij}^{(h)} = \Pi_{ij}^{(1)} = a_{ij}$.

If $\mathbb{P}_{ij}^{(1)}$ is empty, $(i,j) \notin E$ what imposes $a_{ij} = 0$ because of Definition 2.1. Then, equality (3.6) holds trivially.

If $\mathbb{P}_{ij}^{(1)} \neq \emptyset$, it follows that $\mathbb{P}_{ij}^{(1)} = \{(i,j)\}$. Consequently, $\sum_{P \in \mathbb{P}_{ij}^{(1)}} \prod_{(l,m)\in P} a_{lm} = a_{ij} = \Pi_{ij}^{(1)}$, which proves (3.6) for $h = 1$.

Induction Step Let (3.6) hold for $h-1$, that is

$$\forall i,j \in \{1, \dots, K\} : \Pi_{ij}^{(h-1)} = \sum_{P \in \mathbb{P}_{ij}^{(h-1)}} \prod_{(l,m)\in P} a_{lm}. \tag{A1}$$

Moreover, define:

$$E_i := \{k \in \{1, \dots, K\} : (i,k) \in E\} \quad \Rightarrow \quad \mathbb{P}_{ij}^{(h)} = \bigcup_{k \in E_i} \bigcup_{P \in \mathbb{P}_{kj}^{(h-1)}} [\{(i,k)\} \cup P]. \tag{A2}$$

This just means that any path from vertex i to vertex j of length h can be decomposed into a tuple (i,k) and a path from vertex k to vertex j of length $(h-1)$ for all $k \in \{1, \dots, K\}$ with $a_{ik} \neq 0$ and $\mathbb{P}_{kj}^{(h-1)} \neq \emptyset$. Using this, we get:

$$\Pi_{ij}^{(h)} = \sum_{k=1}^{K} a_{ik} \Pi_{kj}^{(h-1)}$$

$$\begin{aligned}
&\overset{(A1)}{=} \sum_{k=1}^{K} a_{ik} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} \prod_{(l,m)\in P} a_{lm} \\
&= \sum_{k \in E_i} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} a_{ik} \prod_{(l,m)\in P} a_{lm} \\
&= \sum_{k \in E_i} \sum_{P \in \mathbb{P}_{kj}^{(h-1)}} \prod_{(l,m)\in[\{(i,k\}\cup P]} a_{lm} \\
&\overset{(A2)}{=} \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m)\in P} a_{lm}.
\end{aligned}$$

■

A.2. Proof of Lemma 3.1

Proof. We prove both directions of the if-and-only-if statement (3.7):

"$\overset{(A.3)}{\Rightarrow}$" Let $\mathbb{P}_{ij}^{(h)} \neq \emptyset$. We prove $\Pi_{ij}^{(h)} \neq 0$ by mathematical induction.

Base Case Let $h = 1$. Then, it is $\mathbb{P}_{ij}^{(1)} = \{(i,j)\}$, that is, $a_{ij} \neq 0$ due to Definition 2.1. Consequently, it is $\Pi_{ij}^{(1)} = a_{ij} \neq 0$.

Induction Step Let one direction of (3.7) hold for $(h-1)$, that is

$$\forall i,j \in \{1, \dots K\} \text{ with } \mathbb{P}_{ij}^{(h-1)} \neq 0, \text{ it is } \Pi_{ij}^{(h-1)} \neq 0. \tag{A3}$$

Now, let $\mathbb{P}_{ij}^{(h)} \neq \emptyset$:

$$\begin{aligned}
&\Rightarrow \exists k \in \{1, \dots, K\} : a_{ik} \neq 0 \wedge \mathbb{P}_{kj}^{(h-1)} \neq \emptyset \\
&\overset{(A.3)}{\Rightarrow} \Pi_{kj}^{(h-1)} \neq 0 \Rightarrow a_{ik}\Pi_{kj}^{(h-1)} \neq 0 \\
&\Rightarrow \Pi_{ij}^{(h)} = \sum_{l=1}^{K} a_{il}\Pi_{lk}^{(h-1)} \neq 0
\end{aligned}$$

because of Assumption 3.1 and $a_{ik}\Pi_{kj}^{(h-1)} \neq 0$.

"$\Leftarrow$" Let $\Pi_{ij}^{(h)} \neq 0$. Assume $\mathbb{P}_{ij}^{(h)} = \emptyset$. Consequently, by Theorem 3.2 it is $\Pi_{ij}^{(h)} = \sum_{P \in \mathbb{P}_{ij}^{(h)}} \prod_{(l,m)\in P} a_{lm} = 0$ because it is an empty sum. This is a contradiction. Thus, $\mathbb{P}_{ij}^{(h)} \neq \emptyset$ has to hold. ■

A.3. Proof of Lemma 3.2

Proof. Without loss of generality, let $i,j \in C_k$. By Definition 2.2, y_i and y_j are mutually reachable. Therefore, it is:

$$\begin{aligned}
&\exists h_i \in \mathbb{N} : \mathbb{P}_{ij}^{(h_i)} \neq \emptyset \overset{\text{Lemma 3.1}}{\Leftrightarrow} \Pi_{ij}^{(h_i)} \neq 0 \overset{\text{Theorem 3.1}}{\Leftrightarrow} y_j \to_{h_i} y_i \Rightarrow y_j \in C\left(y_i\right), \\
&\exists h_j \in \mathbb{N} : \mathbb{P}_{ji}^{(h_j)} \neq \emptyset \overset{\text{Lemma 3.1}}{\Leftrightarrow} \Pi_{ji}^{(h_j)} \neq 0 \overset{\text{Theorem 3.1}}{\Leftrightarrow} y_i \to_{h_j} y_j \Rightarrow y_i \in C\left(y_j\right).
\end{aligned}$$

Consequently, y_i and y_j are mutually causal. ■

DOI: 10.1111/jtsa.12664

A.4. *Proof of Lemma 3.3*

Proof. We prove both directions of the if-and-only-if statement (3.10) whereas the second implication is trivial:

"⇒" Let y_j be a variable that causes y^{I}, that is, $\exists h_1 \in \mathbb{N} : y_j \to_{h_1} y^{\mathrm{I}}$. By Theorem 3.1 it follows:

$$\exists s_1 \in \{1, \dots, p\} : \Pi^{(h_1)}_{\mathrm{I}j,s_1} \neq 0 \Rightarrow \exists i \in \mathrm{I} : \Pi^{(h_1)}_{ij,s_1} \neq 0.$$

For $s_1 = 1$, $\Pi^{(h_1)}_{ij,1} \neq 0$ and thus $\Pi^{(h_1)}_{\mathrm{I}j,1} \neq 0$ follows immediately. So, let $s_1 \in \{2, \dots, p\}$.
Note that one can easily show that $J\mathbf{A}^h = (\Pi^{(h)}_1, \dots, \Pi^{(h)}_p)$ with $J = (I_K, 0, \dots, 0)$ being a $(K \times Kp)$ selection matrix. Consequently, $\Pi^{(h)}_{ab,s} = \mathbf{A}^h_{a,(s-1)K+b}$ holds for all $a, b \in \{1, \dots, K\}$, $s \in \{1, \dots, p\}$ and $h \in \mathbb{N}$.
Therefore, $\mathbf{A}^{h_1}_{i,(s_1-1)K+j} = \Pi^{(h_1)}_{ij,s_1} \neq 0$ holds. Moreover, $\mathbf{A}_{l,l-K} = 1$ for $l = K+1, \dots, Kp$ by construction of the companion matrix (3.9). Applying this argument for $l = (s_1 - 1)K + j$ and using Assumption 3.1, it follows that $\Pi^{(h_1+1)}_{ij,s_1-1} = \mathbf{A}^{h_1+1}_{i,(s_1-2)K+j} \neq 0$. Continuing this argument, one shows that $\Pi^{(h_1+s_1-1)}_{ij,1} = \mathbf{A}^{h_1+s_1-1}_{ij} \neq 0$. So, $\Pi^{(h_1+s_1-1)}_{\mathrm{I}j,1} \neq 0$ is true.

"⇐" Assume that there is a $h \in \mathbb{N}$ such that $\Pi^{(h)}_{\mathrm{I}j,1} \neq 0$. Due to Theorem 3.1, this directly implies $y_j \to_h y^{\mathrm{I}}$. ∎

A.5. *Proof of Theorem 3.3*

Proof. First, we prove the general subset relation $C\left(y^{\mathrm{I}}\right) \subseteq R\left(y^{\mathrm{I}}\right)$.

"⊆" Let $y_j \in C\left(y^{\mathrm{I}}\right)$ and w.l.o.g. $j \in C_{j^*}$. Due to Definition 3.1, either

(a) $y_j \in y^{\mathrm{I}}$, or
(b) $\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}}$ has to hold.

In case (a), $y_j \in C_{j^*} \cap y^{\mathrm{I}}$ and therefore $C_{j^*} \cap y^{\mathrm{I}} \neq \emptyset$. By Definition 2.4, $y_j \in R_{G^{SCC}}\left(C_{j^*}\right)$ has to hold what directly imposes $y_j \in R\left(y^{\mathrm{I}}\right)$.
In case (b):

$$\exists h \in \mathbb{N} : y_j \to_h y^{\mathrm{I}} \overset{\text{Theorem 3.1}}{\Leftrightarrow} \mathbf{A}^h_{\mathrm{I}j} \neq 0 \Leftrightarrow \exists i \in \mathrm{I} : \mathbf{A}^h_{ij} \neq 0 \overset{\text{Theorem 3.2}}{\Rightarrow} \exists i \in \mathrm{I} : \mathbb{P}^{(h)}_{ij} \neq \emptyset.$$

Let $i \in C_{i^*}$. Obviously, j is *reachable* from i in G. Hence, C_{j^*} is also *reachable* from C_{i^*} in G^{SCC}. Consequently, it is $y_j \in R_{G^{SCC}}\left(C_{i^*}\right)$. Because of $i \in C_{i^*}$ and $i \in \mathrm{I}$, it is $y_i \in y^{\mathrm{I}} \cap C_{i^*}$ and therefore $y^{\mathrm{I}} \cap C_{i^*} \neq \emptyset$. Thus, $R_{G^{SCC}}\left(C_{i^*}\right) \subseteq R\left(y^{\mathrm{I}}\right)$ and therefore $y_j \in R\left(y^{\mathrm{I}}\right)$.

"⊇" Let Assumption 3.1 hold, $y_j \in R\left(y^{\mathrm{I}}\right)$ and w.l.o.g. $j \in C_{j^*}$. Due to Definition 2.4, then:

$$\exists i^* \in \{1, \dots, k\} : C_{i^*} \cap y^{\mathrm{I}} \neq \emptyset : y_j \in R_{G^{SCC}}\left(C_{i^*}\right).$$

Again by Definition 2.4, this means that either

(i) $j \in C_{i^*}$, or
(ii) $C_{i^*} \overset{P}{\rightsquigarrow} C_{j^*}$ has to hold.

In case (i):

$$C_{i^*} \cap y^{\mathrm{I}} \neq \emptyset \Rightarrow \exists i \in \mathrm{I} : i \in C_{i^*} \Rightarrow i, j \in C_{i^*} \overset{\text{Lemma 3.2}}{\Rightarrow} y_j \to_h y_i \Rightarrow y_j \in C\left(y^{\mathrm{I}}\right).$$

Note that because of Remark 2.3, $j \in C_{i^*}$ would also imply $i^* = j^*$ and thereby cause $y_j \in C\left(y^{\mathrm{I}}\right)$. In case (ii):

$$\exists i_1 \in C_{i^*} : \exists j_1 \in C_{j^*} : \exists h \in \mathbb{N} : \mathbb{P}^{(h)}_{i_1 j_1} \neq \emptyset. \tag{A4}$$

As Assumption 3.1 holds for companion form (3.9), we can apply Lemma 3.2 and make use of the fact that the upper left $K \times K$ block of the autoregressive matrix of the direct VAR representation of companion form (3.8), $\mathbf{A}^h$, is identical with $\Pi_1^{(h)}$ of the corresponding direct representation of the regular VAR(p) model (2.1) to get:

$$\forall i_1 \in C_{i^*} : \exists h_1 \in \mathbb{N} : y_{i_1} \to_{h_1} y_i \overset{\text{Lemma 3.3}}{\Leftrightarrow} \exists h_2 \in \mathbb{N} : \Pi^{(h_2)}_{i,i_1,1} \neq 0 \Leftrightarrow \mathbf{A}^{h_2}_{i,i_1} \neq 0 \overset{\text{Lemma 3.1}}{\Leftrightarrow} \mathbb{P}^{(h_2)}_{i,i_1} \neq \emptyset,$$
$$\forall j_1 \in C_{j^*} : \exists h_3 \in \mathbb{N} : y_j \to_{h_3} y_{j_1} \overset{\text{Lemma 3.3}}{\Leftrightarrow} \exists h_4 \in \mathbb{N} : \Pi^{(h_4)}_{j_1,j,1} \neq 0 \Leftrightarrow \mathbf{A}^{h_4}_{j_1,j} \neq 0 \overset{\text{Lemma 3.1}}{\Leftrightarrow} \mathbb{P}^{(h_4)}_{j_1,j} \neq \emptyset.$$

As we have paths from vertex i to i_1, vertex i_1 to j_1 and vertex j_1 to j, there is also a path from vertex i to j as we can simply connect them, that is

$$\mathbb{P}^{(h+h_2+h_4)}_{ij} \neq \emptyset \overset{\text{Lemma 3.3}}{\Leftrightarrow} \exists h_5 \in \mathbb{N} : \Pi^{(h_5)}_{ij,1} \neq 0 \Leftrightarrow \Pi^{(h_5)}_{Ij,1} \neq 0 \overset{\text{Theorem 3.1}}{\Leftrightarrow} y_j \to_{h_5} y^{\mathrm{I}} \Rightarrow y_j \in C\left(y^{\mathrm{I}}\right). \quad \blacksquare$$

APPENDIX B: DATA

We describe the data used in the empirical illustrations. Raw data for most series are obtained from the FRED database and Table B1 shows the corresponding FRED mnemonics. We construct some variables from splicing two series to obtain long time series: As a measure for the exchange rate, we use the US/DM exchange rate (`EXGEUS`) until 1998Q4. From 1999Q1 we use `EXUSEU` and splice both series accordingly. The resulting variable is called `EXCH`. We follow McCracken and Ng (2016) and use `OILPRICE` (Spot Oil Price) until 1985Q4 and `MCOILWTICO` (Crude Oil Price, Cushing) since 1986Q1, since the former series has been discontinued. The resulting series is labeled `POIL` in our data set. To obtain a crude measure of stock market volatility, we simply use the squared stock market returns, since the time series of volatility indices in FRED are rather short. This series is called `VOLA`. Seasonally adjusted series have been taken from FRED where necessary. The time series on Government Debt (`GFDEBTN`) has been seasonally adjusted by the authors using X-ARIMA-13. The resulting series is `GFDEBTNSA`. The Euro area time series have been added using the update 15 to the AWM database maintained at the ECB. The AWM mnemonics for the real GDP, CPI, and a short-term interest rates are `YER`, `HICP`, and `STN`. `HICP` has been seasonally adjusted by the authors using X-ARIMA-13. We use `EMUGDP`, `EMUHICPSA`, and `EMURS` to denote the three Euro area variables.

The last columns in Table B1 lists the transformation codes 1–6, corresponding to the following transformations of the series y_t: (1) no transformation, y_t, (2) Δy_t, (3) $\Delta^2 y_t$, (4) $400 \times \log(y_t)$, (5) $400 \times \Delta \log(y_t)$, (6) $400 \times \Delta^2 \log(y_t)$.

APPENDIX C: ALGORITHM

Input: raw data set y, lag length p, set of interest I

1. Transform y to get y_{trans} (transformation codes see Table B1) in Appendix B.
2. Estimate LASSO-VAR(p) on y_{trans} to obtain autoregressive matrices $\hat{A}_1^{\text{LASSO}}, \ldots, \hat{A}_p^{\text{LASSO}}$ (Kascha and Trenkler, 2015).

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

Table B1. Variables, data sources and transformations: Graph-VAR

Name	Mnemonic	Transf.Code
Real GDP	GDPC96	5
CPI	CPIAUCSL	6
Federal Funds Rate	FEDFUNDS	2
Real Consumption	PCECC96	5
Real Government Consumption	GCEC1	5
Real Investment	GPDIC1	5
Real Exports	EXPGSC1	5
Real Imports	IMPGSC1	5
Change in Real Inventories	CBIC96	1
Unit Labor Cost	ULCNFB	5
Employment	PAYEMS	5
Unemployment Rate	UNRATE	2
Hours worked	HOHWMN02USQ065S	1
1-year T-Bill Rate	GS1	2
10-year T-Bill Rate	GS10	2
Corporate Bond Spread	AAAFFM	1
Lending Rate to NFCs	TB3MS	3
Mortgage Rate	MORTG	2
M1	M1SL	6
M2	M2SL	6
Government Debt GFDEBTNSA	GFDEBTN. seas.adj: X-13	5
Real Estate Loans	REALLN	5
Consumer Credits	TOTALSL	5
Commercial Loans	BUSLOANS	5
Dollar/Euro Exchange Rate (EXCH)	spliced from EXGEUS and EXUSEU	5
Effective Exchange Rate	NNUSBIS	5
Oil Price (POIL)	spliced from OILPRICE and MCOILWTICO	5
Commodity Prices	CUSR0000SAC	6
Consumer Prices (excl. food, energy)	CPILFESL	6
Producer Price Index	PPIACO	5
House Prices	USSTHPI	6
Real Housing Investment	PRFI	5
Total Share Prices	SPASTT01USQ661N	5
Volatility Index	VIXCLS	5
Capacity Utilization	CUMFNS	2
Consumer Confidence	CSCICP03USM665S	2
Industrial Confidence	BSCICP03USM665S	2
Purchasing Manager's Index	NAPM	1
Real GDP (Euro Area)	AWM mnemonic: YER	5
CPI (Euro Area)	AWM mnemonic: HICP, seas.adj: X-13	6
Short term interest rate (Euro Area)	AWM mnemonic: STN	2

Note: The table shows FRED and AWM database names together with the transformation codes. See Appendix B for a detailed description of the transformations.

DOI: 10.1111/jtsa.12664 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

3. Based on $\hat{A}_1^{\text{LASSO}}, \ldots, \hat{A}_p^{\text{LASSO}}$ find SCCs $C_1, \ldots, C_k$ (Tarjan, 1972).
4. Based on $C_1, \ldots, C_k$ find minimal VAR $R\left(y_{\text{trans}}^{\text{I}}\right)$ (Definition 2.4).
5. Estimate an unrestricted VAR(p) in levels on $R\left(y^{\text{I}}\right)$.
6. Conduct impulse response analysis.

Output: relevant variables $R\left(y_{\text{trans}}^{\text{I}}\right)$, estimated impulse response functions

APPENDIX D: ADDITIONAL RESULTS

D.1. Selection Uncertainty and Impulse Response Intervals

The construction of impulse response confidence intervals in the selected VAR of Section 4.2 does not account for the uncertainty in the variable selection process. While a theoretical analysis of this issue is beyond the scope of this article, we try to get some indication of the underlying uncertainty in our empirical application. To this end, we use a residual bootstrap procedure, which is designed to empirically capture the selection uncertainty. The approach is similar to the 'endogenous lag order bootstrap', which has been used for capturing uncertainty in VAR lag selection (see e.g. Kilian, 1998). Using the parameter estimates and residuals of the Selected VAR, we generate $B = 999$ bootstrap samples. In each bootstrap replication, we repeat the variable selection approach of the article to determine the set of relevant variables. Based on the VAR in the newly selected variables, we compute the impulse responses of interest. The confidence intervals are constructed using Hall's percentile method (see Hall, 1992) and Appendix D.3 in Lütkepohl (2005)) with nominal coverage of 95%. As the selection is repeated in each bootstrap replication, the results give some indication of the selection uncertainty.

Using our empirical application from Section 4.2, we illustrate the effect of accounting the selection uncertainty with the above procedure. The first two columns in Figure D2 reproduce the result for the baseline sample

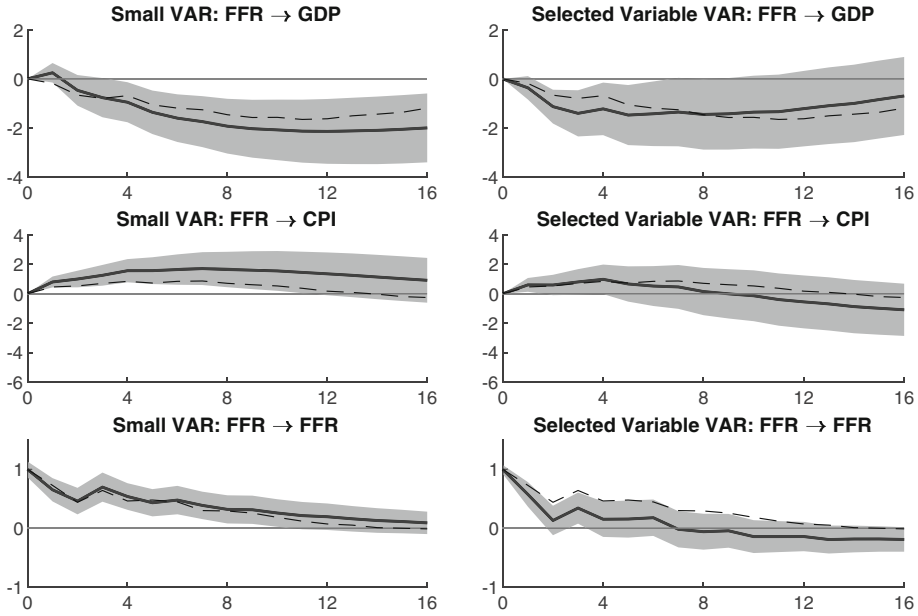


Figure D1. Impulse responses in small VAR and selected variable VAR. Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Right: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables. The dashed line shows the estimated responses from a LASSO-VAR(4). Sample period: 1975Q1–2014Q4

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664

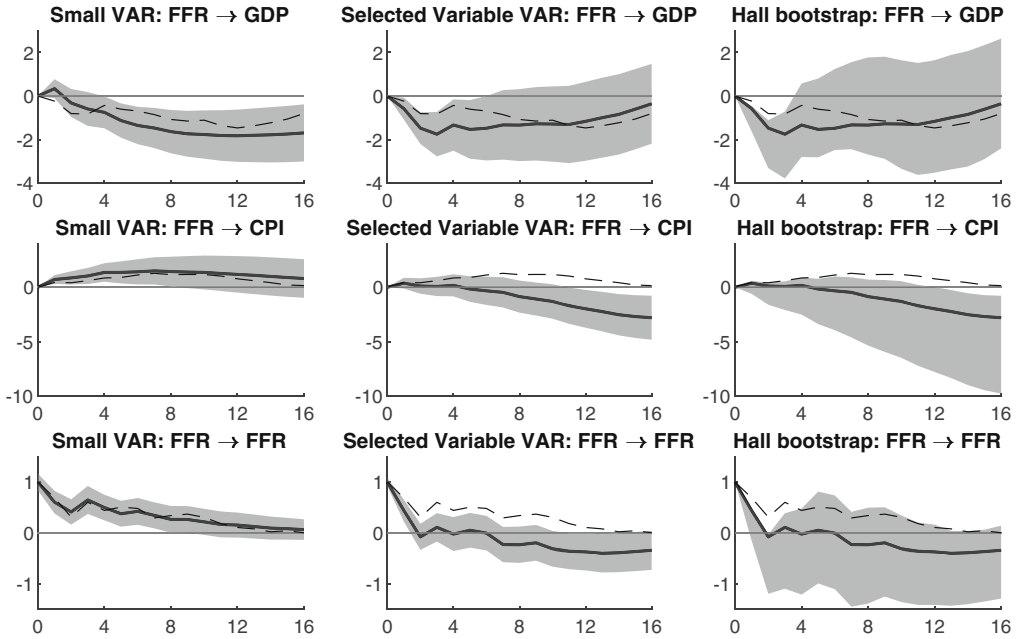


Figure D2. Selection uncertainty in impulse responses from selected variable VAR. Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Middle: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables. Right: Responses to a shock in FFR in selected VAR with bootstrap intervals accounting for selection uncertainty. The dashed line shows the estimated responses from a LASSO-VAR(4). Sample period: 1975Q1–2008Q4

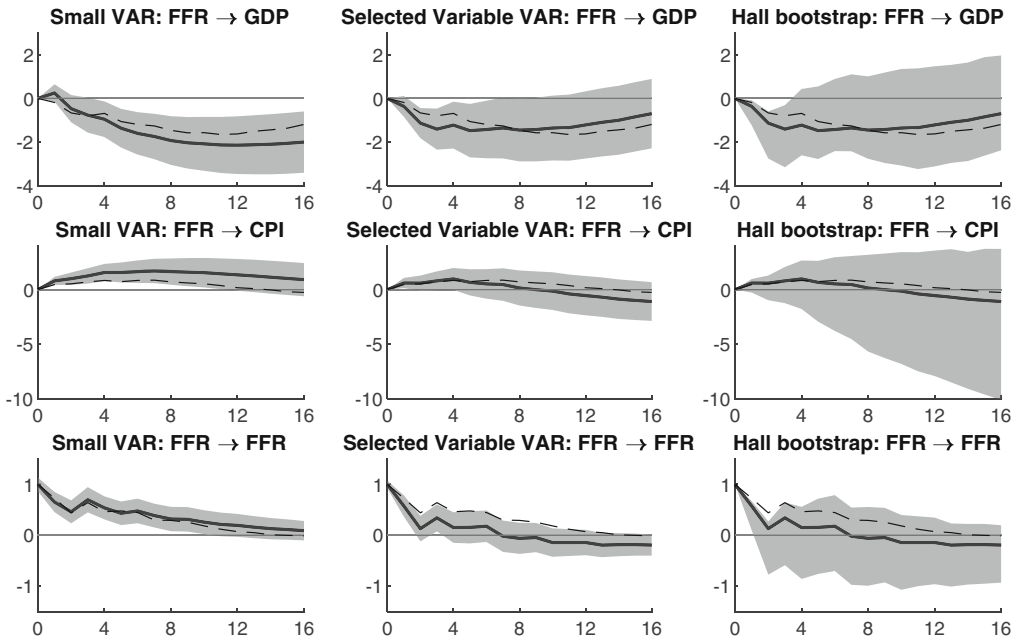


Figure D3. Selection uncertainty in impulse responses from selected variable VAR. Left: Responses to a shock in FFR in 3-variable (small) VAR(4) including GDP, FFR, and CPI. Middle: Responses to a shock in FFR in 9-variable VAR(4) with six selected additional variables. Right: Responses to a shock in FFR in selected VAR with bootstrap intervals accounting for selection uncertainty. The dashed line shows the estimated responses from a LASSO-VAR(4). Sample period: 1975Q1–2014Q4

1975Q1–2008Q4 from Figure 3, while the last column shows the confidence intervals accounting for selection uncertainty. As expected, these intervals are somewhat wider. We note, however, that the economic results are qualitatively very similar. For the extended sample (1975Q1–2014Q4, originally shown in Figure D1), a similar pattern can be observed (see Figure D3). Interestingly, for this extended sample, accounting for selection uncertainty increases the widths of the intervals more than in our baseline sample. Nevertheless, against the background of these results, we still find that our approach successfully mitigates and helps to solve the price puzzle.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12664