Identifying the Causal Structure of Directed Acyclic Graphs (DAGs)

Arnab Bhattacharjee*
Spatial Economics & Econometrics Centre (SEEC)
Heriot-Watt University, UK

23rd December 2017

Abstract

Recursive structures defined on DAGs (or Bayesian networks) are about the most general class of graphs, in the sense of highest graph density, on which causal models can be based. We develop an identification result based on relative variances that recovers the recursive causal graph for a collection of random variables placed at the vertices. The recursive structure assumption can be untenable for some applications. Hence, we extend our results to partially recursive structures, which are recursive between different blocks of a partition of the vertices, but without any edges between vertices belonging to the same block. Inferences are developed for recursive and restricted recursive structures based on panel data on random variables considered as panel units. The results also precisely highlight the exact role that permutations and Cholesky decomposition play in identification of macroconomic models. Applied to data on stock returns across several countries, our methods uncover interesting new evidences on the (contemporaneous) causal linkages between the markets.

Keywords: Spatial weights matrix; Directed acyclic graphs; Partial identification of structural VARs; Causal inference; Panel data.
JEL classification: C31, C33, G17.

*Corresponding author: A. Bhattacharjee, Spatial Economics & Econometrics Centre (SEEC), Heriot-Watt University, Room 1.06, Mary Burton Building, Edinburgh EH14 4AS, Scotland, United Kingdom. email: a.bhattacharjee@hw.ac.uk.
Sincere thanks are due to Anindya Banerjee, Atanas Christev, David Hendry, Sean Holly, Toru Kitagawa and Mike Wickens for helpful comments and encouraging suggestions. The usual disclaimer applies.

1 Introduction

The central topic of this paper is the identification of causal networks between units: economic agents or spatial units within a network, macroeconomic variables or cross-section units in panel data. Causal inferences can only be drawn if the units form a DAG (directed acyclic graph), also called a Bayesian network.

Definition 1 *(Harary, 1994). A* ***directed acyclic graph (DAG)*** *is a finite directed graph with no directed cycles, that is, it consists of finitely many vertices (or units) and edges (or causal links), with each edge directed from one vertex to another, such that there is no way to start at any vertex v and follow a consistently-directed sequence of edges that eventually loops back to v again.*

There are no causal loops in a DAG (Harary, 1994; Pearl, 2009) and hence causal inferences can be drawn. However, such inferences require knowledge of the causal links. Inferences on the causal links themselves is a more complex problem (Tian and Pearl, 2002; Evans, 2017) and has been a less active area in the literature. One specific DAG for which such identification results and corresponding algorithms are available is the causal poly-tree; see, for example, Rebane and Pearl (1987) and Dasgupta (1999). This is the main reason why poly-trees have been prominent in Bayesian networks (Cowell et al., 1999) and causal graphs (Pearl, 2009).

Definition 2 *(Rebane and Pearl, 1987). A* ***polytree*** *is a DAG whose underlying undirected graph is a tree. In other words, if we replace its directed edges with undirected edges, we obtain an undirected graph that is both connected and acyclic.*

However, in many application contexts, the definition of a poly-tree is quite narrow. In a tree, any two vertices are connected by exactly one path, and thus a poly-tree has only a limited number of edges. This can be restrictive because the underlying causal patterns may be supported on a larger number of links, and hence causal effects inferred on poly-trees may be exaggerated.

At the other end of the spectrum, we have a recursive DAG, which we define as a graph that represents a recursive ordering. A directed graph is a DAG if and only if it has a topological ordering, an ordering of the vertices such that the starting endpoint of every edge occurs earlier in the ordering than the ending endpoint of the edge (Sedgewick and Wayne, 2011, p.598-9). In general, the topological ordering of a graph is not unique.

Definition 3 *Let G denote a DAG that supports a unique topological ordering of its vertices. Then, the corresponding* ***recursive DAG (R-DAG)*** *is defined as the transitive closure of G. The transitive closure of a directed graph is the*

graph with the most directed edges that represents the same reachability relation. In other words, it has an edge for every related pair $u \leq v$ of distinct elements in the reachability relation of G; that is, it has an edge $u \longrightarrow v$ whenever u can reach v.

Thus, a recursive DAG has the most complete set of directed edges consistent with its (unique) topological ordering. This corresponds exactly to recursive structural equations models and identification schemes in econometrics; we will provide a specific definition in the context of our models later. Recently, Basak et al. (2017) provided an identification result for the recursive ordering of variables or units. However, in practise, recursive networks too appear to be too restrictive. Hence we propose an alternate causal structure based on a partial ordering of the vertices.

Definition 4 *Let G^* denote a DAG that does not have a unique topological ordering of its vertices, but which supports a partition of the vertices (into several blocks) with unique topological ordering between these blocks. In other words, the k vertices of G^* are supported by a partition $y^{[R]} = \left(y_{[1]}, \ldots, y_{[r]}\right)$, with $r \leq k$, such that the following reachability relations hold: (a) $u \leq v$ if and only if $u \in y_{[i]}$ and $v \in y_{[j]}$ such that $i < j$; and (b) u and v are not reachable from each other ($u <> v$) if they belong to the same partition, that is $\{u, v\} \in y_{[i]}$. Then, the corresponding* ***partially recursive DAG (PR-DAG)*** *is defined as the transitive closure of G^*.*

In many applications, data are consistent with there being multiple units or variables that occupy the same place in the recursive order, and it is often very difficult to infer on the relative order of these units. It would then appear that a PR-DAG (partially recursive DAG) may often be appropriate, where the observation units are partitioned into several blocks, and there is inter-block recursive ordering while within each block, there are no directed edges; we will provide a more specific definition in the context of our models later in the paper.

In this paper, we develop identification results for recursive and partially recursive networks and corresponding inference using panel data. We provide several motivating examples in section 2, followed by technical treatment for the general case in section 3. Section 4 develops an application to data on weekly stock returns across 19 different markets worldwide, and section 5 concludes.

2 Context and motivation

In this section, we provide some broader context and intuition to our work. First, we specialise the context to panel data, using an illustrative example, focusing on

asset return spillovers across global equity markets. The application considered later in the paper relate to the same example. We also draw connections with structural macroeconometric and spatial econometrics literatures. Second, we provide intuition for our identification results using a hypothetical example of two urban housing markets, which leads into the following section on identification results.

2.1 Structural ordering in SVAR models

This paper proposes identification of causal recursive and partially recursive DAG structures using panel data and based on a structural vector autoregressive (SVAR) model (Hamilton, 1994; Stock and Watson, 2001). Panel data are not really necessary for our identification and inference methods; here, the temporal dimension only provides replications. There is a long and established tradition in structural macroeconomic modelling where SVARs are used to identify the endogenous relationship between variables. In some contexts, the underlying macroeconomic variables are latent and represented by statistical factors accumulating information from a large number of underlying variables; in this case, we have factor-augmented (structural) vector autoregression models (Bernanke et al., 2005; Mumtaz and Surico, 2009). The resulting SVAR and FAVAR-SVAR models are widely used for studies in financial market contagion, macroeconomics analyses and policy design; see, for example, Christiano et al. (2007), Bekaert et al. (2009), Diebold and Yilmaz (2009) and Mumtaz and Surico (2009).

Different identification structures are used for the contemporaneous structural part of such SVAR models: commonly recursive schemes (Grilli and Roubini, 1995; Eichenbaum and Evans, 1995; Christiano et al., 1999, 2007) or nonrecursive schemes with zero restrictions (Cushman and Zha, 1997; Kim and Roubini, 2000; Uhlig, 2005; Christiano et al., 2007). To be more specific, consider a SVAR(p) model

$$A_0 y_t = a + \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t,\ t = 1, \ldots, T, \tag{1}$$

where y_t is an $k \times 1$ vector, ε_t a $k \times 1$ vector white noise process, normally distributed with mean zero and variance-covariance matrix $\Sigma = diag\left(\sigma_1^2, \ldots, \sigma_k^2\right)$ is a $k \times k$ positive definite diagonal matrix. Note that we assume the idiosyncratic structural shocks to be uncorrelated, as is common in the SVAR literature. The structural parameters $A_0, A_1, \ldots, A_p$ are (at least partially) unknown $k \times k$ matrices, and a is an unknown $k \times 1$ vector. Following convention, we rescale the model to allow for heteroscedastic variances by setting the diagonal elements of A_0 to unity. Then, we write $A_0 = I_k - W$, where I_k is the $k \times k$ identity matrix and W is a $k \times k$ structural matrix with zero diagonal elements.

The reduced form VAR representation of the model (1) is

$$y_t = b + \sum_{j=1}^{p} B_j y_{t-j} + u_t, \tag{2}$$

where $b = A_0^{-1}a$, $B_j = A_0^{-1}A_j$, for $j = 1, \ldots, p$, $u_t = A_0^{-1}\varepsilon_t$, and $E(u_t u_t') = \Omega = A_0^{-1}\Sigma\left(A_0^{-1}\right)'$. Under fairly general conditions, the reduced form parameters $b, B_1, \ldots, B_p$ are usually identified. But identification of the underlying structural parameters $a, A_0, A_1, \ldots, A_p$ require assumptions on the structure of the SVAR, typically either recursive or nonrecursive zero restrictions. Specifically, one needs to impose constraints to identify the underlying contemporaneous structural matrix ($A_0 = I_k - W$) of the SVAR(p) model (1).

In the recursive scheme, the contemporaneous impulse response matrix, or the impact matrix, A_0^{-1}, is lower triangular. As an illustrative example, Mumtaz and Surico (2009) study the impact of external shocks on the UK economy based on FAVAR-SVAR models with the following variables: a "foreign" block $\{\Delta Y_t^*, \Pi_t^*, \Delta M_t^*, R_t^*\}$, where ΔY_t^* represents an international real activity factor, Π_t^* denotes an international inflation factor, ΔM_t^* is an international liquidity factor, and R_t^* denotes comovements in international short-term interest rates; a "domestic" block, where the dynamics of the UK variables are captured by domestic factors F_t^{UK}; and the UK short-term interest, R_t. Then, their recursive identification scheme posits the following structural implication in the reduced form VAR errors:

$$\begin{pmatrix} u_{\Delta Y^*} \\ u_{\Pi^*} \\ u_{\Delta M^*} \\ u_{R^*} \\ u_{F^{UK}} \\ u_R \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \times & 1 & 0 & 0 & 0 & 0 \\ \times & \times & 1 & 0 & 0 & 0 \\ \times & \times & \times & 1 & 0 & 0 \\ \times & \times & \times & \times & 1 & 0 \\ \times & \times & \times & \times & \times & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_{\Delta Y^*} \\ \varepsilon_{\Pi^*} \\ \varepsilon_{\Delta M^*} \\ \varepsilon_{R^*} \\ \varepsilon_{F^{UK}} \\ \varepsilon_R \end{pmatrix}. \tag{3}$$

The marks "$\times$" represent freely estimated parameters. It is clear that the recursive model (3) implies the following topological ordering: $\Delta Y_t^* \to \Pi_t^* \to \Delta M_t^* \to R_t^* \to F_t^{UK} \to R_t$. Recursive identification in small-scale models is typically associated with a number of anomalies such as the price and liquidity puzzles, and the exchange rate and forward discount puzzles. These empirical facts are anomalies because they are inconsistent with the predictions of a number of, though not all, theories. A possible interpretation of the anomalies is that the recursive scheme is unsuited for recovering correctly a policy shock, and in this context our proposed partially recursive ordering can be useful. In the context of our model (1), the above kinds of topological ordering are exactly reflected in the structural matrix W. This connection is made precise in the following definition.

Definition 5 *Consider a structural VAR model (1). The collection of units (or variables) in y is said to have a* ***recursive structure*** *if there exists some permutation of the elements, say* $y^{[P]} = \left(y^{[1]}, \ldots, y^{[k]}\right)$*, for which the corresponding spatial weights matrix* $W^{[P]}$ *is a lower triangular* $k \times k$ *matrix with zero diagonal elements and non-zero principal subdiagonal. That is,*

$$W^{[P]} = \left(\left(w_{ij}^{[P]} : w_{ij}^{[P]} = 0 \quad if \quad j \geq i \quad and \quad w_{ij}^{[P]} \neq 0 \quad if\, j = i-1\right)\right)_{i,j=1,\ldots,k}.$$

Now, consider the ***partially recursive structure****. Here, we have a partition of the k units into r blocks as* $y_{[1]} \rightarrow y_{[2]} \rightarrow \ldots \rightarrow y_{[r]}$*, where recursive ordering holds between the blocks but within each partition, all elements of W are zeroes. In other words, we have:*

$$W^{[P]} = \begin{bmatrix} W_1 & 0 & \ldots & 0 \\ \times & W_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \times & \times & \ldots & W_r \end{bmatrix}, W_i = 0 \quad if\, i = 1, \ldots, r.$$

Note that a block can constitute a single unit as well, and recursive ordering simply refers to the case where all blocks have unit cardinality.

Clearly, the restrictions underlying the partially recursive structure are less stringent that those for the recursive structure. As discussed in section 1, the above R-DAG assumption may be too strong in many applications. Specifically, there may be situations where some of the units may occupy similar positions in the causal order, but it may be difficult to order the variables unambiguously based on finite sample data. Thus, in order to improve the identification of the monetary shock, several authors have proposed alternative schemes ranging from nonrecursive to sign restrictions. For example, Mumtaz and Surico (2009) consider the following nonrecursive scheme:

$$\begin{pmatrix} u_{\Delta Y^*} \\ u_{\Pi^*} \\ u_{\Delta M^*} \\ u_{R^*} \\ u_{F^{UK}} \\ u_R \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \times & 1 & 0 & 0 & 0 & 0 \\ \times & \times & 1 & \times & 0 & 0 \\ 0 & 0 & \times & 1 & 0 & 0 \\ \times & \times & \times & \times & 1 & 0 \\ \times & \times & \times & \times & \times & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_{\Delta Y^*} \\ \varepsilon_{\Pi^*} \\ \varepsilon_{MD^*} \\ \varepsilon_{MS^*} \\ \varepsilon_{F^{UK}} \\ \varepsilon_R \end{pmatrix}. \tag{4}$$

Following Sims and Zha (2006), the third and fourth rows identify money demand and money supply shocks in the rest of the world, respectively. There are considerable contemporaneous causal implications of the nonrecursive scheme (4). The transmission of shocks originate from world activity and then pass on to world

inflation. World money demand and money supply occupy middle positions in the causal order, but their relative ordering is ambiguous. The shocks then pass on to the domestic economy, followed finally by the policy rate. Then, this suggests a weaker notion of partially recursive ordering.

Specifically, in section 3 we show that, for a SVAR(0) model with no lags, but with recursive contemporaneous causation, and with idiosyncratic shocks that are homoscedastic across the variables, the position at the top of the causal order is taken by the variable with the smallest variance. Once this variable is partialled out, the second position is occupied by the variable with the smallest partial variance; and so on. Further, ties in the variances at the top of the causal order at any stage of the iterative procedure correspond to variables (or units) that lie within the same partition, that is $\{u, \upsilon\} \in y_{[i]}$.

However, the homoscedasticity assumption may be strong in many applications. Then, we also show that, for an SVAR(p) model with PR-DAG contemporaneous dependence, the standard deviation of the idiosyncratic shocks can be inferred from Cholesky decompositions of the error covariance matrix of a reduced form VAR computed over all permutations of the units. Clearly, this corresponds closely to the use of permutations and Cholesky decompositions in the modeling of structural orders in the current literature, towards which we now turn.

2.1.1 Contagion across financial markets

There is substantial recent research interest in using panel data to estimate spillovers and contagion effects in stock returns and volatilities across different countries and markets. Bekaert et al. (2009) documented return spillovers across 23 countries, while Diebold and Yilmaz (2009) developed a spillover index based on VAR models, and used this to study the evolution of return and volatility spillovers across 19 stock markets. They argue that the variance decompositions define weighted and directed networks, and hence this is closely related to our work. Our empirical analysis is based on the context and data from Diebold and Yilmaz (2009), which we briefly discuss to highlight the approach and context of using Cholesky decompositions and permutations across different orderings of endogenous units.

Consider a reduced form VAR representation of the data as above (2). By covariance stationarity, the moving average representation of the VAR exists and is given by

$$\begin{aligned} y_t &= \Theta(L)u_t = A(L)\varepsilon_t \\ \Theta(L) &= (I - \Phi L)^{-1}; A(L) = \Theta(L)Q_t^{-1}, \end{aligned}$$

where $E(\varepsilon_t\varepsilon_t') = I$ and Q_t^{-1} is the "unique" lower-triangular Cholesky factor of the covariance matrix of u_t. In the literature, this relation is then often used to justify interpreting ε_t as the underlying structural shocks. If this interpretation

were correct, one could then potentially go ahead with constructing an index of spillovers or, for that matter, structural interpretation of the models. However, the uniqueness of Q_t^{-1} depends on two additional conditions. First, it requires the assumption that there is an underlying recursive ordering of variables in the nature of a R-DAG. Second, and even more precisely, one would need to assume that the order of units in y_t is indeed the correct causal ordering.

Of course, in practise, one cannot ensure a correct ordering, except through theory or, as in this paper, through appropriate inference on the recursive order. Hence, in their effort to construct a spillover index, Diebold and Yilmaz (2009) consider averaging over all the possible permutations. This is computationally very intensive, since the number of possible permutations, 19!, is very large. Hence, they consider a small number of (randomly chosen) permutations, and verify that their empirical findings are robust. In later work, Klößner and Wagner (2013) argue that handling a large number of permutations is not impossible, and they provide an algorithm to explore all VAR orderings.

The above application is only illustrative at this stage, but it is very useful in introducing our data and context, as well as in emphasizing current best practise in the area. Indeed, the above idea of Cholesky factorisation over "all" permutations is standard in the literature. However, it misses the central point that the methodology relies crucially on an underlying SVAR model with recursive structure. Further, it does not emphasize enough, in our view, why the Cholesky decomposition is useful. We make these issues more clear through our identification results. Specifically, we pose the question: is the recursive ordering, and the weaker partially recursive ordering, identified from the data?

2.1.2 Structural macroeconomics and spatial econometrics

The close connections of our work with current structural macroeconometric modelling is clear from the above discussions. We note two further aspects of this connection. First, structural assumptions are necessary for identification. However, such assumptions are not easily verifiable. Mumtaz and Surico (2009) use identification based on recursive, nonrecursive and sign restrictions, while being somewhat agnostic about which of these is most appropriate in their context.

Then, this paper contributes to the current debate within the literature as to empirical validation of structural assumptions underlying panel data and macroeconomic models. There is some resurgence of research in this area; see, for example, Giacomini and Kitagawa (2015) and Stock and Watson (2015). Our identification results contribute to this literature by providing inferences on the specific recursive or partially recursive ordering that is supported by the data.

Second, our identification result for R-DAG and PR-DAG are based on relative variability of the variables in the SVAR. In this sense, our work is related to

previous literature on identification by relative variances and also to conventional wisdom in SVAR modelling that variables with smaller variances should appear towards the top of the causal order (Rubio-Ramírez et al., 2010; Sims, 2012); see also Lippi and Reichlin (1994).

Finally, identification of causal order relates to the structure of W, and this also has connections with inferences on spatial weights in the recent spatial econometrics literature. This literature has highlighted that the spatial weights matrix W is in general not identified except under structural assumptions, such as symmetry (Bhattacharjee and Jensen-Butler, 2013), sparsity (Bailey et al., 2016) or moment conditions (Bhattacharjee and Holly, 2013). Our work shows that W is also identified under recursive and partially recursive structures and further, that the ordering itself is identified from the data; see also Basak et al. (2017).

2.2 A tale of two cities

To provide intuition for our identification results, we consider first a hypothetical question in the simple case with $k = 2$. Consider house prices in two neighbouring cities, say Edinburgh and Glasgow, denoted y_E and y_G respectively. There is potential for price spillover between the two cities, and such spillovers are captured by an index, similar to Diebold and Yilmaz (2009), and denoted w. Initially, we assume zero autocorrelation ($p = 0$) and homoscedasticity of the innovations: $\sigma_E^2 = \sigma_G^2 = \sigma^2$. Under the assumption of recursive structure, and ignoring the time subscript, the true model is either

$$\begin{pmatrix} y_E \\ y_G \end{pmatrix} = \begin{pmatrix} a_E \\ a_G \end{pmatrix} + \begin{bmatrix} 0 & 0 \\ w & 0 \end{bmatrix} \begin{pmatrix} y_E \\ y_G \end{pmatrix} + \begin{pmatrix} \varepsilon_E \\ \varepsilon_G \end{pmatrix}, \; E\left(\varepsilon\varepsilon'\right) = \sigma^2 I,$$

or,

$$\begin{pmatrix} y_E \\ y_G \end{pmatrix} = \begin{pmatrix} a_E \\ a_G \end{pmatrix} + \begin{bmatrix} 0 & w \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y_E \\ y_G \end{pmatrix} + \begin{pmatrix} \varepsilon_E \\ \varepsilon_G \end{pmatrix}, \; E\left(\varepsilon\varepsilon'\right) = \sigma^2 I.$$

But, which one is it? It is clear that, if Edinburgh precedes Glasgow in causal ordering (the former model), then $V\left(y_E\right) = \sigma^2 \leq \sigma^2\left(1 + w^2\right) = V\left(y_G\right)$, with the equality holding if and only if $w = 0$. On the other hand, if y_E follows y_G in causal ordering then, $V\left(y_E\right) > V\left(y_G\right)$ if $w \neq 0$. The case $w = 0$ is also interesting because this corresponds to partially recursive ordering. Here, the variances would be exactly equal, that is, $V\left(y_E\right) = V\left(y_G\right)$.

Admittedly, the above analysis ignores potential differences in the variance of idiosyncratic errors (that is, $\sigma_E^2 \neq \sigma_G^2$), as well as temporal dynamics ($p > 0$). Using methods described below, we can obtain estimates σ_E and σ_G using a reduced form VAR(p) model for y_C and y_E, subject to starting with the correct ordering in the first place. Then, iterating over all possible permutations, as in

Klößner and Wagner (2013), we can estimate the standard deviations. Assuming $p = 1$, we have

$$\begin{aligned}
\begin{pmatrix} y_{Et}/\widehat{\sigma}_E \\ y_{Gt}/\widehat{\sigma}_G \end{pmatrix} &= \begin{pmatrix} a_E/\widehat{\sigma}_E \\ a_G/\widehat{\sigma}_G \end{pmatrix} + W \begin{pmatrix} y_{Et}/\widehat{\sigma}_E \\ y_{Gt}/\widehat{\sigma}_G \end{pmatrix} + B \begin{pmatrix} y_{E,t-1}/\widehat{\sigma}_E \\ y_{G,t-1}/\widehat{\sigma}_G \end{pmatrix} + \begin{pmatrix} \varepsilon_E/\widehat{\sigma}_E \\ \varepsilon_G/\widehat{\sigma}_G \end{pmatrix}, \\
W &= \begin{bmatrix} 0 & w\frac{\widehat{\sigma}_E}{\widehat{\sigma}_C} \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 0 \\ w\frac{\widehat{\sigma}_C}{\widehat{\sigma}_E} & 0 \end{bmatrix}.
\end{aligned}$$

We are now back to the homoscedasticity case, and now contemporaneous causation can be inferred from the estimated covariance matrix of the scaled vector $\begin{pmatrix} y_{Et}/\widehat{\sigma}_E \\ y_{Gt}/\widehat{\sigma}_G \end{pmatrix}$. So, we can conclude whether the appropriate R-DAG is $y_E \rightarrow y_G$ or $y_G \rightarrow y_E$. Further, we can also examine whether the true model is indeed the PR-DAG where neither y_E nor y_G contemporaneously cause the other. Also, estimates of the above VAR(1) model indicate whether either of the two Granger cause the other, or the other way round, or indeed whether Granger causation runs both ways. With this backdrop, we can now proceed to our identification results.

3 Models and methodology

Consider again the SVAR(p) model (1) but now expressed in terms of the structural matrix W:

$$y_t = a + Wy_t + \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t,\ E\left(\varepsilon_t \varepsilon_t'\right) = \Sigma = diag\left(\sigma_1^2, \ldots, \sigma_k^2\right),$$

where W is a $k \times k$ matrix with zero diagonal elements. Then the reduced form is the following:

$$\begin{aligned}
y_t &= \left(I_k - W\right)^{-1} a + \sum_{j=1}^{p} \left(I_k - W\right)^{-1} A_j y_{t-j} + u_t, \\
E\left(u_t u_t'\right) &= \left(I_k - W\right)^{-1} \Sigma \left(I_k - W\right)^{-1\prime}.
\end{aligned}$$

Now, we make one of the following two structural assumptions required for our identification result.

Assumption 1. *(Recursive Structure) There exists some permutation of the variables in y_t, say $y_t^{[P]}$, for which the corresponding spatial weights matrix $W^{[P]}$ is a lower triangular $k \times k$ matrix with zero diagonal elements and non-zero principal subdiagonal, which implies a* ***recursive DAG (R-DAG)****.*

Alternatively, we may consider the following weaker assumption.
Assumption 1a. *(Partially Recursive Structure) There exists some partition of the variables in y_t, say $y_t^{[R]}$, for which the transitive closure network graph is a* ***partially recursive DAG (PR-DAG)****.*

Under either **Assumption 1** or **Assumption 1a**, $W^{[P]}$ is a lower triangular $k\times k$ matrix with zero diagonal elements. That is, $W^{[P]} = \left(\left(w_{ij}^{[P]} : w_{ij}^{[P]} = 0 \text{ if } j \geq i\right)\right)_{i,j=1,\ldots,k}$. Then, we have an important result.
Lemma 1. $(I_k - W)^{-1} = I_k + \sum_{i=1}^{k-1} W^i$.

The lemma is useful for later results. Further, it is also very useful for calculating "direct and indirect effects". Note further that, under **Assumption 1** (or **Assumption 1a**), the reduced form is always identified. This is because $W^k = 0$, and hence $(I-W)^{-1} = I + W + W^2 + \ldots + W^{k-1}$ always exists. Hence, for identification, we do not need the spatial granularity condition (Pesaran, 2006), which is standard in the literature and closely related to conditions for spatial stationarity in Kelejian and Prucha (1998) and Lee (2004). However, this condition is similar to ergodicity and is useful for obtaining asymptotic normality results. In practise, this requires that strong dependence in the data are modelled *a priori*, using for example, a factor structure (Bai, 2009; Pesaran and Tosetti, 2011). Then, under the pure-SAR model (**??**), denoting covariance matrix by $Cov(.)$, we have:

$$\begin{aligned} Cov(Y_t) &= Cov[(I-W)^{-1}\epsilon_t] \\ &= (I-W)^{-1}[Cov(\epsilon_t)]((I-W)')^{-1} \\ &= (I-W)^{-1}\Sigma_\epsilon((I-W)')^{-1} \end{aligned} \tag{5}$$

where Σ_ϵ is the covariance matrix of ϵ_t. Since the components of ϵ_t are independent by assumption, $\Sigma_\epsilon = diag\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right)$.
Assumption 2. *(Spatial granularity condition): The row and column norms of W are bounded (in absolute value) by 1.*

Now, following the relative variances intuition in the $k = 2$ case from the previous section, we can present our first main result.

Proposition 1. *Consider the SVAR(p) model (1) with $k > 2$, with no lag structure ($p = 0$), homoscedasticity of the shocks ($\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$), and where* ***Assumption 1*** *holds. Then the variable with the smallest variance ($y^{[1]}$) comes at the top of the causal order. Construct the partial covariance matrix of the other variables, after partialling out $y^{[1]}$. The variable with the smallest partial variance ($y^{[2]}$) occupies the second position in the causal order. This iterative procedure recovers the causal order $y_t^{[P]} = \left(y_t^{[1]}, \ldots, y_t^{[k]}\right)$ for the entire vector y_t* in the case

that the underlying causal structure is R-DAG. Further, if the causal structure were PR-DAG, then the variances at the top of the order will be equal at the corresponding steps of the iterative procedure. Then, this procedure identifies the correct PR-DAG partition as well.

Proof. *First consider the case $k = 3$. Consider the recursive structure*

$$W = \begin{bmatrix} 0 & 0 & 0 \\ \theta & 0 & 0 \\ v & w & 0 \end{bmatrix}.$$

Then, by Lemma 1

$$\begin{aligned} E(yy') &= \sigma^2 \left(I + W + W^2\right)\left(I + W' + \left(W^2\right)'\right) \\ &= \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ \theta & 1 & 0 \\ v + \theta w & w & 1 \end{bmatrix} \begin{bmatrix} 1 & \theta & v + \theta w \\ 0 & 1 & w \\ 0 & 0 & 1 \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} 1 & \theta & v + \theta w \\ \theta & 1 + \theta^2 & \theta(v + \theta w) + w \\ v + \theta w & \theta(v + \theta w) + w & 1 + (v + \theta w)^2 + w^2 \end{bmatrix}. \end{aligned}$$

It is clear that the variable with the smallest variance comes at the top of the causal order. However, the relative order of the other two variables is not clear. Partialling out the first first variable, we have

$$\begin{aligned} E\left(y^{[-1]}\left(y^{[-1]}\right)'\right) &= \sigma^2 \left\{ \begin{bmatrix} 1 + \theta^2 & \theta(v + \theta w) + w \\ \theta(v + \theta w) + w & 1 + (v + \theta w)^2 + w^2 \end{bmatrix} - \begin{pmatrix} \theta \\ v + \theta w \end{pmatrix} \begin{pmatrix} \theta & v + \theta w \end{pmatrix} \right\} \\ &= \sigma^2 \begin{bmatrix} 1 & w \\ w & 1 + (v + \theta w)^2 \end{bmatrix}. \end{aligned}$$

It is now clear that the second position in the causal order is taken by the second variable. Hence, the procedure recovers the correct causal order. In the general case, we run a proof by induction. This is done in two parts. First, we show that, at any step of the iteration, once the correct element at the top of the causal order has been identified, partialling this element from the covariance matrix provides the correct covariance matrix for the remaining elements, Second, we show that the smallest diagonal element of this partial covariance matrix corresponds to the first element in the true causal (recursive) order for the remaining elements.

For the first part, the proof is not specific to which step the iterative process of inferring the order the procedure is currently in. Hence without loss of generality,

we consider the first step. We have identified the first element in the true recursive order, that is $y^{[1]}$. *We partition the vector* $y^{[P]}$ *as* $y^{[P]} = \begin{pmatrix} y^{[1]} & y^{[-1]\prime} \end{pmatrix}'$ *and correspondingly partition* W *as*

$$W = \begin{bmatrix} 0 & 0' \\ w_1 & W_2 \end{bmatrix},$$

where $0'$ *is a* $(1 \times \underline{k-1})$ *row vector of zeroes,* w_1 *is a* $(\underline{k-1} \times 1)$ *column vector with arbitrary elements, and* W_2 *is a* $(\underline{k-1} \times \underline{k-1})$ *lower triangualer matrix with zero diagonal elements. Denote* $W^* = \sum_{j=0}^{k-2} {W_2}^j$, *and also note that* $(I-W)^{-1} = \left(I + W + W^2 + \ldots + W^{k-1}\right)$ *since* $W^j = 0$ *if* $j \geq k$. *Correspondingly, the covariance matrix of the reduced form errors is given by*

$$\begin{aligned}
\Sigma &= \sigma^2 (I-W)^{-1} (I-W)^{-1\prime} \\
&= \sigma^2 \left(I + W + W^2 + \ldots + W^{k-1}\right) \left(I + W' + W^{2\prime} + \ldots + W^{k-1\prime}\right) \\
&= \sigma^2 \begin{bmatrix} 1 & 0' \\ W^* w_1 & I + W^* W_2 \end{bmatrix} \begin{bmatrix} 1 & w_1' W^{*\prime} \\ 0 & I + W_2' W^{*\prime} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & w_1' W^{*\prime} \\ W^* w_1 & I + W^* W_2 + W_2' W^{*\prime} \\ & + W^* \left[w_1 w_1' + W_2 W_2'\right] W^{*\prime} \end{bmatrix}.
\end{aligned}$$

Then, partialling out the first element, we have

$$\begin{aligned}
\Sigma_{2.1} &= \sigma^2 \left[I + W^* W_2 + W_2' W^{*\prime} + W^* W_2 W_2' W^{*\prime}\right] \\
&= \sigma^2 \left(I + W^* W_2\right) \left(I + W^* W_2\right)' \\
&= \sigma^2 \left(I + W_2 + W_2^2 + \ldots + W_2^{k-1}\right) \left(I + W_2 + W_2^2 + \ldots + W_2^{k-1}\right)' \\
&= \sigma^2 \left(I + W_2 + W_2^2 + \ldots + W_2^{k-2}\right) \left(I + W_2 + W_2^2 + \ldots + W_2^{k-2}\right)' \\
&= \sigma^2 (I - W_2)^{-1} (I - W_2)^{-1\prime},
\end{aligned}$$

where the penultimate step follows because $W_2^{k-1} = 0$. *The proof of the first part follows by noting that* $\Sigma^{[-1]}$ *is the covariance matrix of* $y^{[-1]} - w_1 y^{[1]} = W_2 y^{[-1]} + \varepsilon^{[-1]}$.

For the second part, consider the $(r+1)$*th step of the iteration. The first* r *positions of* $y^{[P]}$ *have previously been inferred in their correct recursive order; denote this order by* $y^{(1)}$. *The object is to show that the* $(r+1)$*th step correctly identifies* $y^{[r+1]}$ *as being in the* $(r+1)$*th position. Partition* $y^{[P]} = \begin{pmatrix} y^{(1)\prime} & y^{(2)\prime} \end{pmatrix}'$ *and* $\varepsilon = \begin{pmatrix} \varepsilon^{(1)\prime} & \varepsilon^{(2)\prime} \end{pmatrix}'$; *correspondingly partition* W *as*

$$W = \begin{bmatrix} {W^{(1)}}_{r \times r} & 0 \\ {W^{(21)}}_{(k-r) \times r} & {W^{(2)}}_{(k-r) \times (k-r)} \end{bmatrix}.$$

Partialling $y^{(1)}$ from $y^{(2)}$, and using the first part, we obtain:

$$\begin{aligned} y^{[2.1]} &= y^{(2)} - w_1 y^{(1)} \\ &= W_2 y^{(2)} + \varepsilon^{(2)}. \end{aligned}$$

We denote elements of $y^{[2.1]}$ as $y^{[2.1]} = \begin{pmatrix} y_1^{[2.1]} & y_2^{[2.1]} & \dots & y_{k-r}^{[2.1]} \end{pmatrix}$ and corresponding variances as $V_1^{[2.1]}, V_2^{[2.1]}, \dots, V_{k-r}^{[2.1]}$. Then, since W has a triangular (recursive) structure, we have:

$$y_1^{[2.1]} = \varepsilon_1^{(2)}; y_2^{[2.1]} = W_{21}^{(2)} y_1^{[2.1]} + \varepsilon_2^{(2)}; \dots$$

and recursively

$$y_j^{[2.1]} = \sum_{i=1}^{j-1} W_{ji}^{(2)} y_i^{[2.1]} + \varepsilon_j^{(2)}.$$

Correspondingly,

$$\begin{aligned} V_1^{[2.1]} &= \sigma^2; V_2^{[2.1]} = \sigma^2 \left(\left[W_{21}^{(2)} \right]^2 + 1 \right); \\ V_3^{[2.1]} &= \sigma^2 \left(\left[W_{31}^{(2)} \right]^2 + \left[W_{32}^{(2)} \right]^2 \left(\left[W_{21}^{(2)} \right]^2 + 1 \right) + 1 \right); \dots \end{aligned}$$

and recursively

$$V_j^{[2.1]} = \sigma^2 \left(1 + \sum_{i=1}^{j-1} \left[W_{ji}^{(2)} \right]^2 V_i^{[2.1]} \right).$$

This implies that $y_1^{[2.1]} = y^{[r+1]}$ has the smallest partial variance, and the result is proved.
The second part shows that, at any step of the iterative procedure, the element that is next in the true causal order has the smallest partial variance, and the first part showed that the partial covariance matrix retains the required recursive causal order. This recursion runs until the order reduces to 2, and the true recursive order is recovered. The identification for PR-DAG follws simply by noting that within each block of the partition, the partial variances of each unit in the block is exactly equal. ■

There are two important implications to note. First, the partial covariances matrices can be easily estimated by OLS. Hence, in practise, at every step we regress the remaining units on the units above it in the causal order and simply compute the residual covariance matrix. Second, here identification is through relative variances, which is reminiscent of the macroeconomics literature; see, for example, Rubio-Ramírez et al. (2010) and Sims (2012).

The following proposition extends the context to a lag structure and heteroscedastic errors. It provides methodology to estimate the idiosyncratic error variances.

Proposition 2. *Consider the SVAR(p) model (1) with any number of variables and any lag structure. The innovations are potentially heteroscedastic. We make either* ***Assumption 1*** *(recursive structure) or* ***Assumption 1a*** *(partially recursive structure), and further that the variables are in their correct recursive order. Denote the Cholesky decomposition of the reduced form error covariance matrix* $E\left(u_t^{[R]}\, u^{[R]}{}_t'\right) = \Omega$ *in (2) as* $\Omega = LL'$. *Then, the standard deviations of the idiosyncratic shocks constitute the diagonal elements of* L.

Proof. *Consider for simplicity* $p = 1$. *Then model (1) can be written in vector notation as*

$$Y_t = WY_t + \Lambda Y_{t-1} + \epsilon_t, \tag{6}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_k)$ *is a diagonal matrix, and* $\{\epsilon_t\}$ *is potentially heteroscedastic, that is,* $\Sigma_\epsilon = diag\,(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$, . *Then, the reduced form of (6) is given by*

$$\begin{aligned} Y_t &= (I - W)^{-1} \Lambda Y_{t-1} + (I - W)^{-1} \epsilon_t = AY_{t-1} + u_t; \\ A &= (I - W)^{-1} \Lambda, Cov(u_t) = (I - W)^{-1} \Sigma_\epsilon \left(I - W\right)^{-1\prime}. \end{aligned} \tag{7}$$

The reduced form (7) can be estimated as a vector autoregressive (VAR) model (Hamilton, 1994; Stock and Watson, 2001). The reduced form (VAR) estimates provide $\widehat{A}$ *and* $\widehat{Cov}(u_t)$. *Note that the diagonal elements of* $(I - W)^{-1}$ *are ones, and* Λ *is a diagonal matrix. Then, the* i*th diagonal element of* Λ *must be the corresponding diagonal element of* $A \equiv ((a_{ij}))_{i,j=1,\ldots,k}$; *that is,* $\lambda_i = a_{ii}, i = 1, \ldots, k$. *This implies that a consistent estimator for* $(I - W)^{-1}$ *can be obtained by dividing each column of* $\widehat{A}$ *by the corresponding diagonal element. In other words, construct a matrix* $A^* \equiv \left(\left(a_{ij}^*\right)\right)_{i,j=1,\ldots,k}$ *as follows*

$$a_{ij}^* = \widehat{a}_{ij}/\widehat{a}_{jj}, i, j = 1, \ldots, k.$$

Then, $A^* = (I - \widehat{W})^{-1}$ *has unit diagonal elements and constitutes a consistent estimator of* $(I - W)^{-1}$, *and therefore,* $A^{*\,-1} = (I - \widehat{W})$ *is a consistent estimator of* $(I - W)$.

Now, consider $\widehat{Cov}(u_t)$. *Since* $Cov(u_t) = (I - W)^{-1} \Sigma_\epsilon \left(I - W\right)^{-1\prime}$, *the diagonal elements of* $(I - \widehat{W})\widehat{Cov}(u_t)(I - \widehat{W})'$ *provide consistent estimators for the elements of* Σ_ϵ, *that is,* $\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_k^2$. ■

Corollary 1. *Suppose we obtain a consistent estimator* $\widehat{\Omega}$. *Then, the idiosyncratic error standard deviations are consistently estimated by the corresponding diagonal elements of* $\widehat{L}$.

Obtaining the consistent estimator $\widehat{\Omega}$ is simple, needing estimation of the reduced form VAR in standard ways, typically by seemingly unrelated regressions, collecting residuals and compute the residual covariance matrix. Then, **Proposition 2** clearly emphasizes the precise role of the Cholesky decomposition and permutations. Specifically, with the correct ordering, the Cholesky factorisation correctly identifies the standard deviations of the idiosyncratic shocks. However, the correct ordering is likely unknown *a priori*. Hence there is the (potential) need for permutations, or otherwise, consideration of a probability distribution on the space of all permutations.

Finally, the next result provides identification of causal order based on scaled variables.

Proposition 3. *Consider the SVAR(p) model (1) with $k > 2$, with arbitrary lag structure ($p = 0$), arbitrary heteroscedasticity of the innovations, and where **Assumption 1** holds. Scale each variable by its standard deviation estimated using **Proposition 2**. That is: $y_{[S]1t} = y_{1t}/\widehat{\sigma}_1, \ldots, y_{[S]kt} = y_{kt}/\widehat{\sigma}_k$. Estimate the error covariance matrix from the reduced form VAR(p) model based on the standardised variables. Then the variable with the smallest variance ($y_{[S]}^{[1]}$) comes at the top of the causal order. Construct the partial covariance matrix of the other variables, after partialling out $y_{[S]}^{[1]}$. The variable with the smallest partial variance ($y_{[S]}^{[2]}$) occupies the second position in the causal order. This iterative procedure recovers the causal partition $y_{[S]t}^{[R]}$ for the entire vector y_t.*

Proposition 6 ***Proof.*** *Note that, by **Assumption 1** (or **Assumption 1a**), and the fact that the variables are in their true order, W is lower triangular. Now, the reduced form error covariance matrix is $E(uu') = (I - W)^{-1} \Sigma (I - W)^{-1\prime}$. Consider the Cholesky decomposition of the covariance matrix as: $E(uu') = LL'$, where L is a unique lower triangular matrix. Since $(I - W)^{-1}$ is lower triangular and with 1's on the diagonal, and since Σ is a diagonal positive definite matrix, it must be the case that $L = (I - W)^{-1} \Sigma^{1/2}$. Since the diagonal elements of $(I - W)^{-1} \Sigma^{1/2}$ are the idiosyncratic error standard deviations, the result follows.* ■

Note that, **Proposition 3** only holds when the units are in their correct recursive order. Hence, consider the following definition.

Definition 7 *In the context of **Propositions 2** and **3** and an underlying ordering assumption (R-DAG or PR-DAG, **Assumption 1** or **Assumption 1a**), a specific permutation is **admissible** if we start with this permutation, apply **Proposition 2**, and it is then found to consistent with the partition (and corresponding partial order) estimated using **Proposition 3**.*

One crucial implication of **Propositions 2** and **3** is that it allows us to restrict attention to a small set of admissible permutations. We start with a candidate permutation in **Proposition 2** and then this permutation is admissible if, and only if, it matches with the ordering recovered by **Proposition 3**. Then, one can check consistency of structural implications under all such admissible orderings, and if there is only one, this ordering is unique; in this case the corresponding causal model is a R-DAG. In the PR-DAG case, there will be multiple admissible orderings, but they should all be consistent with the underlying correct PR-DAG partition. One can also average over all such admissible orderings, or place a (Bayesian) prior over these depending, for example, on how closely they line up with underlying theory. This is clearly in line with DSGE-VAR (Del Negro and Schorfheide, 2004, 2009). As discussed earlier, validation of structural assumptions underlying panel data and macroeconomic models is of considerable importance; for recent discussions, see Diebold and Yilmaz (2014), Giacomini and Kitagawa (2015) and Stock and Watson (2015). Our work contributes towards this literature.

The above results provide identification of causal ordering not only under recursive ordering, but also under the assumption of partially recursive structure, which is less restrictive than recursive ordering but richer than poly-trees. Hence, this structure may be quite useful in applications like the one we consider in the following section. Note also that the partial order recovered here is scale invariant, and the standardization in **Proposition 3** is only to ensure that we are comparing "like-for-like".

The identification results here are based on second moments and hence inference on ordering achieves $\sqrt{n}$-consistency under reasonable conditions. Inferences under the selected models is straightforward. In fact, under both R-DAG and PR-DAG assumptions, simple OLS delivers consistent and asymptotically Gaussian estimates. The technical results are a bit complex and rely on identifiable uniqueness of the models selected by the above identification results. Then, Pötscher and Prucha (1997) provide an excellent collection of results for M-estimation that can be applied in standard ways to deliver teh above inference results; see also Basak et al. (2017).

4 Data and results

We use the same data as Diebold and Yilmaz (2009) and Klößner and Wagner (2013). The weekly data on returns of 19 stock indices are reported in the data aarchive corresponding to Klößner and Wagner (2013), and we also use their algorithm for exploring all 19! permutations. Since the data are large; $T = 829$ weeks, running consecutively from the week ending Friday, 10 January 1992 to 23 November 2007. Given the large number of observations, we use 1 percent signi-

ficance level for all statistical tests. First, we estimate reduced form VAR models using lag selection to choose a VAR(4) model. Next we iterate over all permutations, applying Propositions 2 and 3 in turn, to reduce attention to a selection of admissible partial orders. For the application of Proposition 3, we need tests for equality of variances. Since we assume Gaussian errors, the standard F-test is optimal in our context (Faust, 1992). However, the literature also points out that the test due to Levene (1960) can perform better in many circumstances and we also use this test for robustness check; see, for example, Lim and Loh (1996). This procedure turns out to be very effective in this application. The unique PR-DAG partition consistent with all admissible permutations is the following (country codes correspond to Table 1):

$$\begin{aligned} CHI \;&\rightarrow\; \{US, TAI, JPN\} \rightarrow PHL \rightarrow \{FRA, ARG\} \rightarrow \{SGP, TUR, AUS\} \\ &\rightarrow\; MYS \rightarrow THA \rightarrow \{UK, GER, MEX, HKG\} \rightarrow IDN \rightarrow BRA \rightarrow KOR. \end{aligned}$$

Table 1: Model estimates

SVAR: W '	CHI	US	TAI	JPN	PHL	FRA	ARG	SGP	TUR	AUS	MYS	THA	UK	GER	MEX	HKG
Chile	**0**	0.133	–	–	–	–	1.051	0.171	–	0.109	–	–	–	–	–	–
US-DJIA		**0**			–	0.926	0.992	–	–	0.209	–	–	0.254	0.603	–	–
Taiwan			**0**		–	–	–	0.168	–	–	–	–	–	–	–	–
Japan				**0**	–	–	–	0.089	–	0.039	–	–	–	0.075	–	–
Philippines					**0**	–	–	0.117	–	–	–	–	–	–	–	0.291
France						**0**	–	0.120	1.055	–	-0.379	–	0.483	0.913	–	–
Argentina							**0**	0.024	–	0.011	–	–	–	–	0.135	–
Singapore								**0**			0.635	0.822	–	–	–	0.393
Turkey									**0**		–	–	–	–	–	–
Australia										**0**	–	–	0.192	–	1.329	1.460
Malaysia											**0**	–	–	–	–	0.122
Thailand												**0**	–	–	–	–
UK-FTSE													**0**			
Germany														**0**		
Mexico															**0**	
Hong Kong																**0**
Indonesia																
Brazil																
Korea																

SVAR: Lags	CHI	US	TAI	JPN	PHL	FRA	ARG	SGP	TUR	AUS	MYS	THA	UK	GER	MEX	HKG
Chile							[1]									
US-DJIA		[2,3,4]			[3]	[1,2]		[1]		[1,4]	[3]		[4]	[2,3,4]	[1]	
Taiwan			[1]						[3]					[1,2]		
Japan				[1]									[4]			
Philippines			[1]		[1]			[2]	[3,4]		[1]					
France		[1,3]				[2]				[4]			[1]	[2]		[2]
Argentina	[1]						[1,2]	[3]							[1]	
Singapore		[1]						[2]		[1]	[1]					[2]
Turkey		[4]							[1]		[1]					
Australia		[2]							[3]		[1,3]			[1]	[3,4]	
Malaysia			[1]		[1,3,4]	[3]	[1]	[3]	[2]	[1,2,4]	[1,3,4]		[4]	[3]		[1]
Thailand										[1]		[1]				
UK-FTSE		[1,4]											[1,3]			
Germany		[1]								[4]			[1]	[1,2,3,4]		
Mexico	[3]						[3]				[2,4]				[1,3]	
Hong Kong	[1]	[1,2]				[1]		[1,2]	[1]	[1,2]	[2,4]	[1]			[1,2]	[1.2]
Indonesia		[3]	[1]		[3]			[4]		[1,3]						
Brazil							[2]									
Korea		[2]						[2,3]								

Model Fit: R^2	CHI	US	TAI	JPN	PHL	FRA	ARG	SGP	TUR	AUS	MYS	THA	UK	GER	MEX	HKG
- SVAR	0.144	0.560	0.434	0.323	0.313	0.696	0.421	0.545	0.434	0.667	0.557	0.325	0.809	0.889	0.668	0.541
- VAR	0.144	0.548	0.433	0.321	0.300	0.570	0.374	0.417	0.407	0.579	0.516	0.226	0.530	0.619	0.512	0.370

In this specific geographical finance context, this partial order makes very good sense. First, the recursive ordering assumption is not supported by the data, and relaxing this to a partially recursive assumption clearly makes sense. Second, the selected partial order or PR-DAG partition also makes good sense, identifying oder of influence in the network. Very importantly, the estimated partial order relates well both to geographical proximity and size of markets, but equally to the diurnal

order in which different markets open and close. In fact, our results capture in a nice way the so-called 'meteor shower' phenomenon documented first in volatility by Engle et al. (1990) and in returns by Hamao et al. (1990); see also Ibrahim and Brzeszczyński (2009). Next, we estimate a SVAR model under the structural (zero) assumptions implied by the above PR-DAG partition. The estimates are reported in Table 1.

In the estimates of the structural network weights matrix W (reported as W') in Table 1, the yellow shaded lower triangle represents the partial order that would have been estimated, using the identification results in Banerjee et al. (2017), if the assumption of R-DAG recursive order were made. However, there are substantial differences of this model from the PR-DAG model above, and these differences are highlighted in the green shaded cells above the diagonal. Further, the statistical significance of estimates of the instantaneous impulse response or contemporaneous network interaction matrix W also indicate some support for yet further rationalisation of the partition. This provides an alternate partition

$$\begin{aligned} CHI &\rightarrow \{US, TAI, JPN, PHL\} \rightarrow \{FRA, ARG\} \rightarrow \{SGP, TUR, AUS\} \\ &\rightarrow \{MYS, THA, UK, GER, MEX\} \rightarrow \{HKG, IDN, BRA, KOR\}, \end{aligned}$$

which is highlighted in blue colour in Table 1. However, combining tests for equality of variances together with the weights estimated by OLS/SURE raises a critical issue related to the Behrens-Fisher problem in this case; see, for example, Moser and Stevens (1992). Hence, we do not impose zeroes implied by this final partition on our estimates. The estimates of the network structure (W) imply strong influences arising from the stock markets of France, US, Australia, Singapore and Chile, which represents different geographical hotspots with both local influences and global contagion effects. There may also be some potential spatial strong dependence issues (Pesaran, 2006) which we have not addresed at the moment.

While the above weights relate to contemporaneous Rubin causation or the potential outcome model (Rubin, 2005), Granger (1969) or temporal causation can be inferred from the estimates of the VAR part of the model. As expected, a predominance of statistically significant lags along the diagonal reflect strong autoregressive effects. However, strong cross-border influences are also observed for some markets, such as Malaysia, Hong Kong and the US. Overall, the estimates make good sense and provide exciting new inferences on spillovers in stock returns. Finally, the goodness-of-fit of the PR-DAG model is much improved over the benchmark reduced form VAR, as indicated by comparison of R^2 values. The MSE of the RR_DAG model is also reduced, in-sample, by 17 percent, which is very promising from a finance point of view.

5 Conclusion

The recent literature has placed considerable attention on the identification of networks through which spillovers in financial markets happen. We propose a partially recursive DAG (PR-DAG) as an alternative to the rather more stringent recursive DAG model considered recently in the literature. This represents a good balance against the poly-tree which potentially provides far too sparse a network in many application context. Under the PR-DAG causal structure, we develop inferences on the partial causal order as a directed partition of the units or variables. Application to weekly data on stock price returns across several markets worldwide provide exciting new evidences on the nature and magnitude of spillover effects and the structure of the network. While the inferences are motivated ny panel data, the methods are also potentially useful for identification and inferences on structural vector autoregressions and factor-based data-rich models in macreoceonomic contexts.

References

[1] Bailey, N., Holly, S. and Pesaran, M. H. (2016). A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence. *Journal of Applied Econometrics* **31**(1), 249-280.

[2] Banerjee, A., Bhattacharjee, A. and Martins, J.M. (2017). The Nature of Causation between Radiative Forcing and Temperature. *Mimeo*. Spatial Economics & Econometrics Centre (SEEC), Heriot-Watt University, UK.

[3] Banerjee, S. and Roy, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*. CRC Press: Boca Raton FL.

[4] Basak, G.K., Bhattacharjee, A. and Das, S. (2017). Causal ordering and inference on acyclic networks: Is there a (contemporaneously) leading portfolio?. *Empirical Economics* (Special Issue in Honour of Ingmar Prucha) (forthcoming).

[5] Baumeister, C. and Hamilton, J.D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica* **83**(5), 1963-1999.

[6] Bekaert, G., Hodrick, R.J. and Zhang, X. (2009). International stock return comovements. *Journal of Finance* **64**, 2591-2626.

[7] Bernanke, B.S., Boivin, J. and Eliasz, P. (2005). Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics* **120**, 387-422.

[8] Bhattacharjee, A. and Holly, S. (2013). Understanding interactions in social networks and committees. *Spatial Economic Analysis* **8**(1), 23-53.

[9] Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics* **43**(4), 617-634.

[10] Boivin, J. and Giannoni, M.P. (2009). Global Forces and Monetary Policy Effectiveness. In: Gali, J. and Gertler, M. (Eds.), *International Dimensions of Monetary Policy*, Chap. 8, 429-478. Chicago: University of Chicago Press.

[11] Canova, F. (2005). The Transmission of US Shocks to Latin America. *Journal of Applied Econometrics* **20**, 229-251.

[12] Christiano, L.J., Eichenbaum, M. and Evans, C.L. (1999). Monetary Policy Shock: What Have We Learned and to What End?. In: Taylor, J.B. and Woodford, M. (Eds.), *Handbook of Macroeconomics*, vol. 1, Part A, Elsevier, 65-148.

[13] Christiano, L.J., Eichenbaum, M. and Vigfusson, R. (2007). Assessing Structural VARs. *NBER Macroeconomics Annual 2006*, 1-106.

[14] Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999). *Probabilistic Networks and Expert Systems*. Springer.

[15] Cushman, D.O. and Zha, T. (1997). Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rate. *Journal of Monetary Economics* **39**, 433-448.

[16] Dasgupta, S. (1999). Learning polytrees. In: *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence* (UAI-99), 134-141.

[17] Dees, S., di Mauro, F., Smith, V. and Pesaran, M.H. (2007). Exploring the International Linkages of the Euro Area: A Global VAR Analysis. *Journal of Applied Econometrics* **22**, 1-38.

[18] Del Negro, M. and Primiceri, G.E. (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *Review of Economic Studies* **82**(4), 1342-1345.

[19] Dickey, D.A., and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427-431.

[20] Diebold, F.X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal* **119**(534), 158-171.

[21] Diebold, F.X. and Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* **182**(1), 119-134.

[22] Edelstein, P. and Kilian, L. (2009). How sensitive are consumer expenditures to retail energy prices?. *Journal of Monetary Economics* **56**(6), 766-779.

[23] Eichenbaum, M. and Evans, C.L. (1995). Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates. *Quarterly Journal of Economics* **110**, 975-1010.

[24] Engle, R., Ito, T. and Lin, W. (1990). Meteor Showers or Heat Waves? Heteroscedastic Intra-daily Volatility in the Foreign Exchange Market. *Econometrica* **58**(3), 525-542.

[25] Evans, R. (2017). Margins of discrete Bayesian networks. *Annals of Statistics* (forthcoming).

[26] Faust, J. (1992). When are variance ratio tests for serial dependence optimal?. *Econometrica* **60**(5), 1215-1226.

[27] Faust, J. and Rogers, J.H. (2003). Monetary Policy's Role in Exchange Rate Behavior. *Journal of Monetary Economics* **50**, 1403-1423.

[28] Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature* **49**(4), 938-960.

[29] Giacomini, R. and Kitagawa, T. (2015). Robust inference about partially identified SVARs. Working paper, University College London.

[30] Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3), 424-438.

[31] Grilli, V. and Roubini, N. (1995). Liquidity and Exchange Rates: Puzzling Evidence from the G-7 Countries. Working Paper, Yale University.

[32] Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal* **3**, 148-161.

[33] Hamao, Y., Masulis, R. and Ng, V. (1990). Correlations in Price Changes and Volatility across International Stock Markets. *Review of Financial Studies* **3**(2), 281-307.

[34] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

[35] Harary, F. (1994). *Graph Theory*. Addison-Wesley: Reading (MA).

[36] Ibrahim, B.M. and Brzeszczyński, J. (2009). Inter-regional and Region-specific Transmission of International Stock Market Returns: The Role of Foreign Information. *Journal of International Money and Finance* **28**(2), 322-343.

[37] Im, K.S., Pesaran, M.H. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* **115**, 53-74.

[38] Inoue, A. and Kilian, L. (2013). Inference on impulse response functions in structural VAR models. *Journal of Econometrics* **177**(1), 1-13.

[39] Kilian, L. (2009). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *American Economic Review* **99**(3), 1053-1069.

[40] Kilian, L. (2013). Structural Vector Autoregressions. In: Hashimzade, N. and Thornton, M. (Eds.), *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Edward Elgar: Cheltenham UK, 515-554.

[41] Kilian, L. and Murphy, D.P. (2012). Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models. *Journal of the European Economic Association* **10**(5), 1166-1188.

[42] Kim, S. (2001). International Transmission of U.S. Monetary Policy Shocks: Evidence from VARs. *Journal of Monetary Economics* **48**, 339-372.

[43] Kim, S. and Roubini, N. (2000). Exchange Rate Anomalies in the Industrial Countries: A Solution with a Structural VAR Approach. *Journal of Monetary Economics* **45**, 561-586.

[44] Klößner, S. and Wagner, S. (2014). Exploring all VAR orderings for calculating spillovers? yes, we can! – a note on Diebold and Yilmaz (2009). *Journal of Applied Econometrics* 29(1), 172-179.

[45] Levene, H. (1960). Robust tests for equality of variances. In: I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow and H.B. Mann (Eds.), *Contributions to probability and statistics* (Stanford Univ. Press, Stanford), 278-292.

[46] Lim, T.S. and Loh, W.Y. (1996). A comparison of tests of equality of variances. *Computational Statistics & Data Analysis* **22**(3), 287-301.

[47] Lippi, M. and Reichlin, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics* **63**(1), 307-325.

[48] Moser, B.K. and Stevens, G.R. (1992). Homogeneity of variance in the two-sample means test. *American Statistician* **46**(1), 19-21.

[49] Mumtaz, H. and Surico, P. (2009). The transmission of international shocks: a factor-augmented VAR approach. *Journal of Money, Credit and Banking* **41**(s1), 71-100.

[50] Mumtaz, H. and Theodoridis, K. (2017). The changing transmission of uncertainty shocks in the US: An empirical analysis. *Journal of Business and Economic Statistics* (forthcoming).

[51] ONS (2007). Consumer Trends. Office of National Statistics, HMSO.

[52] Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. 2nd Edition. Cambridge University Press.

[53] Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**(4), 967-1012.

[54] Pesaran, M.H., Schuermann, T. and Weiner, S.M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economic Statistics* **22**(2), 129-162.

[55] Pötscher, B.M. and Prucha, I.R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer.

[56] Primiceri, G.E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* **72**(3), 821-852.

[57] Ramey, V.A. (2015). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics* **2**.

[58] Rebane, G. and Pearl, J. (1987). The recovery of causal poly-trees from statistical data. In: *Proceedings of the 3rd Conference on Uncertainty in Artificial Intelligence* (UAI-87), 222-228.

[59] Rubin, D. (2005). Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association* **100**(469), 322-331.

[60] Rubio-Ramírez, J.F., Waggoner, D.F. and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies* **77**(2), 665-696.

[61] Sargent, T.J. and Sims, C.A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In: Sims, C.A. (Ed.), *New Methods in Business Cycle Research: Proceedings from a Conference*, Minneapolis MN: Federal Reserve Bank of Minneapolis, 45-109.

[62] Scholl, A. and Uhlig, H. (2005). New Evidence on the Puzzles: Results from Agnostic Identification on Monetary Policy and Exchange Rates. *SFB 649 Discussion Paper* **2005-037**, Humboldt University.

[63] Sedgewick, R. and Wayne, K. (2011). *Algorithms* (4th ed.), Addison-Wesley.

[64] Sims, C.A. (2012). Statistical modeling of monetary policy and its effects. *American Economic Review* **102**(4), 1187-1205.

[65] Sims, C.A. and Zha, T. (2006). Were There Regime Switches in US Monetary Policy?. *American Economic Review* **96**, 54-81.

[66] Stock, J.H. and Watson, M.W. (1989). New indexes of coincident and leading economic indicators. In: Blanchard, O.J. and Fischer, S. (Eds.), *NBER Macroeconomics Annual 1989*, vol. 4, Cambridge MA: MIT Press, 351-394.

[67] Stock, J.H. and Watson, M.W. (2001). Vector autoregressions. *Journal of Economic Perspectives* **15**, 101-115.

[68] Stock, J.H. and Watson, M.W. (2015). Factor models and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics* **2**.

[69] Thurman, W.N. and Fisher, M.E. (1988). Chickens, eggs, and causality, or which came first?. *American Journal of Agricultural Economics* **70**(2), 237-238.

[70] Tian, J. and Pearl, J. (2002). On the testable implications of causal models with hidden variables. In: *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence* (UAI-02), 519-527.

[71] Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics* **52**(2), 381-419.