Causal ordering and inference on acyclic networks

CrossMark

Causal ordering and inference on acyclic networks

Gopal K. Basak[1] · Arnab Bhattacharjee[2] · Samarjit Das[3]

© The Author(s) 2018

Abstract This paper develops a new identification result for the causal ordering of observation units in a recursive network or directed acyclic graph. Inferences are developed for an unknown spatial weights matrix in a spatial lag model under the assumption of recursive ordering. The performance of the methods in finite sample settings is very good. Application to data on portfolio returns produces interesting new evidences on the contemporaneous lead–lag relationships between the portfolios and generates superior predictions.

Keywords Spatial weights matrix · Directed acyclic graphs · Recursive ordering · M-estimation · Portfolio returns · Partial identification

JEL Classification C63 · G14 · C31 · C58 · G17

✉ Arnab Bhattacharjee
  a.bhattacharjee@hw.ac.uk

  Gopal K. Basak
  gkb@isical.ac.in

  Samarjit Das
  samarjit@isical.ac.in

[1] Stat-Math Unit, Indian Statistical Institute, Kolkata, India

[2] Spatial Economics and Econometrics Centre (SEEC), Heriot-Watt University, Edinburgh, UK

[3] Economics Research Unit, Indian Statistical Institute, Kolkata, India

1 Introduction

Considerable attention has been placed in the recent literature to the estimation of an unknown spatial weights matrix (W) in spatial regression models, particularly the spatial autoregressive (SAR, or spatial lag) model or the spatial error model (SEM) (Anselin 1988, 2001). This paper considers estimation of W in contexts where the observation units form a directed acyclic graph, with an application to financial market returns. Denoting Y as a k-dimensional outcome variable and X as exogenous (or weakly exogenous) regressors observed over k distinct units (either points or regions in a spatial domain or vertices in a network), the SAR model is described as

$$Y_t = \lambda W Y_t + X_t \beta + \epsilon_t, \tag{1.1}$$

where WY_t is the spatial lag of Y_t, λ is the spatial autoregressive parameter, ϵ is a vector of idiosyncratic errors, and W is a $k \times k$ matrix with zero diagonal elements. The subscript t ($t = 1, 2, \ldots$) has the interpretation of replications and, specifically for panel data, can represent time periods. The corresponding SEM model with spatial autoregressive errors omits the spatial lag of Y_t, but includes a spatial lag of the error term.

In spatial applications, off-diagonal elements of W represent the (directed) interaction between outcomes at a pair of locations and are typically measured either by contiguity or inverse functions of distances (Anselin 1988; Fingleton 2003). In network applications (such as Ballester et al. 2006), W represents the strength of interactions between economic agents and are typically measured by observed indicators of reciprocity or membership of social groups. Then, the SAR or SEM models are estimated either using GMM methods (Kelejian and Prucha 1998, 1999), or maximum likelihood (ML)/ quasi-ML (Lee 2004; Gupta and Robinson 2015), or Bayesian methods (LeSage and Pace 2009).

Accuracy of the above estimation is quite severely conditioned on correct measurement of W, at the same time as there is substantial uncertainty in applications as to the correct measures of the spatial weights (Fingleton 2003; Giacomini and Granger 2004). In this context, estimation of spatial weights can be particularly useful. Bhattacharjee and Jensen-Butler (2013) show that, in a spatial lag or spatial error model, W is only partially identified from an estimated spatial variance–covariance matrix. Hence, additional structural assumptions are required for inference, and empirical verification of these assumptions is an important research question.

Beenstock and Felsenstein (2012) and Bhattacharjee and Jensen-Butler (2013) proposed estimation of W in a spatial panel setting under the assumption of symmetric spatial weights. Bhattacharjee and Holly (2011) and Pesaran and Tosetti (2011) discuss estimation in contexts where there is, in addition to stationary W-based spatial dependence, also spatial strong dependence modelled through a factor structure. Bhattacharjee et al. (2012) extend the above methods to pure cross section (spatial) data setting, and Bhattacharjee and Holly (2013) propose estimation under moment conditions drawing upon connections with the system GMM literature (Arellano and Bond 1991; Blundell and Bond 1998). Finally, Ahrens and Bhattacharjee (2015) and Bailey et al. (2016) propose estimation under the assumption of a sparse spatial weights

matrix. Most of the above inference methods are based on sample moments. They are therefore closely related to GMM estimators in Kelejian and Prucha (1998, 1999) and Kapoor et al. (2007). In fact, the statistical properties of these estimators can be developed using asymptotic results in Pötscher and Prucha (1997) and Kelejian and Prucha (2001).[1]

This paper is a contribution towards the same literature but with a different identification restriction. Specifically, we consider a SAR model (1.1) where the interactions, modelled by W, represent a directed acyclic graph (DAG). While causal inference on DAG is not new to the literature, our main contribution is towards identification of the "true" DAG structure from the data. A DAG is a finite directed graph with no directed cycles (Harary 1994; Pearl 2009). We consider the most general form of DAG, where the units have a recursive ordering. In related work, Rebane and Pearl (2013) proposed an identification procedure for poly-trees and Bhattacharjee (2017) developed a method to recover recursive ordering under homoscedastic errors. Our identification result and methodology are different. In addition, we extend inferences in several ways: (a) provide inferences on recursive ordering when error variance varies by unit (or location); (b) develop estimators and inference for W; and (c) propose indices to quantify the degree to which a network is recursive (acyclic). The Monte Carlo performance of the methods in finite samples is very good. Finally, we apply our methods to data on returns of 25 stock portfolios constructed by Fama and French (1993), and find evidence of (contemporaneous) lead–lag relationships between the portfolios in the nature of recursive ordering, which in turn aids identifying leading and trailing portfolios in the network. Forecasts from the estimated model are far superior relative to a benchmark VAR. The rest of the paper is structured as follows. Section 2 discusses our model and inference methods, followed by a Monte Carlo study in Sect. 3 and the application to the Fama–French portfolios in Sect. 4. Finally, Sect. 5 concludes.

2 Identification and inference

In this section, we discuss our theoretical results on identification and inference, and their implications for empirical analysis. We collect assumptions and theory in the first subsection and discuss the use of this theory for applications in the second.[2]

2.1 Assumptions and theoretical results

Consider initially the pure SAR model (without regressors) as

$$Y_t = WY_t + \epsilon_t, \tag{2.1}$$

---

[1] Pötscher and Prucha (1997) is an excellent monograph providing very general theory on M-estimation. This theory can be used to develop many of the moment based results from the current literature in spatial econometrics.

[2] The structure and improved presentation of the section owes much to the kind suggestions of an anonymous reviewer.

where ϵ_t is a vector of idiosyncratic errors for $t = 1, 2, \ldots, n$, with zero mean vector and uncorrelated components, and W is an unknown spatial weights matrix on which we aim to conduct inference. In comparison with (1.1), there are no regressors X. Further, since W is unknown in the current context, λ and W are not separately identified, and therefore without loss of generality, we set $\lambda \equiv 1$ (Bhattacharjee and Jensen-Butler 2013). The reduced form of (2.1) is $Y_t = (I - W)^{-1}\epsilon_t$. Then, our key identification restriction—recursive structure—is expressed in the following assumption.

Assumption 1 (*Recursive Structure*) There exists some permutation of the elements (spatial/observation units) in Y_t, say $Y_t^{[P]} = \left(Y_t^{[1]}, \ldots, Y_t^{[k]}\right)$, for which the corresponding spatial weights matrix $W^{[P]}$ is a lower triangular $k \times k$ matrix with zero diagonal elements. That is, $W^{[P]} = \left(\left(w_{ij}^{[P]} : w_{ij}^{[P]} = 0 \text{ if } j \geq i\right)\right)_{i,j=1,\ldots,k}$. Denote the permutation associated with the vector $Y_t^{[P]}$ by $\tau = ([1], [2], \ldots, [k])$, where τ is a permutation of the finite sequence $1, 2, \ldots, k$. Further, we assume that the permutation τ is unique, that is, all elements in the principal subdiagonal of $W^{[P]}$ are nonzero.

Assumption 1 is the key assumption in this paper and requires several qualifying comments. First, the recursive structure above constitutes the most general form of a DAG as well as a generalisation of a causal poly-tree (Cowell et al. 1999). Here, we provide a very brief discussion of properties of DAGs and related graphs that are useful for our work.[3] A DAG is a directed graph, which means that the edges have arrows indicating the direction of flow between vertices. A DAG consists of finitely many vertices and edges, with each edge directed from one vertex to another, such that there is no way to start at any vertex and follow a directed sequence of edges that eventually loops back to the same vertex again. In the context of Assumption 1, each unit $i = 1, 2, \ldots, k$ represents a vertex, and w_{ij} represents a directed edge from unit j to unit i.

A causal graph is typically represented as a DAG, where the directed edges represent the direction of causal influences between variables, which are represented as vertices. Clearly, estimation based on solutions to linear equations, such as OLS or Yule–Walker equations (Yule 1927; Walker 1931; Brockwell and Davis 1991), can provide causal inference only if there is no directed cycle. This is the central reason why DAGs have been very useful in causal models.

Consider a DAG as defined above. Obviously, for a pair of units i and j, only one of w_{ij} or w_{ji} can be nonzero; otherwise, there will be a cycle between the two units. Further, a DAG implies more than just the above restriction, because a cycle can potentially involve multiple units: for example, $h \rightarrow i \rightarrow j \rightarrow h$. In order to negate any such cycles, a DAG should have a topological ordering, a sequence of the vertices such that every edge is directed from earlier to later in the sequence. However, the ordering need not be complete; for example, two vertices can in principle have

---

[3] We are grateful to an anonymous reviewer for encouraging us to include in the paper a short discussion of DAGs and their use in applied work. For further details and discussion in an applied economics context, we refer to Bessler and Lee (2002) and for a classic treatment in the graph theory context, please see Harary (1994).

a common child (a vertex that comes later in the ordering), but these two vertices may not be topologically connected, that is, they may be neither parent nor child of each other. By contrast, the recursive structure represents a complete ordering, $\tau = ([1], [2], \ldots, [k])$, of the units $i = 1, 2, \ldots, k$. In this sense, recursive structure is a generalisation of a DAG where, by restricting some of the lower triangular elements of $W^{[P]}$ to zero, one can generate any DAG from a conformable recursive structure. A poly-tree is a specific kind of DAG whose underlying undirected graph is a tree; hence, the recursive structure is a generalisation of the poly-tree as well.

Second, as the above discussion highlights, recursive ordering is quite general and useful in many contexts. However, while causal inference typically assumes knowledge of the true ordering of variables in $Y_t^{[P]}$, this paper develops a method to identify this ordering from the data, together with inferences on the corresponding weights matrix $W^{[P]}$. The second part of the Assumption 1 is important and provides a necessary and sufficient condition for $W^{[P]}$ to map to a unique ordering. Consider, for example, the case $k = 2$ and $W = 0$. In this case, both orderings are consistent with W, and there is no unique mapping from $W^{[P]}$ to τ. Obviously, this possibility needs to be ruled out.

Third, and as emphasised above, Assumption 1 negates the possibility of causal loops. Hence, if the correct recursive ordering $Y_t^{[P]}$ were known, the corresponding weights matrix $W^{[P]}$ is identified and can be estimated either by solving the Yule–Walker equations or by OLS, which are closely related. For this reason, DAGs and in particular recursive identification have been popular in much applied work: in macroeconomics (Bessler and Lee 2002; Bernanke et al. 2005), in Bayesian networks (Cowell et al. 1999; Spiegler 2016) and causal graphs (Spirtes et al. 1993; Pearl 1995, 2009). However, while most of this literature takes the DAG (or recursive structure) as a given and then conducts causal inference, our work is fundamentally different in that we propose inference to identify the structure of the DAG itself.

Finally, note that under Assumption 1, 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 to assume 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 results; hence, we make Assumption 2.

Assumption 2 (*Spatial granularity condition*) The row and column norms of W are bounded (in absolute value) by 1.

In practice, Assumption 2 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 (2.1), denoting covariance matrix by Cov(.), we have:

$$
\begin{aligned}
\text{Cov}(Y_t) &= Cov[(I - W)^{-1} \epsilon_t] \\
&= (I - W)^{-1} [\text{Cov}(\epsilon_t)]((I - W)')^{-1} \\
&= (I - W)^{-1} \Sigma_\epsilon ((I - W)')^{-1}
\end{aligned}
\tag{2.2}
$$

where Σ_ϵ is the covariance matrix of ϵ_t. Since the components of ϵ_t are independent by assumption, $\Sigma_\epsilon = diag\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right)$.

Theorem 1 *Let $\{Y_t\}$ be generated by the pure-SAR model* (2.1). *Under Assumption* 1, *if $\{\epsilon_t\}$ is homoscedastic, that is, $\Sigma_\epsilon = \sigma^2 I$, where I is the identity matrix (of the same order as that of ϵ), then W can be retrieved fully from the covariance matrix of Y.*

Proof For notational simplicity, we omit the suffix t. Consider the precision matrix (inverse of the covariance matrix) of Y, that is, $(\text{Cov}(Y))^{-1} = (I - W)'\Sigma_\epsilon^{-1}(I - W) = \sigma^{-2}(I - W)'(I - W) = \sigma^{-2}(I - (W' + W) + W'W)$. Since W is lower triangular (or, some rearrangement of its rows and columns makes it a lower triangular matrix), exactly one of its columns, say i_k, is a zero vector. Hence, exactly one of the diagonal element of $W'W$ is zero, since the diagonal element of $W'W$ is the inner product of the columns with itself. The corresponding diagonal element of the precision matrix of Y is σ^{-2}, which is less than any other diagonal, since diagonal elements of $(W' + W)$ are all zero. The implication is that residual variance of the i_kth element is the smallest, meaning that the i_kth element of Y depends highly on the other elements of Y. Hence, the k th element in the ordering is correctly identified.

Now, delete the i_kth row of Y and find the variance of the reduced Y (say, Y_{-i_k}) vector. Then, its inverse would be of the same form, that is, $(\text{Cov}(Y_{-i_k}))^{-1} = \sigma^{-2}(I - (W'_{-i_k} + W_{-i_k}) + W'_{-i_k}W_{-i_k})$, where W_{-i_k} is found deleting i_kth row and i_kth column of W. We again find the minimum diagonal element which is σ^{-2}, since exactly one diagonal element of $W'_{-i_k}W_{-i_k}$ is zero. Take this diagonal element and say it is i_{k-1}th element. Delete that from the Y_{-i_k} vector and continue recursively, till we arrive at the final element of Y, which we denote as the i_1th element. Thus, the downward ordering will be $i_1, i_2, \ldots, i_k$ and the corresponding rearrangement rows and columns of W is a lower triangular matrix.

Once the order of Y is known, for which W is lower triangular, one can now solve Yule–Walker type equations to solve for W. For example, it is now clear that $\text{Var}(Y_{i_1}) = \text{Var}(\epsilon_{i_1}) = \sigma^2$, $\text{Cov}(Y_{i_2}, Y_{i_1}) = w_{i_2,i_1}\text{Var}(Y_{i_1}) = w_{i_2,i_1}\sigma^2$, as ϵ_{i_2} and ϵ_{i_1} are uncorrelated which in turn implies ϵ_{i_2} and Y_{i_1} are uncorrelated. Thus, $w_{i_2} = \text{Cov}(Y_{i_2}, Y_{i_1})/\text{Var}(Y_{i_1}) = \text{Cov}(Y_{i_2}, Y_{i_1})/\sigma^2$, and $\text{Var}(Y_{i_2}) = w_{i_2,i_1}^2\text{Var}(Y_{i_1}) + \text{Var}(\epsilon_{i_2}) = w_{i_2,i_1}^2\sigma^2 + \sigma^2$. Further, for $j \geq 3$,

$$\text{Cov}(Y_{i_j}, (Y_{i_1}, \ldots, Y_{i_{j-1}})) = \text{Cov}((w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}})(Y_{i_1}, \ldots, Y_{i_{j-1}})', (Y_{i_1}, \ldots, Y_{i_{j-1}}))$$

$$\text{since } \epsilon_{i_j} \text{ and } (Y_{i_1}, \ldots, Y_{i_{j-1}}) \text{ are uncorrelated}$$

$$= (w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}})\text{Cov}((Y_{i_1}, \ldots, Y_{i_{j-1}})') \tag{2.3}$$

Hence $(w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}}) = \text{Cov}(Y_{i_j}, (Y_{i_1}, \ldots, Y_{i_{j-1}}))(\text{Cov}((Y_{i_1}, \ldots, Y_{i_{j-1}})'))^{-1}$. and

$$\text{Var}(Y_{i_j}) = \text{Var}((w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}})(Y_{i_1}, \ldots, Y_{i_{j-1}})') + \text{Var}(\epsilon_{i_j})$$

$$= (w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}})\text{Cov}((Y_{i_1}, \ldots, Y_{i_{j-1}})')((w_{i_j,i_1}, \ldots, w_{i_j,i_{j-1}})') + \sigma^2$$

$$= \text{Cov}(Y_{i_j}, (Y_{i_1}, \ldots, Y_{i_{j-1}}))(\text{Cov}((Y_{i_1}, \ldots, Y_{i_{j-1}})'))^{-1}\text{Cov}(Y_{i_j}, (Y_{i_1}, \ldots, Y_{i_{j-1}})') + \sigma^2. \tag{2.4}$$

Hence, the result follows.

The constructive proof of Theorem 1 provides our estimator of recursive ordering. Next, we discuss large sample properties of the estimator. First, consider the simplest case, $k = 2$. Our estimated ordering is based on relative variances; for finite sample data, if $s_2^2 > s_1^2$, then $\widehat{\tau} = (1, 2)$, or otherwise if $s_2^2 < s_1^2$, then $\widehat{\tau} = (2, 1)$, where the sample covariance matrix is

$$\widehat{\mathrm{Cov}}(Y) = \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix}.$$

In the first case, $\widehat{w}_{12} = 0$, $\widehat{w}_{21} = s_{12}/s_1^2$, and in the latter case, $\widehat{w}_{12} = s_{12}/s_2^2$, $\widehat{w}_{21} = 0$.[4] Without loss of generality, assume $\tau_0 = (1, 2)$, so that by the Yule–Walker equations the true W_0 satisfies $w_{12} = 0$, $w_{21} = \sigma_{12}/\sigma_1^2$. Clearly, $\widehat{W}$ is consistent because the moment estimators are consistent, so that in large samples, the correct ordering between σ_1^2 and σ_2^2 is evident from the sample moments. Also, as a function of sample moments, $\widehat{W} = \vartheta\left(s_1^2, s_2^2, s_{12}\right)$, the estimate of the nonzero element is a continuous function of the arguments, but is not differentiable at the margin $s_1^2 = s_2^2$.

This has important implications for asymptotic and finite sample properties of the estimator. Continuity provides standard asymptotic distributions but non-differentiability ensures that the finite sample distributions are non-regular. Specifically, since estimation is by solving Yule–Walker equations under the estimated ordering which is correctly identified in large samples, as $n \to \infty$, $\widehat{w}_{12}$ will have a degenerate distribution at zero, while $\widehat{w}_{21}$ will be asymptotically Normal with the correct mean. However, in finite samples, there will be instances when $s_2^2 < s_1^2$ because of sampling variations, and the finite sample distribution will be bimodal, where the "wrong" mode will recede with increasing sample size.

It is clear from the above discussion that the sampling distributions are bimodal in finite samples, and hence the derivation of asymptotic results is somewhat complex in this case. A full development of the asymptotic theory is beyond the scope of this paper. Below we provide an intuitive discussion focusing on estimation by least squares, but outline the nature of technical arguments that are required for fully developed theory. Here, the correct ordering τ may be viewed as a nuisance parameter, and the data are dependent and endogenous. As mentioned before, Pötscher and Prucha (1997) provide an excellent compendium of results for M-estimation that are useful in this context, and which we will rely heavily upon.

Setting our notations in line with Pötscher and Prucha (1997), consider a data generating process $\{z_t : t \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathbb{U}, P)$, with z_t taking values in a non-empty measurable space $(Z, \Im)$, and let (B, ρ_B) and (T, ρ_T) be non-empty metric spaces. Here, z_t has the interpretation of current and past values of endogenous and exogenous variables. In the case of our pure-SAR model (2.1), $z_t \equiv Y_t$, but one can have additional regressors, both exogenous or weakly exogenous (such as lags of Y_t). Z is a Euclidean space or a subset thereof, B is the space of parameters

---

[4] The indeterminate case, $s_2^2 = s_1^2$, has zero measure in finite samples and is negated in the population by the uniqueness of τ (Assumption 1).

of interest,[5] and T is the space of nuisance parameters; B and T need not be subsets of a Euclidean space, which is important in our case because the nuisance parameter is the permutation τ. Let $Q_n(Y_1, Y_2, \ldots, Y_n, \tau, W)$ be a real valued function defined on $Z^n \times T \times B$, where n denotes the sample size. Also assume that Q_n is $\mathbb{U}$-measurable for all $(\tau, W) \in T \times B$. Then, an M-estimator $\widehat{W}_n$ is one that satisfies the relation $Q_n(\widehat{\tau}_n, \widehat{W}_n) = \inf_{W \in B} Q_n(\widehat{\tau}_n, W)$.[6] Now, by Assumption 1, the mapping from W_0 to τ is unique, and once $\widehat{\tau}$ has been estimated, this determines the structural zeroes in $\widehat{W}$, which is then estimated by OLS; hence $\widehat{W} \equiv \widehat{W}(\widehat{\tau})$. Specifically, since estimation is by least squares,

$$Q_n(\widehat{\tau}_n, \widehat{W}_n) = n^{-1} \sum_{t=1}^{n} q_t(Y_t, W(\widehat{\tau}_n))$$

where

$$q_t(Y_t, W(\widehat{\tau}_n)) = [Y_t - W(\widehat{\tau}_n)Y_t]'[Y_t - W(\widehat{\tau}_n)Y_t].$$

Noting that τ_0 is the population analogue of $\widehat{\tau}_n$, define $\overline{Q}_n(\tau_0, W) = E Q_n(\tau_0, W)$, where the expectation is taken over all W's consistent with the ordering τ_0.

Then, the limiting behaviour of $\widehat{W}_n$ can be analysed by relating it to the limiting behaviour of the minimisers $\overline{W}_n$ of $\overline{Q}_n(\tau_0, W(\tau_0))$. Thus, demonstration of consistency basically consists of three steps (Pötscher and Prucha 1997). First, we need to show that $Q_n - \overline{Q}_n \to 0$ uniformly on $T \times B$. Second, the sequence of minimisers $\overline{W}_n$ of $\overline{Q}_n$ need to be identifiably unique, where the definition of identifiable uniqueness is given in Pötscher and Prucha (1997) (Definition 3.1, p.16); see also White (1980) and Domowitz and White (1982). Third, $\overline{Q}_n$ should not be too flat at the minimiser. With the above background, we now proceed to our outline proof of consistency.

Corollary 1 *As in Theorem 1, let $\{Y_t\}$ be given by the pure-SAR model (2.1), let Assumption 1 hold, and assume that $\{\epsilon_t\}$ is homoscedastic, that is, $\Sigma_\epsilon = \sigma^2 I$. Then, the elements of W can be estimated consistently from the data. Order of consistency is the square root of sample size (n).*

Proof Let us first provide an intuitive exposition of the consistency result. Under the assumption of recursive ordering of W, there exists exactly one column of W which is a zero vector, therefore exactly one diagonal element of $W'W$ is zero and all the other diagonals are strictly positive. Therefore

$$(\text{Cov}(Y_t))^{-1} = (I - W)'\Sigma_\epsilon^{-1}(I - W) = \sigma^{-2}(I - W)'(I - W)$$
$$= \sigma^{-2}(I - (W' + W) + W'W)$$

has exactly one diagonal element which is $\sigma^{-2} > 0$ and all the other diagonal elements are of the form $\sigma^{-2}\left(1 + \|w_j\|^2\right)$, where w_j is a nonzero column of W, and hence strictly larger than σ^{-2}. Since the sample covariance of Y is consistent for its

---

[5] For the moment, $W \in B$ is our parameter of interest. We will include additional regressors later.

[6] $Q_n()$ is a function of $Y_1, Y_2, \ldots, Y_n$, but the reference to $Y_1, Y_2, \ldots, Y_n$ is suppressed here for the sake of notational simplicity. Further, $\widehat{W}_n$ needs to satisfy the condition only asymptotically.

population counterpart, the diagonal elements of the inverse of the covariance of Y are also estimated consistently by its sample counterparts, given the parameter values of σ^2 and W. Hence, the minimum is consistently identified from the data. In other words, the element in the kth place is determined consistently. Then, we reduce the dimension by deleting the row given by the minimum of the diagonal and consider the covariance matrix of the reduced dimension and proceed as before to show that the $(k-1)$th element in the recursive order is estimated consistently.

Proceeding sequentially, we recover the entire recursive order consistently. Given the determined order, that is, rearranging the rows of W in triangular form, estimation of W is done through Yule–Walker type of methodology purely based on corresponding sample variances and covariances, which are consistent for their population counterparts. Further, given the correct recursive ordering, the estimates of elements of W are continuous functions of the variances and covariances. Hence, the order of convergence is the same as the sample second moments, which implies $\sqrt{n}$-consistency. Hence the result follows.

The technical proof is based on asymptotic theory discussed in Pötscher and Prucha (1997). The main step is demonstrating identifiable uniqueness of $\overline{W}_n$. Specifically, with the choice of Q_n as outlined above, $\overline{Q}_n$ does not depend on n. This is because $\overline{Q}_n(\tau_0, W) \equiv E R_n(\tau_0, W)$, where

$$R_n(\tau, W) = n^{-1} \sum_{t=1}^{n} [Y_t - W(\tau)Y_t]' [Y_t - W(\tau)Y_t]$$

$$= n^{-1} \sum_{t=1}^{n} \epsilon_t' \epsilon_t + 2n^{-1} \sum_{t=1}^{n} \epsilon_t' [W(\tau) - W_0] Y_t$$

$$+ n^{-1} \sum_{t=1}^{n} Y_t' [W(\tau) - W_0]' [W(\tau) - W_0] Y_t,$$

where the final two terms converge to zero as $n \to \infty$. Then, identifiable uniqueness of $\overline{W}_n \equiv \overline{W}$ implies that $\overline{W}$ is the unique minimiser of $E R_n$. This follows from the Yule–Walker system because in this case, the response function corresponding to $\overline{W}$ is the conditional mean of the outcomes for each unit given the outcomes for other units which come before the index unit in the true recursive order τ; see related discussion in Pötscher and Prucha (1997, p. 17, pp. 23–24). Then, Lemma 3.1 in Pötscher and Prucha (1997) applies and we have consistency, that is, $\widehat{W}_n \xrightarrow{P} W_0$ as $n \to \infty$.   □

Next, we consider asymptotic distribution of the estimator. Here as well, we first present an intuitive proof. Then, we refer to Pötscher and Prucha (1997) for an extensive collection of sufficient conditions for asymptotic normality, which can be used in our context to develop a more technical argument. Here, we abstract from the more technical details.

Corollary 2 *Let $\{Y_t\}$ be given by the pure-SAR model* (2.1) *above, and assume homoscedastic $\{\epsilon_t\}$, that is, $\Sigma_\epsilon = \sigma^2 I$, where I is the identity matrix. If Assumptions* 1

and 2 *hold, then the Yule–Walker estimator of elements of W is jointly asymptotically normally distributed.*

Proof The basic idea behind our intuitive proof is that, given the recursive order of W, the Yule–Walker estimator of W are jointly asymptotically Normal; see, for example, Brockwell and Davis (1991). Now, since the estimated order converges in probability (in fact, it converges almost surely) to τ, asymptotic normality of the estimator follows.

The technical arguments are based on Theorem 11.2 in Pötscher and Prucha (1997) together with Assumptions 11.1, 11.2, 11.3 and 11.4; see also Bierens (1982). The key assumptions are that there exist non-random sequences $\overline{\tau}_n \equiv \tau$ and $\overline{W}_n$ such that $\widehat{\tau}_n \to \tau$ and $\widehat{W}_n \to \overline{W}_n$, together with assumptions on smoothness of the score functions. $\qquad\square$

Next, we relax the assumption of homoscedastic errors. For this purpose, we consider a more general model. Our central example is the popular time-space dynamic model (Anselin 2001; Elhorst 2001; Giacomini and Granger 2004; Baltagi et al. 2014) where, in addition to the spatial lag (WY_t), there is also an additional regressor—a temporal lag ($y_{i,t-1}$) with heterogenous persistence across the units

$$y_{it} = \lambda_i y_{i,t-1} + W_{i.} y_t + \epsilon_{it}, i = 1, \ldots, k, t = 1, \ldots, n,$$

where $W_{i.}$ denotes the ith row of W. The model can be written in vector notation as

$$Y_t = WY_t + \Lambda Y_{t-1} + \epsilon_t, \tag{2.5}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a diagonal matrix. Then, the reduced form of (2.5) is given by

$$Y_t = (I - W)^{-1} \Lambda Y_{t-1} + (I - W)^{-1} \epsilon_t. \tag{2.6}$$

The reduced form (2.6) can be estimated as a vector autoregressive (VAR) model (Hamilton 1994; Stock and Watson 2001). Then, we have the following result.

Theorem 2 *Let* $\{Y_t\}$ *be generated by the time-space dynamic model* (2.5)*. Under Assumption* 1*, if* $\{\epsilon_t\}$ *is heteroscedastic, that is,* $\Sigma_\epsilon = diag\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right)$*, then the error standard deviations can be estimated from the estimates of the reduced form* (2.6)*.*

Proof We write the reduced form (2.6) as

$$Y_t = AY_{t-1} + u_t;$$
$$A = (I - W)^{-1} \Lambda, \mathrm{Cov}(u_t) = (I - W)^{-1} \Sigma_\epsilon (I - W)^{-1'}.$$

The reduced form (VAR) estimates provide $\widehat{A}$ and $\widehat{\mathrm{Cov}}(u_t)$. Note that the diagonal elements of $(I - W)^{-1}$ are ones, and Λ is a diagonal matrix. Then, the ith diagonal element of Λ must be the corresponding diagonal element of $A \equiv \left((a_{ij})\right)_{i,j=1,\ldots,k}$; 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 $\widehat{\mathrm{Cov}(u_t)}$. Since $\mathrm{Cov}(u_t) = (I - W)^{-1} \Sigma_\epsilon (I - W)^{-1\,\prime}$, the diagonal elements of $(I - \widehat{W})\widehat{\mathrm{Cov}(u_t)}(I - \widehat{W})'$ provide consistent estimators for the elements of Σ_ϵ, that is, $\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_k^2$. $\qquad\square$

2.2 Discussion of theoretical results and implementation

The above results suggest an applied approach consisting of three steps. Estimating the idiosyncratic error standard deviations is the first step, and for this we use Theorem 2, and estimate using an unconstrained reduced form VAR (2.6). These estimated standard deviations are then used to convert the model to a homoscedastic error case. In the second step, we infer on the recursive ordering. Finally, in step three, we estimate W, as well as other parameters in the model, by solving Yule–Walker equations where zeroes implied by the estimated recursive order are imposed. The identification result underlying steps 2 and 3 is given in Theorem 1, while consistency and asymptotic normality are established in Corollaries 1 and 2, respectively.[7]

Several comments are in order for our Theorem 1. First, while we establish the result based on a pure SAR model (2.1), we can have additional regressors in the model. This does not change our core order identification result.

Second, once the recursive ordering is estimated, estimation of the elements of W can proceed based on an estimate of $\mathrm{Cov}(Y)$ from sample data. The recursive structure in Assumption 1 suggests simply solving Yule–Walker equations or OLS under the estimated ordering $\widehat{\tau}_n = (i_1, i_2, \ldots, i_k)$; both methods are closely related. Then, our estimator for W is $\widehat{W}(\widehat{\tau}_n)$ estimated by Yule–Walker method under the ordering $\widehat{\tau}_n$.

Third, the identification result in Theorem 1 is based on the diagonal elements of the precision matrix. It is interesting to note that, for the closely related conditional autoregressive (CAR) model, the off-diagonal elements of precision matrix are zeroes precisely for units that are conditionally independent. It has been noted that the CAR and SAR models are observationally very similar (Wall 2004) and difficult to distinguish in the data. Thus, the above result provides an explicit link between the two popular models SAR and CAR and is new to the literature.

---

[7] As pointed out by an anonymous reviewer, we could alternatively use constrained VAR estimation in step three, and then iterate steps 1–3 until convergence. This would be more efficient.

Fourth, alternate identification of recursive structure can be based directly on the covariance matrix, rather than implicitly through the precision matrix as in Theorem 1. An alternate identification result using the structure of the covariance matrix is provided in the working paper version, together with extensive Monte Carlo results; see also Bhattacharjee (2017).

Finally, the homoscedasticity assumption in Theorem 1 is strong, and the same assumption also underlies the identification result in Bhattacharjee (2017). They develop an estimator of the error variance under the correct recursive ordering. Then, they suggest considering all possible permutations of the elements in Y_t and conduct inferences under probability distributions on the model space, that is, distributions over all possible permutations. The idea of considering multiple permutations is also popular in the applied macroeconomics literature (Diebold and Yilmaz 2009; Klößner and Wagner 2014), particularly as a robustness check for recursive ordering suggested by theory. However, this approach can be computationally very expensive, particularly when there are a large number of observation units. In the application considered in this paper, $k = 25$, and hence the above method would require consideration of $k! \approx 1.55 \times 10^{25}$ permutations; clearly, this is not quite feasible.

Then, Theorem 2 shows how the inclusion of an additional covariate aids estimation of the error variances. The initial variables can now be scaled by the estimated standard deviations, and we are back to the homoscedastic error case. Then, we can use Theorem 1 to infer on the recursive order and estimate $\widehat{W}$ using the Yule–Walker representation. The asymptotic properties of the estimator are discussed in Corollaries 1 and 2. There are some important points to note.

First, instead of Y_{t-1}, estimation can use any other regressor as appropriate in an application context. What is important is that this chosen regressor can have potentially heterogenous effects on the different units and should not have spatial Durbin effects, that is, the regressor should have effect on the outcome for the same observation unit, and not other units. For example, in applications with stock return factor models, this could be a Fama–French factor (Fama and French 1993, 1996).

Second, other regressors can, in addition, also be included in the model. In particular, if some regressor is suspected to have spatial Durbin effects, these may be included as additional regressors. Typically, there would also be unit-specific fixed effects, and often, factor variables that account for potential strong spatial dependence (Pesaran 2006; Pesaran and Tosetti 2011).

Third, as the procedure above shows, an unknown W can often be identified from regression coefficients in spatial regression models. This, in our view, is a fact that is known in the spatial econometrics literature but may be not highlighted often enough. In our case, we used an estimator $A^{*-1} = (I - \widehat{W})$ based only on the reduced form estimates. The question is: why should we then aim to estimate W under structural constraints? This is because the above type of estimators is in general not very precise, and the poor estimates are often inadequate for drawing structural inferences.

Fourth, in our setting, there are other competing regression-based estimators for $(I - W)$. For example, one can estimate the structural model (2.1) directly, without explicitly specifying W. This can be done by IV or GMM, using as instruments either

Table 1 Alternate estimators of error variances

$k = 3$	IV-2SLS		IV-3SLS		VAR	
n	500	1000	500	1000	500	1000
Bias	0.105	-0.019	0.097	-0.024	-0.005	-0.003
RMSE	0.292	0.192	0.288	0.193	0.141	0.078

higher-order lags and differences, as suggested by the panel GMM literature (for example, Arellano and Bond 1991; Blundell and Bond 1998), or peripheral units at higher spatial lags (Bhattacharjee and Holly 2013). We consider several such estimators in the Monte Carlo study in the following section. We find that while these estimates for W may be imprecise, they provide consistent estimates for the error variances. Further, the above estimator based on reduced form VAR is more efficient and fit for purpose in our context.

3 Monte Carlo simulations

Based on the above results and discussion, we now proceed to our Monte Carlo studies. First, we evaluate the finite sample performance of three alternate regression-based estimators of the error variances. We consider the time-space dynamic model (2.5) with $k = 3$, Gaussian errors, $\lambda_1 = \lambda_2 = \lambda_3 = 0.6$, initial conditions $Y_0 = 0$, and

$$W = \begin{bmatrix} 0 & 0 & 0.3 \\ 0.1 & 0 & 0.2 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix},$$

so that the true ordering is $\tau_0 = (3, 1, 2)$. Three different estimators are considered: (a) IV-2SLS of the structural model using one lag each of the 3 variables as instruments; (b) IV-3SLS, with the same instruments as above; and (c) unrestricted reduced form VAR as in Theorem 2. Two sample sizes are considered, $n = \{500, 1000\}$, and bias and RMSE of the estimated error variances, based on 500 Monte Carlo replications, are reported in Table 1. Clearly, all the three estimators are consistent, but the VAR estimates are the most efficient. Hence, these VAR-based estimates are used in the application in Sect. 4.

Next, we evaluate the performance of our inferences on recursive order. We consider two models: (i) a spatial model without temporal dynamics and with homoscedastic errors, which we consider as the base model (2.1); and (ii) a time-space dynamic heteroscedastic model (2.5). Data are generated as follows. For model (i), errors (ϵ) are standard Normal, and the elements of the weights matrix (W) are generated from Uniform $(-0.5, 0.5)$. For the time-space heteroscedastic model (ii), the (temporal) persistence coefficient (λ_i) is generated as Uniform $(0, 1)$ independently for $i = 1, \ldots, k$, and likewise error standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_k$ as independent Uniform $(1, 2)$. In the output reported in Table 2, columns represent

Table 2 Monte Carlo study on estimation of recursive order

$n \backslash k$	Spatial homoscedastic model				Time-space heteroscedastic			
	6	10	20	30	6	10	20	30
(a) Correlation								
200	0.288	0.465	0.690	0.821	0.285	0.326	0.518	0.635
500	0.442	0.592	0.834	0.910	0.407	0.437	0.608	0.738
1000	0.522	0.723	0.896	0.947	0.484	0.511	0.679	0.784
2000	0.635	0.790	0.934	0.966	0.589	0.589	0.745	0.850
5000	0.747	0.871	0.961	0.982	0.683	0.715	0.809	0.886
(b) Similarity								
200	0.222	0.166	0.119	0.100	0.273	0.173	0.112	0.084
500	0.278	0.201	0.155	0.141	0.309	0.202	0.134	0.113
1000	0.297	0.241	0.194	0.188	0.361	0.226	0.157	0.139
2000	0.378	0.298	0.260	0.242	0.405	0.283	0.206	0.164
5000	0.457	0.405	0.362	0.348	0.487	0.394	0.264	0.224
(c) Concordance								
200	0.613	0.677	0.761	0.821	0.587	0.596	0.673	0.725
500	0.680	0.732	0.834	0.878	0.643	0.643	0.712	0.773
1000	0.713	0.789	0.874	0.911	0.685	0.678	0.743	0.799
2000	0.769	0.829	0.906	0.933	0.740	0.722	0.783	0.837
5000	0.827	0.878	0.935	0.956	0.789	0.794	0.826	0.868

different choices of $k = \{6, 10, 20, 30\}$ and the rows are different sample sizes: $n = \{200, 500, 1000, 2000, 5000\}$.

Table 2 reports three measures of the accuracy of estimation of the recursive order. Block (a) reports the product moment correlation between the estimated recursive order and the true ordering.[8] In Block (b), we report a similarity index, which is the proportion of cases when our method gets the rank of an unit exactly correct; this is a very demanding criterion, particularly for larger k. Block (c) reports the concordance index, which is the proportion of cases, out of all $^kC_2$ combinations, where our method gets the relative ordering of a pair of units correct. From Table 2, it is clear that, as sample size increases, the recovery of the true recursive order improves, in terms of correlation, similarity and concordance.[9] Overall, the Monte Carlo results are promising, and we can now proceed to our application.

---

[8] Spearman's rank correlation is identical because the random variable is integer valued.

[9] We also conducted, but do not report here, Monte Carlo experiments for larger number of units ($k = 40, 45$) and smaller sample sizes ($n = 50, 100$). The results are very similar. The only additional points to make are: (a) the method is simple to apply, and a larger k does not increase computation intensity substantially; and (b) performance is still quite good even for low sample sizes ($n = 50$). These extended results are available in the working paper version together with an alternate identification result based on the covariance matrix, and comparison of Monte Carlo results for the two methods.
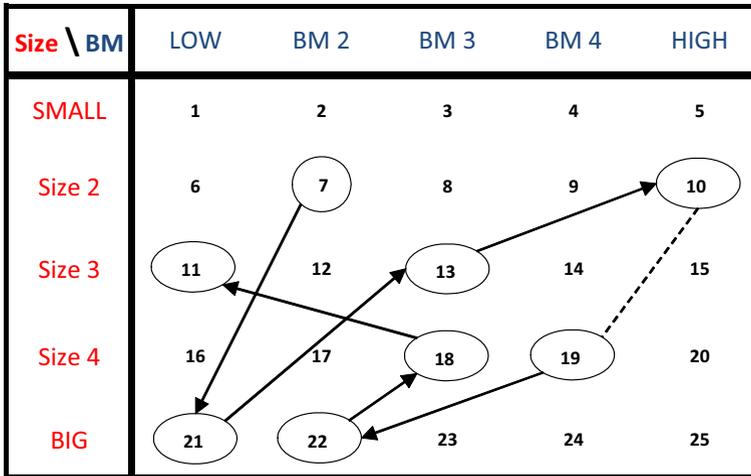
Size \ BM	LOW	BM 2	BM 3	BM 4	HIGH
SMALL	1	2	3	4	5
Size 2	6	7	8	9	10
Size 3	11	12	13	14	15
Size 4	16	17	18	19	20
BIG	21	22	23	24	25

Fig. 1 The 25 Fama–French portfolios, by size (BE) and value (BM, BE/ME)

4 Application

We consider an empirical application to the study of contemporaneous lead–lag relationships in the 25 stock portfolios constructed by Fama and French (1993), the data for which are regularly updated and available from French (2015). Three important considerations guide our choice of application. First, these data have been extensively used by investors and studied in the finance literature, particularly with regard to optimal portfolio construction and inference, and factor models for portfolio returns; see, for example, Davis et al. (2000), Basak et al. (2002), Griffin (2002), and Byun and Rozeff (2003). These Fama/French benchmark portfolios are the intersections of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the ratio of book equity to market equity (BM, that is BE/ME). The size breakpoints for each year are the NYSE market equity quintiles at the end of June of that year. Likewise, the BE/ME breakpoints are NYSE quintiles, where BE/ME for an index year is the book equity for the previous fiscal year end divided by ME for the same period. Monthly returns of the 25 portfolios thus constructed constitute our data; see French (2017) for further details.

Second, we prefer an application with sufficiently many observation units, covered over a substantial period of time, to highlight the use of our methodology as well as to evaluate in-sample and out-of-sample performance of the estimated models. Following Basak et al. (2002), we consider monthly returns for the period July 1963 to December 1991, using data for 300 months (1963m7 to 1988m6) as our estimation sample, and retain 42 months data (1988m7 to 1991m12) for evaluation of the estimated model out-of-sample. The number of portfolios is 25, which implies $k! = 25! > 10^{25}$ possible permutations. The 25 portfolios are graphically summarised in Fig. 1.

Third, much has been written about these data, particularly with regard to factor structure in portfolio returns. For example, Fama and French (1993, 1996) show that

the size (BE) and value (BE/ME) effects subsume the dividend-to-price and earnings-to-price behaviours. In addition, the Fama–French 3-factor model effects (Fama and French 1993) should be important. Hence, if after accounting for factor structure, we can find any evidence of (contemporaneous) lead–lag relationship between the portfolio returns, this can have important implications for theory and practise of finance.

The final point is quite important for interpretation of our results and directions for future research. The standard factor model (Fama and French 1993) posits a strong relation between risk and return, but ignores any possible spillovers of risk between stocks or risk portfolios. Interestingly, contagion risk between banks and markets have been noted in the literature (see, for example, Kodres and Pritsker 2002; Hasman 2013), but not between risk portfolios. However, if there were any such spillovers, this would point to an alternate source of risk propagated not only by the factors, but through spillovers. Then, our assumed recursive structure would help identify portfolios that are influential in propagating risk (the leading portfolios in our estimated ordering) and those that accumulate such risks (the trailing portfolios). In the absence of structural theory, it is difficult to anticipate a priori which portfolios would be leading and which would be trailing. However, we find that the estimated recursive model provides improved prediction of returns, particularly out-of-sample. Then, this indicates evidence that there is substantial propagation of spillover risk, and suggests future research to clarify the nature of such spillovers.

As outlined by our methodology, we conduct our analysis in several steps. First, we account for factor structure in the returns following current best practise from the large panel literature. Specifically, we include two factor variables. First, to account for the heterogenous impact of time-specific factors, we compute cross section average return across the portfolios. This can be interpreted as the return on the market portfolio, but is also an application of the common correlated effects methodology (Pesaran 2006). However, even if the number of time periods here is large ($n = 300$ for the estimation sample), it is not clear that the number of portfolios ($k = 25$) is sufficiently large. Hence, we also take a statistical approach inspired, for example, by Bai (2009), and construct a statistical factor by principal components.[10] For each portfolio, we consider one of the above two as a time-specific factor, depending on whichever has higher correlation with the returns on the portfolio. Second, we also consider a cross section factor, which in our case is simply a portfolio-specific fixed effect. After accounting for the above two factors, residual correlation averaged across the $^{25}C_2 = 300$ distinct combinations is only -0.038, even if the Pesaran CD test statistic (Pesaran 2015) fails the null hypothesis of weak cross section dependence because of large sample sizes. We therefore proceed with our analyses under the assumption of weak cross section dependence.

Second, we build a model for the data in very similar form as Sect. 2,

$$y_{it} = \lambda_i y_{i,t-1} + W_{i.} y_t + \gamma_i' f_t + \epsilon_{it}, i = 1, \ldots, 25, t = 1, \ldots, 300,$$

the only extension being the inclusion of a factor structure $\gamma_i' f_t$ to account for cross section strong dependence. The model is estimated in reduced form and error variances

---

[10] The leading factor accounts for 86% of the total variation in returns across the 25 portfolios.

are estimated. Third, we scale the returns on each portfolio by the estimated standard deviation, and a recursive order is estimated. We do not report the entire sequence of portfolios but point to the leading 4 and final 4 portfolios in Fig. 1; this sequence runs as: $7 \rightarrow 21 \rightarrow 13 \rightarrow 10 \rightarrow \ldots \rightarrow 19 \rightarrow 22 \rightarrow 18 \rightarrow 11$.

A structural interpretation of these portfolios is beyond the scope of this paper. However, we make a few comments. First, the evidence of recursive ordering is very strong, particularly at the beginning and end of the recursive (causal) chain. As emphasised in Theorem 1, recovery of the recursive order is based on relative magnitude of diagonal elements in the precision matrix (inverse of covariance matrix). The ratios of the precision values for the leading two units at each stage can be viewed as an approximate indicator of the strength of recursive order, in the sense that it is related to an F-ratio for comparison of variances. For the 6 comparisons relevant to the 8 portfolios at the ends of the causal chain reported above, the average ratio is 4.55; in fact, for the 4 leading portfolios, this average ratio is even higher at 7.16.[11]

Second, Bhattacharjee (2017) developed a different identification result for recursive order which runs in the opposite direction: from the beginning of the causal chain to the end. We also applied this method to our data and find that the top 4 units in the causal chain come at exactly the same positions using both methods, but the results diverge somewhat in the middle of the network. This provides very strong validation for the methods developed here, particularly since the top of the recursive order is only identified at the end of our procedure. Third, the relatively smaller ratios in the middle of the chain suggest that some elements in the principal subdiagonal may be close to zero. Hence, modelling causal networks using poly-trees may be useful in this context. This is reminiscent of Rebane and Pearl (2013) and Bhattacharjee (2017); further development along these lines is retained for future work.

In the third step of inference, we take the above recursive order as a given and estimate a structural model including lag (VAR) effects and contemporaneous recursive (SVAR) effects. We compare forecasts from the above model (in-sample and out-of-sample) with a standard VAR model. Estimation is standard (Hamilton 1994; Stock and Watson 2001). We do not report the full results, but make some comments.

First, the forecast performance of our recursive model is very good relative to the benchmark VAR. The MSE is reduced in-sample relative to the VAR by about 48% and out-of-sample by 54%. Bias is also substantially reduced, by 94 and 89%, respectively. Importantly, then, predictive performance of the models may also be used as a measure of strength of recursive relationships within a network; see also Lo et al. (2015).

Second, the estimate of the weights matrix W (not reported here) provides exciting details on the nature of the network relationships. The causal effects are positive in many cases, but negative in others. This feature of negative spillovers is also reported in other current research; see, for example, Bhattacharjee and Jensen-Butler (2013) and Bailey et al. (2016). Further, Bhattacharjee and Holly (2013) suggest a measure for the influence of each unit within a network. As expected, portfolios at the top of the network have, in general, the greatest influences. However, there are other portfolios as well. For example, portfolio 3 of small firms has the third highest network influence.

---

[11] An index of the strength of recursive relationship can in principle be developed from this idea. However, statistical theory here may be technically challenging because of dependent data.

Another issue of specific importance for traders is identification of portfolios at the top or the bottom of the recursive order. As expected, small firms (portfolios 1–5) do not occupy these positions. However, it is interesting to note that portfolios in the top-end (south-east in Fig. 1: portfolios 25, 20 and 24) also do not occupy these positions. This is somewhat surprising, since the Fama–French factor model would imply that these portfolios attract the highest return and therefore the highest risk as well. This indicates that transmission of risks between stocks is somewhat more nuanced that what a simple factor model may suggest. More research is necessary to understand such risk spillovers, both in the context of stock and portfolio returns but also more generally in contagion models.

Third, as might be expected, the factor structure is very important, with both time-specific factor (related to market portfolio) and fixed effects (related to firm-specific factors) exerting very strong effects on the returns of all portfolios. However, in the structural model, the persistence parameter is statistically significant only for 3 of the 25 portfolios. In summary, while the application above is mainly illustrative, it provides exciting new evidences on the possibility of better prediction of portfolio returns, better understanding of networks in financial markets, and several pointers towards potential future research.

5 Conclusion

Estimation and inference on spatial weights matrix under various structural constraints has been an active research area within the spatial econometrics literature. We extend this literature by developing a new identification result and inferences under the assumption of recursive ordering. This resulting networks can be very useful for contemporaneous causal inference. Much of the developments in this paper as well as the general area owes itself to previous work on GMM-based methods in spatial econometrics (Kelejian and Prucha 1998, 1999, 2001; Kapoor et al. 2007) and to asymptotic theory for dynamic nonlinear models in the presence of nuisance parameters and dependence (Pötscher and Prucha 1997). Monte Carlo study indicates that finite sample performance of the proposed methods is very good. Applied to data on portfolio returns, our methods offer exciting new evidences on leading and following portfolios and therefore on the flow of information in financial markets.

Several new ideas for future research emerge. Structural models for returns in financial markets may provide exciting new insights and new tools for research and practise. Likewise, new methodological research may focus on inferences under order restrictions, with applications to DAGs and poly-trees, as well as indices of structural ordering.

References

Ahrens A, Bhattacharjee A (2015) Two-step Lasso estimation of the spatial weights matrix. Econometrics 3(1):128–155

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L (2001) Spatial econometrics. In: Baltagi B (ed) A companion to theoretical econometrics. Blackwell, Oxford, pp 310–330

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297

Bai J (2009) Panel data models with interactive fixed effects. Econometrica 77(4):1229–1279

Bailey N, Holly S, Pesaran MH (2016) A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. J Appl Econom 31(1):249–280

Ballester C, Calvó-Armengol A, Zenou Y (2006) Who's who in networks. Wanted: the key player. Econometrica 74:1403–1417

Baltagi BH, Fingleton B, Pirotte A (2014) Estimating and forecasting with a dynamic spatial panel data model. Oxf Bull Econ Stat 76(1):112–138

Basak GK, Jagannathan R, Sun G (2002) A direct test for the mean variance efficiency of a portfolio. J Econ Dyn Control 26(7):1195–1215

Beenstock M, Felsenstein D (2012) Nonparametric estimation of the spatial connectivity matrix using spatial panel data. Geogr Anal 44(4):386–397

Bernanke BS, Boivin J, Eliasz P (2005) Measuring monetary policy: a factor augmented vector autoregressive (FAVAR) approach. Q J Econ 120:387–422

Bessler DA, Lee S (2002) Money and prices: US data 1869–1914 (a study with directed graphs). Empir Econ 27(3):427–446

Bhattacharjee A (2017) Identifying the causal structure of directed acyclic graphs (DAGs). Mimeo. Spatial Economics and Econometrics Centre (SEEC), Heriot-Watt University, UK

Bhattacharjee A, Holly S (2011) Structural interactions in spatial panels. Empir Econ 40(1):69–94

Bhattacharjee A, Holly S (2013) Understanding interactions in social networks and committees. Spatial Econ Anal 8(1):23–53

Bhattacharjee A, Jensen-Butler C (2013) Estimation of the spatial weights matrix under structural constraints. Region Sci Urban Econ 43(4):617–634

Bhattacharjee A, Castro E, Marques J (2012) Spatial interactions in hedonic pricing models: the urban housing market of Aveiro, Portugal. Spatial Econ Anal 7(1):133–167

Bierens HJ (1982) A uniform weak law of large numbers under ϕ-mixing with application to nonlinear least squares estimation. Stat Neerl 36(2):81–86

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Econom 87(1):115–143

Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York

Byun J, Rozeff MS (2003) Long-run performance after stock splits: 1927 to 1996. J Finance 58(3):1063–1085

Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ (1999) Probabilistic networks and expert systems. Springer, Berlin

Davis JL, Fama EF, French KR (2000) Characteristics, covariances, and average returns: 1929 to 1997. J Finance 55(1):389–406

Diebold FX, Yilmaz K (2009) Measuring financial asset return and volatility spillovers, with application to global equity markets. Econ J 119(534):158–171

Domowitz I, White H (1982) Misspecified models with dependent observations. J Econom 20(1):35–58

Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33:119–140

Fama EF, French KR (1993) Common risk factors in the returns of stocks and bonds. J Financ Econ 33:3–56

Fama EF, French KR (1996) Multifactor explanations of asset pricing anomalies. J Finance 51(1):55–84

Fingleton B (2003) Externalities, economic geography and spatial econometrics: conceptual and modeling developments. Int Region Sci Rev 26:197–207

French KR (2017) Detail for 25 portfolios formed on size and book-to-market. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/tw_5_ports.html. Accessed 25 May 2018

Giacomini R, Granger CWJ (2004) Aggregation of space-time processes. J Econom 118:7–26

Griffin JM (2002) Are the Fama and French factors global or country specific? Rev Financ Stud 15(3):783–803

Gupta A, Robinson PM (2015) Inference on higher-order spatial autoregressive models with increasingly many parameters. J Econom 186(1):19–31

Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton

Harary F (1994) Graph theory. Addison-Wesley, Reading, MA

Hasman A (2013) A critical review of contagion risk in banking. J Econ Surv 27(5):978–995

Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. J Econom 140(1):97–130

Kelejian HH, Prucha I (1998) A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Finance Econ 17:99–121

Kelejian HH, Prucha I (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40:509–533

Kelejian HH, Prucha IR (2001) On the asymptotic distribution of the Moran I test statistic with applications. J Econom 104(2):219–257

Klößner S, Wagner S (2014) Exploring all VAR orderings for calculating spillovers? Yes, we can!—a note on Diebold and Yilmaz (2009). J Appl Econom 29(1):172–179

Kodres LE, Pritsker M (2002) A rational expectations model of financial contagion. J Finance 57(2):769–799

Lee L-F (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72:1899–1925

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. Taylor & Francis, New York

Lo A, Chernoff H, Zheng T, Lo SH (2015) Why significant variables aren't automatically good predictors. Proc Nat Acad Sci 112(45):13892–13897

Pearl J (1995) Causal diagrams for empirical research. Biometrika 82:669–710

Pearl J (2009) Causality: models, reasoning and inference, 2nd edn. Cambridge University Press, Cambridge

Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4):967–1012

Pesaran MH (2015) Testing weak cross-sectional dependence in large panels. Econom Rev 34(6–10):1089–1117

Pesaran MH, Tosetti E (2011) Large panels with common factors and spatial correlation. J Econom 161(2):182–202

Pötscher BM, Prucha IR (1997) Dynamic nonlinear econometric models: asymptotic theory. Springer, Berlin

Rebane G, Pearl J (2013) The recovery of causal poly-trees from statistical data. arXiv preprint arXiv:1304.2736

Spiegler R (2016) Bayesian networks and boundedly rational expectations. Q J Econ 131(3):1243–1290

Spirtes P, Glymour C, Scheines R (1993) Causation, prediction, and search. Springer, New York

Stock JH, Watson MW (2001) Vector autoregressions. J Econ Perspect 15:101–115

Walker G (1931) On periodicity in series of related terms. Proc R Soc Lond Ser A 131:518–532

Wall MM (2004) A close look at the spatial structure implied by the CAR and SAR models. J Stat Plan Inference 121:311–324

White H (1980) Nonlinear regression on cross-section data. Econometrica 48(3):721–746

Yule GU (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philos Trans R Soc Lond Ser A 226:267–298