Pattern Recognition 130 (2022) 108789

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

# Time-varying Group Lasso Granger Causality Graph for High Dimensional Dynamic system

Wei Gao*, Haizhong Yang

*School of Statistics, Xi'an University of Finance and Economics, Xi'an, Shannxi, 710010 China*

ARTICLE INFO

*Article history:*
Received 11 May 2020
Revised 13 April 2022
Accepted 11 May 2022
Available online 20 May 2022

*Keywords:*
Time-varying Granger causality
Feature selection
Group Lasso
Financial market network

ABSTRACT

Feature selection is a crucial preprocessing step in data analysis and machine learning. Since causal relationships imply the underlying mechanism of a system, causality-based feature selection methods have gradually attracted great attentions. For a high dimensional system undergoing dynamic transformation, because of the non-stationarity and sample scarcity, modeling the causal structure among these features is difficult. In this paper, we propose a time-varying Granger causal networks to capture the causal relations underlying high dimensional time-varying vector autoregressive models with high order lagged dependence. A kernel reweighted group lasso method is proposed, which overcomes the limitations of sample scarcity and transforms the problem of Granger causal structural learning into a group variable selection problem. The asymptotic consistency of the proposed algorithm is proved. We apply the time-varying Granger causal networks to simulation experiments and real data in the financial market. The study demonstrates that the method provides an efficient tool to detect changes and analysis characters of causal dependency structure in network evolution.



## 1. Introduction

In complex multivariate systems, such as those encountered in biology, finance and neural science, the dependence of variables often contains important information about system features. Dependence-based feature selection has received a significant amount of attention from several disciplines [1–3]. Graphical models are a favored formalism for modeling the dependency structures, where nodes of the network stand for features(variables), and edges correspond to the interactions among them. The feature extraction component includes network estimation and topological feature extraction based on graph analysis. The connection between two variables can be established by considering different connectivity metrics, including Pearson's correlation coefficient, classical linear Granger causality and mutual information, among which Granger causality [4] is particularly useful as a directional measures for feature discovery [5–8] in many fields. When Granger causality interactions are modeled among features or series, the graphical models are extended to Granger causality graphs or causal networks [9]. Network theory provides a new research perspective for the study of intricate linkages [10,11].

With recent advances in sensor technology, many research topics present very high dimensional data. Much work in the statistics and machine learning communities has focused on the questions of the sparse techniques for variable or feature selection, which are crucial to general scientific discovery in the big data age. In sparse feature selection methods, the cardinality of the set of the selected variables is significantly smaller than the entire number of variables [12]. Due to the advantages over other methods in variable selection, many effective sparsity inducing feature selection methods based on Lasso [13], Group Lasso [14], Elastic Net [15] etc., have been proposed, for example the feature selection method based on graph-based feature representations and the Fused Lasso framework [16], the multiple structural interacting elastic net model(MSIEN) [17,18]. And many efficient regularization techniques [13–15] have been extensively studied for Granger causality graphs [19–23]. A briefly review about the state-of-the-art graph-based feature selection methods and regularization-based feature selection methods is given in [18]. The proposed approach is most relevant to the time-varying Ganger causality graph model.

In many real world systems, due the evolution of the variables with time, the system is undergoing dynamic transformation. Time-varying networks are required to capture the dynamic causal influences between variables. There are mainly three categories methods for time-varying causality networks: methods based on

* Corresponding author.
*E-mail addresses:* gaoweistat@163.com (W. Gao), h.yang@163.com (H. Yang).

https://doi.org/10.1016/j.patcog.2022.108789
0031-3203/

time-varying coefficients vector autoregressive(VAR) models [24–26], Bayesian methods [27,28] and Lasso methods [29,30].

During the process of financial globalization, the existing economic systems are more and more interconnected on the basis of interdependence through cross-border lending, investment networks, trade relations, and supply chains. Then the financial market can be considered a complex time-varying system consisting of multiple interacting financial components [31], e.g., the stock trade price and return rate. One way to detect the changes for a system consisting of complex interactions is to represent multiple co-evolving financial time series as a family of time-varying financial network, for example, undirected network based on correlation relations [11,32–34] and causal network based on Granger-causality [10].

However, the methods mentioned above have some drawbacks. Firstly, interaction among multivariate variables is more complex than pairwise variables. Classical Granger causality based on pairwise analysis checks whether lagged information of a variable $Y$ provides any significant statistical information about a variable $X$ in the presence of lagged $X$ without considering the influence of other variables. In high dimensional variables, Granger causality based on pairwise variables may result to spurious causality. Secondly, when the causal dependence is caused by high order lag variables, existing methods based on VAR(1) models will lead to significant information loss. Thirdly, an ordinary Lasso approach does not capture the group structure connections in the VAR model. In the Granger causality test, we are interested just in the effects of the entire past of one series as a group to a given series. This observation suggests that the method considering the group information of variables will lead to more effective variable selection technologies.

In this paper, we propose the kernel reweighted group lasso method and extend the Granger causality graph to a high dimensional time-varying VAR(p) model. The objective is to improve the efficiency of time-varying Granger causality graphical modeling through the group structure of temporal variables, for the high dimension systems with high order lagged dependence. The method relies on the assumption that the underlying network structures are sparse and vary smoothly across time. Furthermore, we apply the proposed method to financial time series, analyse the time-evolving features of relations among companies and individuals. The main contributions of this work are as follows.

- We propose time-varying Granger causality networks based on conditional Granger causality for high dimensional VAR(p) models, considering the direct relations between two variables when the influence of the other variables are given. To the best of our knowledge, the proposed graphical models are the first to consider time-varying multivariate Granger causality with high order lagged dependence.
- We introduce kernel reweighted group Lasso method to estimate the time-varying Granger causality networks. The problem of structural learning is transformed into the variable selection problem. The group structure of the lagged temporal variables is naturally determined by the time series they belong to. Then, we apply group Lasso [14] to promote group-structured sparsity patterns by penalizing the sum of the $l_2$ norms of the predetermined sub-vectors. Thus both the accuracy and efficiency of the algorithm are improved.
- We apply the proposed algorithm to a financial dataset and construct time-varying Granger causal networks. Based on the characteristics of the variables dependency structure, we further analyse the fluctuations of trading network structures in different periods of economic development. Due the directionality of the causality, the networks can also show the out degree and in degree distribution features of the time-varying Granger causality graphs.

The rest of the paper is organized as follows. Section 2 reviews the evolution of Granger causality graphs methods. In Section 3, we propose the time-varying Granger causality graph and an algorithm to recover the graphical model at each time epoch. Section 4 summarizes the results of the numerical experiments and applications of real data. Conclusions and future work are drawn in Section 5.

## 2. Related work

In the research of Granger causality graph models, Granger causality test is an important problem. In many cases, each series is directly Granger-caused by only a small subset of the other series. For analysing increasingly large datasets, traditional VAR-based Granger causality test methods are limited by the intensive computation in parameters estimation [35]. As a subset selection problem, the VAR model is overparameterized for modeling Granger causal relations. One solution is to assume a sparse structure for high-dimensional time series [36] and solve the problem by Lasso methods. Arnold et al. [19] proposed the causality concept based on the Graphical Lasso Granger (GLG) method. Lozano et al. [20] proposed the Graphical group Lasso Granger (GgrLG) method to recover gene causality networks based on multivariate Granger causality concept. Bolstad et al. [21] inferred a sparse causal network modeling by VAR processes. Basu et al. [22] introduced a group lasso regression regularization method to learn the framework of Granger causal models, under the sparsity assumption of the edges and inherent grouping structure of the nodes. Yang et al. [23] proposed a group lasso nonlinear conditional Granger causality (GLasso-NCGC) method to handle the nonlinearity and directionality of complex network systems. Furthermore, neural network methods are applied to detect nonlinear Granger causality. For example, Chivukula et al. [37] introduced a Deep Neural Network (DNN) and a Recurrent Neural Network (RNN) to discover causal relations and causal graphs from Granger causality tests.

Most traditional approaches to Granger causality detection assume linear time series dynamics and treat time as invariant. However, at the heart of the system lies the components exchange mechanism, which establishes a time-evolving network. In many real world systems, there is time-varying dependence between series, and using invariant linear models may lead to inconsistent estimates of Granger causal interactions [25,38].

Therefore, time-varying graphical models have attracted much attention from researchers. Jethava [24] surveyed recent methods for reconstructing time-varying biological networks from a modeling perspective, such as gene interaction networks based on time series node observations (e.g., gene expressions). One class method applies time-varying coefficients VAR models to detect time-varying Granger causality interactions in time series [25,26]. The proposed causal detection framework [25,26] are suitable for causal inference between bivariate time-varying systems. As indicated in [26], sparse representation algorithms like Lasso may improve the efficiency of the time-varying Granger causal analysis to reduce the computation time. Another class of time-varying Granger causality network is based on Bayesian method. Wang et al. [27] treated network structures and parameters in a time-varying dynamic Bayesian network (DBN) as random processes whose values at each time point determine a stationary DBN model, then used this DBN model to specify the distribution of a data sequence at a given time. Iacopini and Rossini [28] provided an innovative Bayesian nonparametric (BNP) time-varying graphical framework for high dimensional time series. Because of the limitation of the number of parameters, the Bayesian graph-

ical models for time-varying parameters VAR in [27,28] only are used in the case of VAR(1) model with several nodes. Based on the sparsity assumption on the network, Lasso method is used to learn time-varying directed graphical models. Song et al. [29] proposed a time-varying dynamic Bayesian network (DBN) to model the structurally varying directed dependency structures underlying non-stationary time series. Guo et al. [30] used the time-varying DBN in [29] to model the time-varying effective connectivity between pairs of ECoG electrodes selected by high gamma power. The kernel reweighted $l_1$ regularized method [29] analyse high dimensional data through time-varying VAR(1) models and still need some improvement, for example considering the structure features of data and other regularized method.

Meanwhile, the study of an evolving financial market network is of great interest. Ye et al. [32] presented a novel method to characterize the evolution of time-varying complex networks by adopting a thermodynamic representation of network structure. Chen et al. [11] focused on volatility spillover effects and considered the issue of how to measure the connectedness of networks among financial firms. Bai et al. [33] used the commute time matrix to develop a new quantum-inspired kernel for dynamic financial network analysis. However, the quantum kernel cannot reflect the most mutually correlated time series specified by the vertices, and will influence the effectiveness [34]. To overcome this problem, Bai et al. [34] developed a new kernel-based approach to measure the similarity between time-varying financial networks and to understand the structural evolution of financial networks over time, and performed the proposed method on time-varying financial networks extracted from New York Stock Exchange (NYSE) dataset.

The research mentioned above is focused on undirected correlation relations among financial time series. Ye et al. [39] used an approximate network von Neumann entropy measure to define a new characterization of network structure to capture the distribution of entropy across the edges of a network, and histogramed the edge entropy using a multi-dimensional array for both undirected and directed networks. For a node, the directed edges is going to its nearest neighbor points, thus the graph is directed. The application of the method [39] to NYSE network is still on undirected and unweighted graphs. Gong et al. [10] constructed the causal complex network of financial institution based on the Granger-causality network and the principal component analysis, and further analyzed the network topology structure characteristics using centrality indicators.

Finally, it worth noting the differences between our method and those closely related to it.

- The time-varying Granger causality networks based on Bayesian method [27,28] and Lasso method [29] are mainly for VAR(1) model. VAR(1) model can only model the relations from $X_{t-1}$ to $X_t$, it is unable to identify the relationships from $X_{t-p}, \cdots, X_{t-2}$, to $X_t$. However, in many cases, there are high order lagged dependence between the individuals. Many traditional methods for VAR(1) models are difficult to estimate the large number of parameters of VAR(p) models ($p \times d^2$ parameters for $d$ variables). In this paper, we extend the time-vary Granger causality graphs from VAR(1) to VAR(p) models.
- Song et al. [29] presented a sparse structure estimation method for a time-varying VAR(1) model based on Lasso [13]. Although the procedure can be easily extended to learn the structure of VAR(p) models, the number of autoregressive coefficients will increase rapidly with the lag order $p$, which will affect the accuracy and efficiency of the models. In this paper, we introduce group-structured sparsity patterns and apply group Lasso [14] to select significant variables. We also carry out experiments on time series models with structure breaks and verify the performance of the methods when the assumption of smoothly varying are violated.
- Most time-varying financial networks are generated from the Pearson correlation [34] metric. The Pearson correlation is a directionless measure without indicating the causal direction. Moreover, The time-varying Granger causality networks proposed in [10] are based on Granger causality relation between pairwise time series. Pairwise correlation of two variables without eliminating the effect of other variables may result in spurious correlation. The time-varying Granger causality graph models proposed in this paper are based on conditional Granger causality, considering the effect of other series in the financial system. As directed relations, the time-varying Granger causality graphs have rich indicators to characterize both the global evolution features of networks and the different functions of individual nodes in the graph.

## 3. Proposed method

We start this section with discussing the theoretical aspects of the time-varying Granger causality graph. Next the structural learning algorithm of the proposed graphical model is provided.

### 3.1. Time-varying Granger causality graph

Granger causality is based on linear regression modeling of stochastic processes and is normally used to check whether one economic variable can help to forecast the other economic variable [4].

#### 3.1.1. Granger causality

Classical Granger causality captures causal relationships between two time series. We first introduce conditional Gragner causality for high dimensional time series.

For a $d$ dimensional ($d > 2$) time series $\{Y_i(t)\}_{i=1}^d, t \in \mathbb{Z}$, series $Y_j$ is said to Granger cause $Y_i$ if the past values of $Y_j$ ($Y_j(s), s < t$) yield a more accurate prediction of $Y_i(t)$ when the past values of other series ($Y_k(s), s < t, k \neq i, j$) are also included in the prediction. One common tool for modeling Granger causality is the VAR model with a lag order $p$:

$$\boldsymbol{Y}(t) = \boldsymbol{A}^1\boldsymbol{Y}(t-1) + \boldsymbol{A}^2\boldsymbol{Y}(t-2) + \cdots + \boldsymbol{A}^p\boldsymbol{Y}(t-p) + \boldsymbol{\epsilon}(t), \quad (1)$$

where $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t), \cdots, Y_d(t))^T, \boldsymbol{A}^k = \{a_{i,j}^k\}, i, j = 1, 2, \cdots, d, k = 1, 2, \cdots, p$ are $d \times d$ autoregressive coefficient matrices, $a_{i,j}^k$ describes the influence of $Y_j(t-k)$ on $Y_i(t)$, and $\boldsymbol{\epsilon}(t) = [\epsilon_1(t), \cdots, \epsilon_d(t)]^T$ is Gaussian white noise, $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_d)$.

Granger causality can be revealed by checking the components of the coefficient matrix $A^k$. Series $Y_j$ Granger-causes $Y_i$ with respect to $\mathscr{F}$ if and only if $\exists\ k \in \{1, 2, ..., p\}$ such that $a_{ij}^k \neq 0$, where $\mathscr{F}$ is the $\sigma$-algebra generated by the linear span of the past and present values of $Y$. The Granger causality is denoted by $Y_j \longrightarrow Y_i\ [\mathscr{F}]$.

#### 3.1.2. Time-varying Granger causality

In many real environments, the relationships among time series may change over time. Model (1) can be extended to the time-varying VAR(p) model,

$$\begin{aligned}\boldsymbol{Y}(t) = {} & \boldsymbol{A}^1(t)\boldsymbol{Y}(t-1) + \boldsymbol{A}^2(t)\boldsymbol{Y}(t-2) + \cdots \\ & + \boldsymbol{A}^p(t)\boldsymbol{Y}(t-p) + \boldsymbol{\epsilon}(t), \end{aligned} \quad (2)$$

where $A^k(t) = \{a_{i,j}^k(t)\}, i, j = 1, 2, \cdots, d, k = 1, 2, \cdots, p$ are $d \times d$ autoregressive coefficient matrices at time $t$, $t \in \mathbb{Z}$ and $\boldsymbol{Y}(t)$, $\boldsymbol{\epsilon}(t)$ are the same as in model (1).

Similar to Granger causality, time-varying Granger causality can also be revealed by checking the components of the coefficient matrix $A^k(t)$. At time $t^* \in \mathbb{Z}$, series $Y_j$ Granger-causes $Y_i$ with respect

to $\mathscr{F}_{t^*}$ if and only if $\exists\ k \in \{1, 2, ..., p\}$ such that $a_{ij}^k(t^*) \neq 0$, where $\mathscr{F}_{t^*}$ is the $\sigma$-algebra generated by the linear span of the past and present values of $Y$ at time $t^*$. The Granger causality is denoted by $Y_j \longrightarrow Y_i \ [\mathscr{F}_{t^*}]$.

*3.1.3. Time-varying Granger causality graph*

Given $d$ variables, a graphical model or network is generated by calculating a certain kind of connectivity metric for every pair of variables. Granger Causality metric generates directed networks that have asymmetric adjacency matrices of size $d \times d$. Combining time-varying Granger causality with graphical models, we generate time-varying Granger causality graphs as follows.

Let $\{Y_i(t)\}_{i=1}^d, t \in \mathbb{Z}$ be a process generated by the time-varying VAR(p) model (2). Then $G(t) = (V, E(t)), t \in \mathbb{Z}$ is the time-varying Granger causality graph, where the vertex set $V = \{1, 2, \cdots, d\}$ denotes the component series $Y_i, i \in \{1, 2, \cdots, d\}$, and edges set $E(t)$ satisfies the condition that for all $j \in V, j \neq i$,

$$j \longrightarrow i \in E(t) \iff Y_j \longrightarrow Y_i \ [\mathscr{F}_t], \qquad t \in \mathbb{Z}. \tag{3}$$

### 3.2. Structural learning of time-varying Granger causality graph

In this section, we consider how to recover the time-varying Granger causality relations(directed edges) for a single series(node) at one time point(without loss of generality, take node $i$ at time $t^*$, for example), and then join the Granger causal relations of the separate nodes to achieve the desired result.

*3.2.1. Problem statement*

For establishing the time-varying Granger causality graphs, as vertex set $V$ is known and fixed, the problem is to determine the edge set $E(t)$. In many networks, each node is directly influenced by only a small subset of other nodes. The VAR model (2) is over-parameterized in such cases. Following Bolstad et al. [21], we define the sparse time-varying VAR models which have the form as model (2), but include an extra parameter $S_i(t)$ to eliminate over-parameterization.

Let $S_i(t)$ be the subset of nodes that causally influence node $i$ at time $t$, i.e., $S_i(t) = \{j \in \{1, \cdots, d\} : \exists\ 1 \leq k \leq p, a_{ij}^k(t) \neq 0\}$. That is, at time $t$, if node $j$ influences node $i$, then $j \in S_i(t)$, otherwise $j \notin S_i(t)$ and $a_{i,j}^k(t) = 0$ for all time indices $k$. The sparse time-varying VAR($p$) model for node $i$ is given by

$$Y_i(t) = \sum_{j:j\in S_i(t)} \sum_{k=1}^{p} a_{i,j}^k(t) Y_j(t-k) + \epsilon_i(t), i = 1, 2, \cdots, d. \tag{4}$$

We are primarily interested in the networks, $\min_t |\cup_{i=1}^d S_i(t)| \leq md$, for some constant $m > 1$. In such cases, the main challenge to accurate inferences is to reliably identify the set $S_i(t)$.

Identifying $S_i(t)$ is a subset selection problem. In a Granger causality graph, the question we are interested in is whether the entire past of series $Y_j$ affects the current value of $Y_i$. Clearly, as a structural learning method, the relevant variable selection question is not whether each lagged variable $Y_j(t-k), k = 1, 2, \cdots, p$ should be included in the regression, but whether the lagged variables for a given time series $(Y_j(t-1), Y_j(t-2), \cdots, Y_j(t-p))$ as a group should be included. Thus, in the variable selection process, a more faithful implementation is to incorporate the group structure imposed by time series into the modeling methods.

The temporal variables $Y_j(t-k), j \in S_i(t), k = 1, 2, \cdots, p$ in Eq. (4) have a naturale group structure that is imposed by the time series they belong to. Let $\boldsymbol{X}_j(t) = \left(Y_j(t-1), \cdots, Y_j(t-p)\right) \in \mathbb{R}^{1\times p}$ be a vector consisting of the lag variables from series $Y_j, j = 1, \cdots, d$. Then, in Eq. (4), the lag variable set $\left\{Y_j(t-k), j = 1, 2, \cdots, d, k = 1, 2, \cdots, p\right\} \in \mathbb{R}^{1\times dp}$ has a group structure with $d$ groups $(\boldsymbol{X}_1(t), \cdots, \boldsymbol{X}_d(t))$.

The estimation problem in Eq. (4) is reformulated as a regressive problem,

$$Y_i(t) = \boldsymbol{X}(t)\boldsymbol{\alpha}_i(t) + \epsilon_i(t), \tag{5}$$

where $\boldsymbol{X}(t) = (\boldsymbol{X}_1(t), \boldsymbol{X}_2(t), \cdots, \boldsymbol{X}_d(t))$, $\boldsymbol{\alpha}_i(t) = \begin{pmatrix} \boldsymbol{\alpha}_i(t)^1 \\ \boldsymbol{\alpha}_i(t)^2 \\ \vdots \\ \boldsymbol{\alpha}_i(t)^d \end{pmatrix}$, $\boldsymbol{\alpha}_i(t)^j = \left(a_{ij}^1(t), a_{ij}^2(t), \cdots, a_{ij}^p(t)\right)^T$.

*3.2.2. Kernel weighted group Lasso method*

The next problem is to estimate parameter vector $\boldsymbol{\alpha}_i(t)$ in Eq. (4). The main challenge in the inference of time-varying Granger causality is the unavailability of multiple observations at every instant $t$. Usually, one or at most a few observations are available for each instant. To make the problem statistically tractable, Song et al. [29] relied on the assumption that the underlying network structures are sparse and vary smoothly over time. If coefficient matrices $\boldsymbol{A}^k(t), k = 1, 2, \cdots, p$ vary smoothly over time, then the temporally adjacent Granger causality graphs are more likely to share common edges than temporally distant graphs. Song et al. [29] proposed a kernel reweighted $l_1$-regularized autoregressive procedure to estimate the parameters of time-varying $VAR(1)$ models. In this paper, we combine the kernel reweighted procedure and group Lasso to determine the Granger causality of a time-varying $VAR(p)$ model.

At time $t^*$, we reweight the observations $\boldsymbol{X}(t)$ according to their distance from $t^*$. The weight of an observation at time $t$, $\omega^{t^*}(t)$ is defined as follows,

$$\omega^{t^*}(t) = \frac{K_h(t - t^*)}{\sum_{t=1}^n K_h(t - t^*)} \tag{6}$$

where $K_h(\cdot) = K(\frac{\cdot}{h})$ is a symmetric nonnegative kernel function and $h$ is the kernel bandwidth. In our simulation, we use a Gaussian kernel function,

$$K_h(t) = \exp\left(-\frac{t^2}{h}\right)$$

.

Then the problem Eq. (5) is transformed into estimating the regression coefficients of model,

$$\boldsymbol{W}^{\frac{1}{2}}(t^*)\boldsymbol{y}_i = \boldsymbol{W}^{\frac{1}{2}}(t^*)\boldsymbol{X}\boldsymbol{\alpha}_i(t^*) + \boldsymbol{W}^{\frac{1}{2}}(t^*)\boldsymbol{\epsilon}_i, \tag{7}$$

where

$$\boldsymbol{y}_i = \begin{pmatrix} Y_i(p+1) \\ Y_i(p+2) \\ \vdots \\ Y_i(n) \end{pmatrix},$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1(p+1) & \boldsymbol{X}_2(p+1) & \cdots & \boldsymbol{X}_d(p+1) \\ \boldsymbol{X}_1(p+2) & \boldsymbol{X}_2(p+2) & \cdots & \boldsymbol{X}_d(p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_1(n) & \boldsymbol{X}_2(n) & \cdots & \boldsymbol{X}_d(n) \end{pmatrix},$$

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_i(p+1) \\ \epsilon_i(p+2) \\ \vdots \\ \epsilon_i(n) \end{pmatrix},$$

and the weight matrix is a $(n-p)\times(n-p)$ diagonal matrices.

$$\boldsymbol{W}^{\frac{1}{2}}(t^*)=\begin{pmatrix}\sqrt{\omega^{t^*}(p+1)} & 0 & \cdots & 0\\ 0 & \sqrt{\omega^{t^*}(p+2)} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sqrt{\omega^{t^*}(n)}\end{pmatrix}.$$

Simple subset selection problems can be solved using the well-known Lasso procedure [13]. In many situations, natural groupings exist among variables, and variables belonging to the same group should be either selected or eliminated together. Yuan and Lin [14] used the group Lasso method to deal with the question of group selection. The group Lasso method divides a coefficient vector into predetermined sub-vectors and penalizes the sum of the $\ell_2$ norms of the sub-vectors, whereas the Lasso method penalizes the $\ell_1$ norm of each coefficient. Such a penalty is beneficial when each group of coefficients is believed to be either all zero or all non-zero, and the solution contains only a small number of nonzero coefficient groups.

Based on the group structure of the predictor matrix $\boldsymbol{X}$, the group Lasso estimate of Eq. (7) solves the optimal problem as follows.

$$\hat{\boldsymbol{\alpha}}_i(t^*)=\min_{\boldsymbol{\alpha}_i(t^*)}\frac{1}{n}\|\boldsymbol{W}^{\frac{1}{2}}(t^*)\boldsymbol{y}_i-\boldsymbol{W}^{\frac{1}{2}}(t^*)\boldsymbol{X}\boldsymbol{\alpha}_i(t^*)\|^2+\lambda\sum_{j=1}^{d}\|\boldsymbol{\alpha}_i^j(t^*)\|_2 \tag{8}$$

where $\boldsymbol{\alpha}_i^j(t^*)=\left(a_{ij}^1(t^*),\cdots,a_{ij}^p(t^*)\right)$ is the coefficient vector corresponding to the $j$-th group vector $\boldsymbol{X}_j(t^*)=\left(Y_j(t^*-1),\cdots,Y_j(t^*-p)\right)$. By adopting $\|\boldsymbol{\alpha}_i^j(t^*)\|_2$ as the intra-group penalty, the group Lasso method encourages the coefficients for the variables of a given group $\boldsymbol{X}_j(t)$ to be either all zero($Y_j$ is not Granger causality of $Y_i$ at time $t^*$) or all non-zero($Y_j$ is Granger causality of $Y_i$ at time $t^*$), and the solution contains only a small number of non-zero coefficient groups.

We follow the sub-gradient approach proposed by Liu and Ye [40] to solve optimal problem Eq. (8). An efficient implementation of the algorithm in MATLAB interfaced module called SLEP is also supplied in Liu et al. [41]. The procedure to estimate a time-varying Granger causality graph is summarized in Algorithm 1.

**Algorithm 1** Procedure for estimating a time-varying Granger causality graph.

**Require:** Time series data $\{Y_i(t)\}_{t=1}^n, i=1,2,\cdots,d$, regularization parameter $\lambda$ and kernel parameter $h$.
  **for** $t^*=p+1,\cdots,n$ **do**
    **for** $i=1,\cdots,d$ **do**
      Compute $\boldsymbol{W}(t^*),\boldsymbol{y}_i,\boldsymbol{X}$
      Estimate $\hat{\boldsymbol{\alpha}}_i(t^*)^T$ for Eq.(8)
      **for** $j=1,\cdots,d$ **do**
        **if** $\hat{\boldsymbol{\alpha}}_i^j(t^*)\neq\boldsymbol{0}$, $j\neq i$ **then**
          Place an edge $j\to i$ into $E(t^*)$
        **end if**
      **end for**
    **end for**
  **end for**
**Ensure:** Time-varying Granger causality graph $G(t^*)=(V,E(t^*)), V=\{1,2,\cdots,d\}, t^*=p+1,\cdots,n$

**Definition 1. (Time-varying group Lasso Granger causal graphs,TV-gL-GCG)** Let $\{Y_i(t)\}_{i=1}^d, t\in\mathbb{Z}$ be a process generated by the time-varying VAR(p) Model (2). Then, the time-varying group Lasso Granger causality graphs of $\{Y_i(t)\}_{i=1}^d, t\in\mathbb{Z}$ are the graphs $\hat{G}(t)=(V,\hat{E}(t))$ with vertex set $V=\{1,\cdots,d\}$ and edges set $E(t)$, which satisfies that, for all $j\in V, j\neq i$,

$$j\longrightarrow i\in\hat{E}(t)\iff\hat{\boldsymbol{\alpha}}_i^j(t)\neq\boldsymbol{0},\quad t\in\mathbb{Z} \tag{9}$$

where $\hat{\boldsymbol{\alpha}}_i^j(t)$ are the kernel reweighted group Lasso estimations from Eq. (8).

### 3.2.3. Consistency of the algorithm

In this subsection, we study the statistical consistency of the estimation (Eq. (8)) in Section 3.2.2. Song et al. [29] founded asymptotic consistency of the Lasso estimation for TV-VAR(1) model. Bolstad et al. [21] derived conditions under which the group Lasso procedure can consistently estimate time invariant sparse network structure. Our structural consistency result for TV-gL-GCG estimation procedure combines the proof strategy of [21] for the group Lasso and [29] for time-varying Lasso. In the following, we state our assumptions and theorem.

Define $\boldsymbol{X}_{S_i(t)}$ and $\boldsymbol{X}_{\overline{S_i(t)}}$ as submatrices of $\boldsymbol{X}$ composed of the matrices $\boldsymbol{X}_j, j\in S_i(t)$ and $\boldsymbol{X}_j, j\notin S_i(t)$, respectively. The covariance matrix $\boldsymbol{R}_{S_i(t),S_i(t)}=E\left[\boldsymbol{X}_{S_i(t)}^T\boldsymbol{X}_{S_i(t)}\right]$.

**A1**. $d=\mathrm{O}(n^{c_1})$, $m=\mathrm{O}(n^{c_2})$, $p=\mathrm{O}(n^{c_3})$, and $\lambda=\Theta((n^{-c_4/2})$, with $c_i>0, i=1,2,3,4, c_2<c_4, c_3+c_4<1$; $m\lambda^2=o(1)$ and $\frac{p}{n\lambda^2}=o(1)$.

**A2**. The elements of the matrix $A^k(t), k=1,\cdots,p,\ t\in\mathbb{Z}$ are smooth functions with bounded second derivatives, i.e., there exists a constant $L>0$ such that

$$|\frac{\partial}{\partial t}a_{ij}^k(t)|<L\quad and\quad |\frac{\partial^2}{\partial t^2}a_{ij}^k(t)|<L. \tag{10}$$

**A3**. The minimum absolute value of the nonzero element of the matrix $A^k(t)$ is bounded away from zero at observation points, and this bound tends to zero as we observe more and more samples, i.e.,

$$\alpha_{min}=\min_{t\in\{1,2,\cdots,n\}}\min_{i\in\{1,\cdots,d\},j\in S_i(t)}\|\boldsymbol{\alpha}_i^j(t)\|_2\geq C_{con}>0. \tag{11}$$

**A4**. Let $\boldsymbol{\Sigma}(t^*)=\boldsymbol{W}(t^*)\mathbb{E}[\boldsymbol{X}^T\boldsymbol{X}]\in\mathbb{R}^{pd\times pd}$ be the joint covariance matrix and be invertible,

$$\max_{i\in\{1,\cdots,d\}}\sigma_i^2(t^*)\leq C_{power}<\infty, t\in\{1,2,\cdots,n\}.$$

**A5**.

$$\max_i\left\|\boldsymbol{\Sigma}(t^*)^{-1}_{S_i(t^*),S_i(t^*)}\right\|_2\leq C_{min}^{-1}<\infty,\qquad i=1,2,\cdots,d,t^*\in\mathbb{Z},$$

$$\max_i\|\boldsymbol{\Sigma}(t^*)_{S_i(t^*),S_i^c(t^*)}\|_2\leq C_{max}<\infty,\qquad i=1,2,\cdots,d,t^*\in\mathbb{Z},$$

where $\boldsymbol{\Sigma}(t^*)_{S_i(t^*),S_i(t^*)}$, $\boldsymbol{\Sigma}(t^*)_{S_i(t^*),\overline{S_i(t)}}$ are the corresponding subsets of $\boldsymbol{\Sigma}(t^*)$.

**A6**. For $j\in\overline{S_i(t^*)}, i=1,2,\cdots,d,t^*\in\mathbb{Z}$,

$$\psi_{j\to i}^{FC}(t)=\left\|\sum_{k\in S_i(t)}\Psi_{j,k}^T(t)\frac{\boldsymbol{\alpha}_i^k(t)}{\|\boldsymbol{\alpha}_i^k(t)\|}\right\|_2\leq C_{fcs}<1,$$

where $\Psi_{j,S_i(t)}=\boldsymbol{\Sigma}(t^*)^{-1}_{S_i(t^*),S_i(t^*)}\boldsymbol{\Sigma}(t^*)_{S_i(t),j}$.

**A7**. The kernel $K(\cdot):\mathbb{R}\mapsto\mathbb{R}$ is a symmetric function and has bounded support on [0,1]. There exists a constant $M_K$ that upper bounds the following quantities: $\max_{x\in\mathbb{R}}|K(x)|$ and $\max_{x\in\mathbb{R}}K(x)^2$.

Assumption A1 specifies how the parameters of the time-varying models grow as a function of the number of observations $n$. Assumption A2 requires the coefficients to vary smoothly over time and have finite first and second derivatives. Assumptions A3-A5 show that there will be no false negatives for sufficiently small $\lambda$. Assumption A5 ensures that each observation in $S_i(t^*)$ contains some independent information and that any influence due to the

nodes in $S_i(t^*)$ cannot be easily generated by the nodes in $\overline{S_i(t^*)}$. Assumption A6 is used to show that the probability of declaring a nonzero connection when none exists goes to zero for large $n$. Assumption A7 ensures that the weighted observations satisfy Assumptions A3-A6.

**Theorem 1.** *Assume that A1-A7 hold. If* $\lambda = O(\sqrt{(\log p)/nh})$, $\alpha_{min} \geq 2\lambda$, $h = O(n^{1/3})$, *and* $\frac{s\log d}{nh} = o(1)$, *then*

$$\mathbb{P}\Big[\cup_{i=1}^{d}\hat{S}_i(t^*) = \cup_{i=1}^{d}S_i(t^*)\Big] \to 1.n \to \infty, \quad \forall t^* \in \{1, \cdots, n\}, \tag{12}$$

Problem (8) is a non-differentiable convex optimization problem, for which the classical tools of convex optimization lead to the following optimality conditions: KKT(Karush-Kuhn-Tucker) conditions (details of the proof are in Yuan and Lin [14]).

**Lemma 1.** *A vector* $\boldsymbol{\alpha}_i(t^*) \in \mathbb{R}^{pd}$ *with sparsity pattern* $S_i(t^*) = \{j, \boldsymbol{\alpha}_i^j(t^*) \neq \mathbf{0}\}$ *is optimal for problem (8) if and only if*

$$\boldsymbol{X}_j^T\boldsymbol{W}(t^*)\big(\boldsymbol{y}_i - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_i(t^*)\big) = \frac{\lambda n\hat{\boldsymbol{\alpha}}_i^j(t^*)}{\parallel \hat{\boldsymbol{\alpha}}_i^j(t^*) \parallel}, \qquad \forall j \;\; s.t.\hat{\boldsymbol{\alpha}}_i^j(t^*) \neq \mathbf{0}, \tag{13}$$

$$\Big\|\boldsymbol{X}_j^T\boldsymbol{W}(t^*)\big(\boldsymbol{y}_i - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_i(t^*)\big)\Big\| \leq \lambda n, \qquad \forall j \;\; s.t.\hat{\boldsymbol{\alpha}}_i^j(t^*) = \mathbf{0}, \tag{14}$$

In the following, we state a sketch of the proof, but leave the detailed proof in the supplementary material.

**Proof.** Firstly, we follow the proof strategy of Bostad et al. [21] for the group lasso estimation of Granger causality graphs and Song et al. [29] for kernel reweighted lasso estimation of time-varying Granger causality graphs, combining with the KKT condition (13). Results

$$\mathbb{P}[S_i(t^*) \subseteq \hat{S}_i(t^*)] \to 1, \quad n \to \infty,$$
$$\text{and} \quad \mathbb{P}[\hat{S}_i(t^*) \subseteq S_i(t^*)] \to 1, \quad n \to \infty$$

are proved respectively.

Eq. (8) recovers the correct parents of node $i$ (the set $S_i(t^*)$) with a probability converging to 1 as $n \to \infty$, that is

$$\mathbb{P}[S_i(t^*) = \hat{S}_i(t^*)] \to 1, \quad n \to \infty.$$

Then, for the union bound,

$$\mathbb{P}\Big[\cup_{i=1}^{d}\hat{S}_i(t^*) = \cup_{i=1}^{d}S_i(t^*)\Big] \leq \sum_{i=1}^{d}\mathbb{P}\Big[\hat{S}_i(t^*) = S_i(t^*)\Big]$$
$$\to 1, \quad \forall t^* \in \{1, \cdots, n\}, \tag{15}$$

which completes the proof of Theorem 1. □

In our present context, this theorem can be directly used to derive a proposition on the consistency of a time-varying group Lasso Granger causality graphs.

**Proposition 1.** *Suppose the true time-varying Granger causality graphs be* $G(t), t \in \mathbb{Z}$, *associated stochastic model* $\{Y_i(t)\}_{i=1}^d, t \in \mathbb{Z}$ *(defined in Model (2)). If the assumptions in Theorem 1 are fulfilled, then from the time series observations, Algorithm 1 will output time-varying group Lasso Granger causal graphs* $\hat{G}(t)$ *that are consistent with the true graphs* $G(t), t \in \mathbb{Z}$ *with a probability converging to 1 as n tends to* $\infty$.

## 4. Experiments results

In this section, we conduct experiments to verify the effectiveness of the proposed method. Our test data come from both simulation and financial time series. Section 4.1 demonstrates the performance of the structural learning algorithm for TV-VAR(p) models. The methods for constructing time-varying Granger causality graphs reported in references [27–29] are used mainly for VAR(1) model, and the approaches proposed in [27,28] are only for 2 and 4 dimensional variables. The kernel reweighted group Lasso method proposed in this paper is successfully used for high dimension $VAR(p)$, $p \geq 2$ models. Thus we mainly compare our method with kernel weighted Lasso [29], group Lasso with invariant structure [22], sliding window group Lasso and sliding window correlation-based network [39]. We also apply the proposed kernel reweighted group Lasso method to time series models with structure breaks and compare with the group Lasso method based on true change points. The experiments show that the proposed method performs well when the assumption of smoothly varying is violated.

Section 4.2 applies the proposed algorithm to financial network which is extracted from New York Stock Exchange dataset and further analyse the fluctuations in trading network structures caused by global political or economic events, which is tested in [32,34]. We also show the out degree and in degree distribution features of the time-varying Granger causality graphs and analyze the evolution of the cause and effect relations between stocks. Gong el at. [10] constructed causal networks of China's financial market of 24 companies and analysed the network topology structure characteristics. The networks are based on pairwise causality between two companies without considering the effect of the other companies. In addition, there are only one crisis period (from September 26, 2007 to December 31, 2009) during the sample period(from September 26, 2007 to December 28, 2017). The New York Stock Exchange dataset in this paper has several different financial crises [32,34]. And time-varying group Lasso Granger causality graphs method can not only better recover the underlying networks, but also provide information of financial contagion mechanism for the underlying financial dataset.

### 4.1. Synthetic data

In this subsection, the performance of the proposed kernel weighted group Lasso method will be evaluated under different simulation scenarios, smoothly varying and piecewise stationary model structures respectively. In both scenarios, 100 data sets are randomly generated and all time series have mean zero. For the synthetic data, we evaluate the method using the so-called F1 score, which is the harmonic mean of precision and recall,i.e.,

$$F1 := \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{16}$$

where the average score pair precision and recall are defined as,

$$Precision = \frac{\#\{\ true\ \ positives\}}{\#\{\ edges\ \ identified\}}, \qquad Recall = \frac{\#\{\ true\ \ positives\}}{\#\{\ true\ \ edges\}}.$$

The F1 score is a natural choice of performance measure, as it tries to balance precision and recall. F1 is high only when both precision and recall are high.

#### 4.1.1. Case 1: Smoothly varying structure

In this experiment, we generate several synthetic time series using second-order autoregressive models with smoothly varying model structures. To demonstrate the advantage of the proposed approach in modeling time-varying Granger causality structures, we compare our structure learning methods based on kernel weighted group Lasso methods with methods based on kernel weighted Lasso [29], group Lasso with invariant structure [22], sliding window group Lasso and sliding window correlation-based network [39].

The procedures for generating synthetic data are similar to those in Song et al. [29], but have a lag order $p > 1$. The synthetic time series are generated according to a time-varying VAR(2) model as follows,

$$\boldsymbol{Y}(t) = \boldsymbol{A}^1(t)\boldsymbol{Y}(t-1) + \boldsymbol{A}^2(t)\boldsymbol{Y}(t-2) + \boldsymbol{\epsilon}(t), \tag{17}$$

**Table 1**
F1 scores of the estimated time-varying Granger causality graphs of Model (17).

| Methods | $d$=10 | | | $d$=20 | | | $d$=50 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n=350 | n=1050 | n=1750 | n=350 | n=1050 | n=1750 | n=350 | n=1050 | n=1750 |
| KWgL-VAR(2) | **0.6209** | **0.6972** | **0.7514** | **0.5094** | **0.6225** | **0.6740** | **0.4099** | **0.5644** | **0.6154** |
| KWL-VAR(2) | 0.5953 | 0.6477 | 0.6798 | 0.4533 | 0.5605 | 0.6001 | 0.3507 | 0.4829 | 0.5306 |
| SWgL-VAR(2) | 0.5737 | 0.6295 | 0.6537 | 0.4314 | 0.5204 | 0.6065 | 0.3147 | 0.4794 | 0.5595 |
| VAR(2) [22] | 0.4989 | 0.5136 | 0.5206 | 0.3425 | 0.3863 | 0.4075 | 0.2052 | 0.2912 | 0.3314 |
| KWL-VAR(1) [29] | 0.4692 | 0.4756 | 0.5225 | 0.3289 | 0.3529 | 0.3745 | 0.2341 | 0.3011 | 0.3208 |
| SWcorr [39] | 0.4528 | 0.4751 | 0.4771 | 0.2955 | 0.2984 | 0.3000 | 0.1395 | 0.1408 | 0.1412 |

with noise $\epsilon(t) \sim \mathcal{N}(0, \boldsymbol{I}_d)$.

We first generate the coefficient matrices $A^k(t), k = 1, 2, t = 1, 2, \cdots, n$, as follows. At first, we randomly select $2d$ elements from each pair of matrices $A^k(t_i)$, $k = 1, 2, i \in \{1, 2, \cdots, 8\}$ and assign values drawn from uniform distribution $U[-1, 1]$. The eight pairwise different anchor transition matrices obtained through above mentioned method each corresponds to a random graph of node size $d$ with an average degree of $2d$ are generated. Next we evenly space these eight pairwise anchor matrices as follows: $t_1 = 1, t_2 = \frac{n}{7}, t_3 = \frac{2n}{7}, \cdots, t_8 = n$. The intermediate matrices are interpolated to match the number of observations $n$ as follows,

$$a_{ij}^k(t) = a_{ij}^k(t_{m-1}) + \frac{a_{ij}^k(t_m) - a_{ij}^k(t_{m-1})}{t_m - t_{m-1}}(t - t_{m-1}), \quad t_{m-1} < t < t_m, \tag{18}$$

where $m = 2, 3, \cdots, 8, k = 1, 2, i, j = 1, 2, \cdots, d$.

Then, with the sequence of time-varying coefficient matrices of $\{A^1(t), A^2(t)\}(t = 1, 2, \cdots, n)$, we simulate the time series according to the time-varying *VAR* model in Eq. (17).

We generate $n = 350, 1050$, and 1750 observations for different dimensions $d = 10, 20$ and 50 according Eq. (17), and recover the underlying time-varying Granger causality graphs using the kernel reweighted group Lasso method in Algorithm 1. In the simulation, let the lag order $p = 2$ and the bandwidth parameter of the Gaussian kernel $h = \frac{n^2}{49}$. The regularization parameters $\lambda$ is the one to ensure optimal F1 scores. For each simulation, we obtain the $n - p$ coefficient matrices $\hat{\boldsymbol{A}}^{(k)}(t), k = 1, 2, t = p + 1, \cdots, n$, Granger causality graphs $\hat{G}(t), t = p + 1, \cdots, n$ and their $F1$ scores respectively. The $F1$ score of the simulation is the mean values of the $n - p$ $F1$ scores. The experiment is repeated 100 times. The estimation results are summarized in Table 1.

In Table 1, the results are compared with the performance of methods based on kernel weighted Lasso [29], group Lasso with invariant structure[22], sliding window group Lasso and sliding window correlation-based network[39]. In this section, we use the abbreviations a) KWgL-VAR(2) to denote kernel reweighted group Lasso method for VAR(2) model, b) KWL-VAR(2) to denote kernel reweighted Lasso method for VAR(2) model, c) SWgL-VAR(2) to denote sliding window group Lasso, d)VAR(2) to denote group Lasso with invariant structure [22], e)KWL-VAR(1) to denote kernel reweighted Lasso method for VAR(1) model [29], f)SWcorr to denote sliding window correlations [39]. The main conclusions are as follows.

Table 1 shows that for the synthetic data generated from Eq. (17), $F1$ scores of the KWgL method rapidly improves as the number of observations increases. This is consistent with the consistency of the estimation method in Section 3.2.3. Comparing with the other methods, our method (KWgL-VAR(2)) achieves the highest $F1$ values on all scenarios. The advantages become obvious with the increase of dimension and sample size.

The KWgL-VAR(2) achieves higher $F1$ scores than KWL-VAR(2) and SWgL-VAR(2). The results verify that for time series models with high-order dependence and smoothly varying structure, group Lasso based method offers distinct advantages over Lasso based method, and kernel weighted method performs better than sliding window methods. Method based on VAR(1) model can not reflect the relations due to high order lag variables. Then the $F1$ scores of KWL-VAR(1) are worse than these of KWL-VAR(2). The results of invariant structure VAR(2) based on group Lasso [22] are worse than those methods based on time-varying coefficients. These results verify that the time-varying structure contains information which can not be captured by the invariant structure models. SWcorr shows the lowest $F1$ values on all scenarios and changes not obviously with the increase of sample size. These results indicate that the correlation based network can not show the Granger causality correctly.

In general, the results in Table 1 further verify the advantages of the proposed kernel reweighted group Lasso method for high dimension time series with high order lagged and time-varying dependence.

#### 4.1.2. Case 2: Structural breaks

The proposed kernel reweighted group Lasso algorithm relies on the assumption that the underlying network structures are varying smoothly across time. In the follows, we carry out experiments on time series models with structural breaks and verify the performance of the methods when the assumption about smoothly varying are violated.

Consider a piecewise stationary VAR(2) model

$$\boldsymbol{Y}(t) = \sum_{j=1}^{M+1} \left[A^1(t_j)\boldsymbol{Y}(t-1) + A^2(t_j)\boldsymbol{Y}(t-2)\right] I(t_{j-1} \le t < t_j) + \boldsymbol{\epsilon}(t), \tag{19}$$

with noise $\epsilon(t) \sim \mathcal{N}(0, \boldsymbol{I}_d)$. The time instincts $\{t_1, \cdots, t_M\}$ with $1 = t_0 < t_1 < \cdots < t_{M+1} = n + 1$ denote the change points when the parameters $A^k(t_j)$ changes to $A^k(t_{j+1}), k = 1, 2$ at time $t_j$.

The coefficient matrices $A^k(t_1), k = 1, 2$ are generated as the same as section 4.1.1. For $A^k(t_j), k = 1, 2, j = 2, \cdots, M + 1$, we randomly select nonzero coefficients in $A^k(t_{j-1}), k = 1, 2$ and make them zero(delete edges), or randomly select zero coefficients and make them nonzero(add edges), to generate $A^k(t_j), k = 1, 2$. In the experiments, we consider different change proportion(that is the percentage of changed edges in total edges at each change point), 10%, 20%, 50%.

We consider the different number of change points, $M = 1$ and $M = 7$. The $M$ change points are evenly distributed in $n$ observations. With dimension $d = 50$, we generate $n = 200, 1000$, and 2000 observations according Eq. (19), and recover the underlying time-varying Granger causality graphs using the kernel reweighted group Lasso method in Algorithm 1.

Table 2 shows the average $F1$ scores of 100 simulation experiments. We also estimate the Granger causality networks by group Lasso at each segmentation through true change points. The method is denoted as abbreviation "True $t_j$s gL-*VAR*(2)" in Table 2.

In Table 2, when $M = 1$, the performance of the two methods is similar. The $F1$ scores of our method(denoted as KWgL-

**Table 2**
F1 scores of the estimated time-varying Granger causality graphs of Model (19), $d = 50$.

| Change | Sample | $M = 1$ | | $M = 7$ | |
|---|---|---|---|---|---|
| percentage | Size | KWgL-VAR(2) | True $t_j$s gL-VAR(2) | KWgL-VAR(2) | True $t_j$s gL-VAR(2) |
| 10% | $n = 200$ | **0.7239** | 0.7194 | **0.6952** | 0.3820 |
| | $n = 1000$ | **0.8588** | 0.7675 | **0.8186** | 0.7117 |
| | $n = 2000$ | **0.9105** | 0.8625 | **0.8377** | 0.7852 |
| 20% | $n = 200$ | **0.7223** | 0.7049 | **0.6141** | 0.3789 |
| | $n = 1000$ | **0.8544** | 0.8359 | **0.7492** | 0.7159 |
| | $n = 2000$ | **0.9075** | 0.8738 | 0.7645 | **0.7943** |
| 50% | $n = 200$ | **0.7100** | 0.7093 | **0.5301** | 0.3764 |
| | $n = 1000$ | **0.8500** | 0.8294 | 0.6758 | **0.7431** |
| | $n = 2000$ | 0.8542 | **0.8727** | 0.7310 | **0.8445** |

VAR(2)) are slightly better than True $t_j$s gL-$VAR(2)$, except percentage 50%, $n = 2000$. The results of both methods become better with the increase of sample size. For different change percentages, there is no significant difference in the results of both methods.

For $M = 7$, change percentages 10% and 20%, the results of KWgL-VAR(2) are better than True $t_j$s gL-$VAR(2)$, except percentage 20%, $n = 2000$. For change percentages 50%, True $t_j$s gL-$VAR(2)$ have higher $F1$ scores than KWgL-VAR(2), except $n = 200$. Small changes percentages such as 10% and 20% of the network structure are closer to the assumption of smoothly varying than big changes percentage 50%. And kernel reweighted methods can borrow more adjacent observation. Then KWgL-VAR(2) perform better than True $t_j$s gL-$VAR(2)$ for small sample size such as $n = 200$ and small changes percentages such as 10% and 20%.

This experiment shows that the proposed kernel reweighted group Lasso method for smoothly varying networks may have good performance in some cases of structural breaks. However, it should be noted that we use true number and locations of change points in this paper, which need to be estimated in application. While the number and locations of change points are not required to be estimated advance in KWgL-VAR(2). In practice, for instance, for resting-state fMRI signals, correlation analysis reveals both slowly varying and abruptly changing characteristics corresponding to modularities in brain functional networks [42]. Therefore, the proposed method can be used when the two types of time-varying features exist simultaneously to estimate potential time-varying network structure.

### 4.2. Stock market networks analysis

In this subsection, we apply the proposed method to a financial dataset to construct the time-varying Granger causality network of stock returns and analyse the structural features of the network. The advantages of Granger causality for recognition of the complex and high dimensional relationships between variables are further verified.

#### 4.2.1. Experimental settings and graph theoretic metrics

In this experiment, we empirically validate the effectiveness of the proposed approach for reconstructing the time-varying financial networks using time series of the New York Stock Exchange (NYSE) dataset [32]. We select 44 stocks (31 from the Dow Jones Industrial Average index and 13 from the Dow Jones Utility Average index) from the NYSE dataset. Our sample consists of the daily closing prices of these stocks over 3817 transaction days from January 1996 to February 2011. We adopt the logarithm of return, i.e., $Y_i(t) = \log P_i(t) - \log P_i(t-1)$, where $P_i(t)$ is the closing price of the $i-th$ stock at day $t$. Then we get a sample observations with $d = 44, n = 3816$. All the dataset were obtained from the public financial dataset on Yahoo (http://finance.yahoo.com).The dataset is representative, encompassing both tranquil periods and crisis periods with global events such as the 1997 Asian financial crisis (October 27, 1997), 1998 Russian financial crisis (August 17,1998), the Dot-com bubble burst (March 13, 2000), September 11 attacks (September 11, 2001), Iraq War (March 20, 2003), the New century financial bankruptcy (April 4, 2007), the Lehman crisis (September 15, 2008) [32].

The causal network is composed of firms as vertices and causality directions as edges. The causality direction is measured by the Granger causality test, which is the most popularly used method for evaluating causality effects of time series. If there is an arrow between any two companies, it means the companies at time $t$ may Granger cause the returns of another company at time $t+1$. In the Granger causality test, the AIC criterion is used to select the order of the VAR($p$) model, which is roughly determined to be 2 for our dataset. The regularization parameter is selected to satisfy such that the sparsity of the networks are around 0.1. Then Algorithm 1 is performed to recover the underlying varying Granger causality graphs for a time-varying VAR(2) model.

Then we obtain $n-2$ Granger causality graphs $G(t) = (V, E(t)), t = 3, 4, \cdots, n$ at each time points and apply the networks to describe the financial network connections and evolution. To demonstrate the connectedness and the structural features, we employ several measures based on network [10] as follows.

The degree of Granger causality (DGC) is represented as the fraction of significant causality relations in all $d(d-1)$ pairs of nodes, where $d = |V|$ is the number of vertices in the graph,

$$DGC = \frac{\sum_{i=1}^{d}\sum_{j\neq i}(j \to i)}{d(d-1)}. \tag{20}$$

Node degree expresses the number of linkages connected to the node. The out degree measure counts the number of nodes significantly caused by the node $j$, or to say, it measures the summation of causality directions from individual firm $j$ to others.

$$g_j^{out} = \frac{\sum_{j\neq i}(j \to i)}{d-1}. \tag{21}$$

The in degree measure counts the number of institutions that significantly cause the node $j$, or to say, it counts the incoming causality directions from other firms to firm $j$.

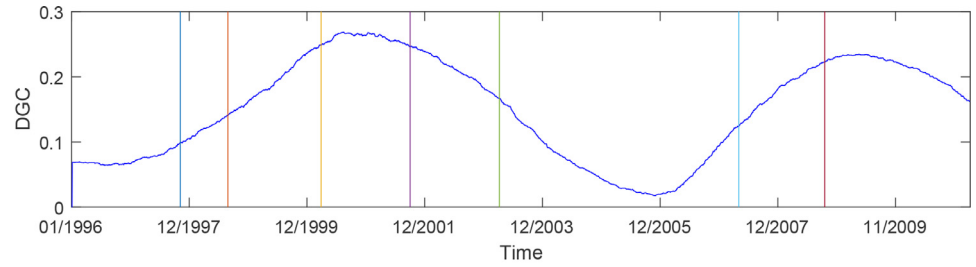
$$g_j^{in} = \frac{\sum_{i\neq j}(i \to j)}{d-1}. \tag{22}$$

The degree centrality measures the connectivity of nodes in the graph through computing the average of out degree and in degree.

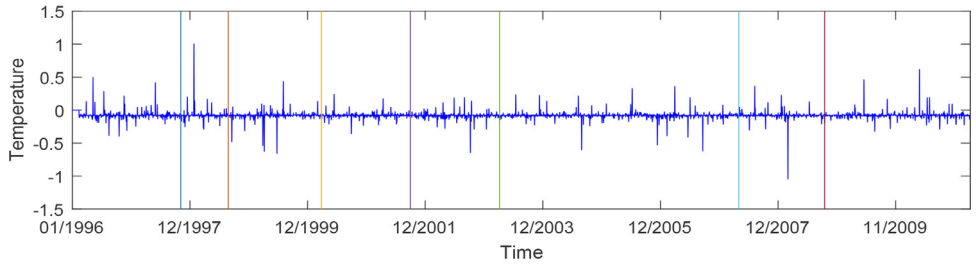
$$g_j = \frac{g_j^{out} + g_j^{in}}{2} = \frac{\sum_{i\neq j}(i \to j) + (j \to i)}{2(d-1)}. \tag{23}$$

#### 4.2.2. Network structure evolution

Then we investigate the structure evolution characteristics of the financial networks. We compute the degree of Granger causality based on Eq. (20) of the time-varying Granger causality graphs.




(a) Results based on our method




(b) Results based on the method in [32]

**Fig. 1.** The degree of Granger causality of the time-varying Granger causality graphs and the temperature of the stock correlation network versus time respectively.

The results are shown in Fig. 1(a). Fig. 1 presents the evolutionary behavior of the NYSE stock market network, in which the x-axis corresponds to the date(time), the y-axis corresponds to the different measures, and the vertical lines correspond to the financial crises.

To have a better comparison, we explore the evolutionary behaviour of the dataset by the thermodynamic characterization method of the correlation-based dynamic networks [32]. For each time window of 28 days, we compute the cross correlation coefficients between the time-series for each pair of stocks, and create connections between them if the maximum absolute value of the correlation coefficient is among the highest 5% of the total cross correlation coefficients. This yields a time-varying stock market network. We apply the thermodynamic characterization method to the dynamic networks. For convenience, we adopt $Temperature_t = \frac{T_t}{max(T)}$, where $T_t$ is the temperature variables according Eq.(30) in [32]. The results are shown in Fig. 1(b).

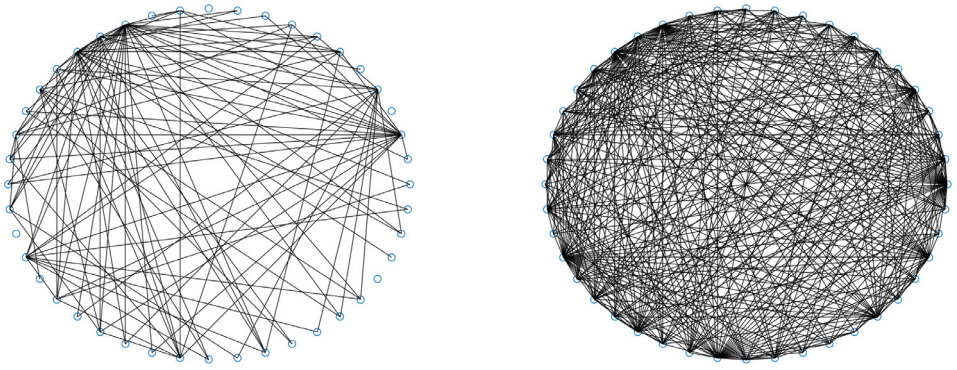
Fig. 1 (a) shows that most of the significant degrees of Granger causality(the 3rd, 4th and 7th vertical lines) successfully correspond to some realistic serious financial crises. The temperature measures in Fig. 1(b) are not significant higher during crisis than that of steady periods, which is not consistent with the financial observations that crises tighten the interactions between stocks. The NYSE dataset in [32] is consisted of 347 stocks and 5976 trading days. Since the smaller sample size and the number of nodes, in Fig. 1(b), most realistic serious financial crisis are not as significant fluctuations as that in [32]. From Fig. 1(a) we can see that the degree of change is large in crisis period and small in stable period. Particularly, due to the outbreak of the global recession and financial crisis that began in 2007, Fig. 1(a) shows there are higher values in recent years, which is also founded in temperature series [32]. The results further verify the effectiveness of our method. In addition, the thermodynamic characterization of network structure characterizes the evolution of time-varying complex networks. Ye et al. [32] showed that the thermodynamic variables can be computed in terms of simple network characteristics, e.g., the total number of nodes and node degree statistics for nodes connected by edges. The degree of Granger causality in Fig. 1(a) is a node degree statistics. The thermodynamic characterization of directed network structure in the similar way will be more comparable. We leave these work for future research.

The financial market networks are not invariant over time, but rather are context-dependent and can undergo systematic variety at each time point. Here, we choose two representative points of time among the whole sample period: one is before the crisis and the other is during the crisis. The results are shown in Fig. 2. In this way we can find a dynamic network structure through the sample period. For convenience, we omits the direction of edges in Fig. 2. Fig. 2(a) shows the network structure on October 25, 1996, which is before the crisis. We obtain a sparse network, which indicates that firms did not connect closely with each other in the steady period. Fig. 2(b) presents the network structure on March 13, 2000, where the network was very intensive during crisis time. The firms in the network were heavily connected with each other. Fig. 2 shows that the network structure was generally sparse in a stable period but became tighter during the crisis period. The degree of Granger causality for the two networks are 0.0655 and 0.2484, respectively. Our study confirms that during a normal period the market is very modular and heterogeneous, whereas during an instability (crisis) the market is more homogeneous, highly connected and less modular [43].

#### 4.2.3. Out degree and in degree distributions

Furthermore, Granger causality relation provides information about the direction of the connection. The topology of causal network can be discovered on the basis of degree distribution. We compute the out degree(Eq. (21)) and the in degree (Eq. (22)) of the time-varying Granger causality networks $G(t)$. The distributions of the in degree and out degree of the Granger causality graph on crisis period are shown in Fig. 3, in which the x-axis represents the degrees and the y-axis represents the frequency of the in degree and out degree. In the remainder of this subsection, we only show results of some time points during different period, the results of other events, such as 1998 Russian financial crisis (August 17,1998), September 11 attacks (September 11, 2001) and Iraq War (March 20, 2003) are similar. For comparison, Fig. 4 presents the distributions during steady period. Main results are as follows.

- The four histograms in Fig. 3 show the similar characteristics. The frequencies of nodes with 0 out degree are less than 0.4



(a) Network before crisis (10/25/1996) (b) Network during Dot-com bubble burst(03/13/2000)

**Fig. 2.** Network representatives of stock market.

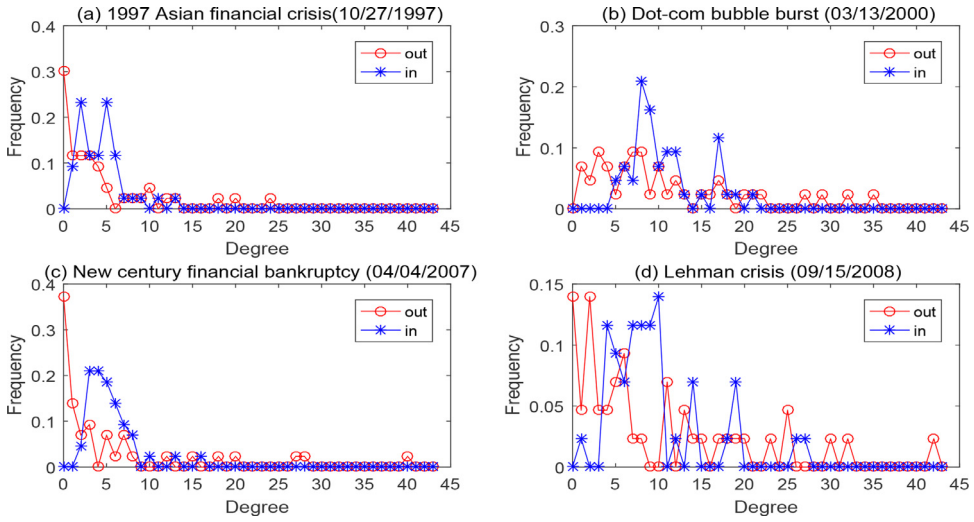



**Fig. 3.** Indegree and outdegree distribution of the time-varying Granger causality graphs for stock market during crisis.

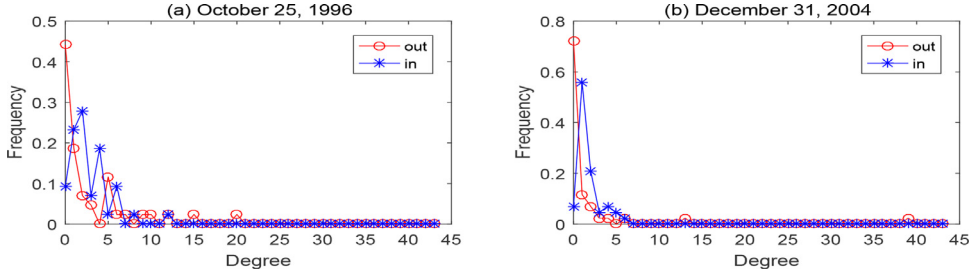



**Fig. 4.** Indegree and outdegree distribution of the time-varying Granger causality graphs for stock market during steady period.

in Fig. 3(a) and (c), and the frequencies are even smaller(0.15) in Fig. 3(b) and (d). While in Fig. 4(a) and (b), the frequencies of nodes with 0 out degree are about 0.45 and 0.7 respectively. The results show that more firms effected to others during crisis.

- The in degree distribution is similar to that of the out degree. In Fig. 3, the value range of in degree is about $[0, 15]$ (Fig. 3(a),(c)) and $[0, 25]$ (Fig. 3(b), (d)), while in Fig. 4 the range is within $[0, 5]$. The results show that firms affected by more others during crisis.
- The comparison between Fig. 3 and Fig. 4 is consistent with the financial observations that crises tighten the interactions between stocks and the influence of dominant stocks becomes stronger. It can be discovered that the network interconnectedness is affected by the shocks. The financial system becomes highly interconnected during crisis. Lu et al.[44] calculated the differential networks, namely, the difference between the adjacency matrices corresponding to the scenarios before and in crisis durations, to measure quantitatively the structural changes of the entities network. A total of 30 stocks that are used to construct the Do Jones index are considered in the period from 1994 to 2013. The conclusion of [44] shows that the influences of the financial crises share some features, for example in the crises the entities are tightly linked into dense clusters. At the same time, the influence of each financial crisis has its own features. The four subgraphs in Fig. 3 show different

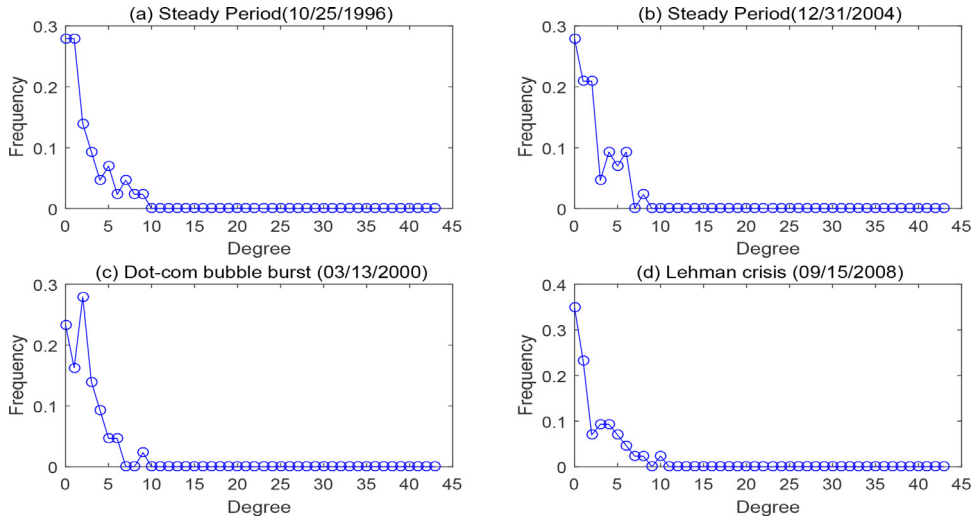



**Fig. 5.** Degree distribution of the correlation-based dynamic network [39] for stock market.

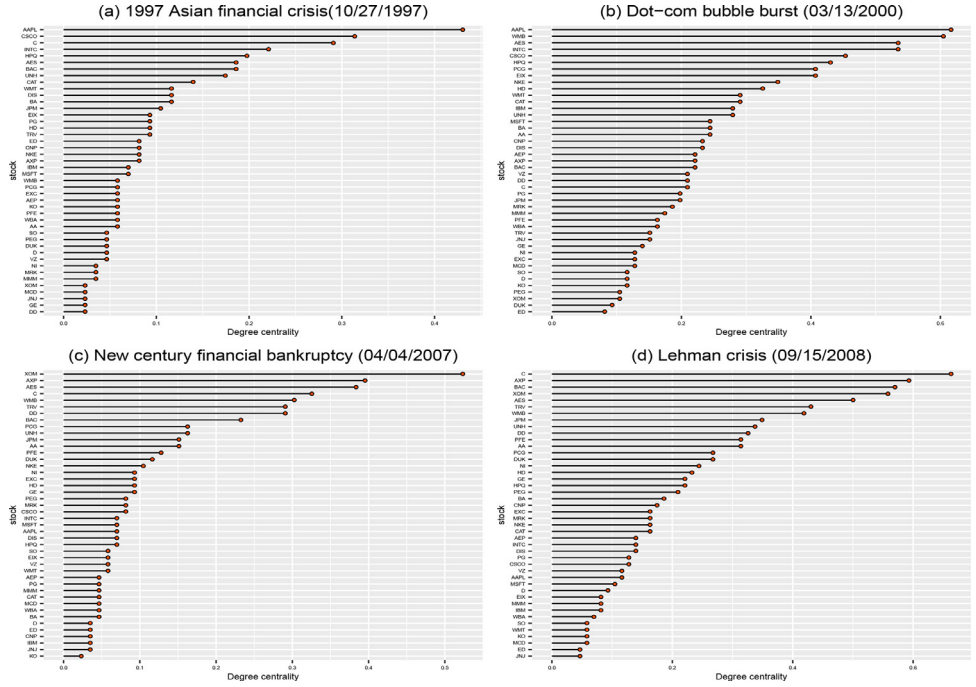



**Fig. 6.** Degree Centrality of nodes during crisis.

features. The results of analysis based on the proposed graphical models accord with the results of real observations[44]. The causality network can serve as important systemic risk indicator.

In order to compare, we also compute the degree distributions of the undirected dynamic network based on sliding window correlation coefficients [39]. The results are shown in Fig. 5, where the degree $g_i = \sum_{j \neq i}(i - j)$. There are not significant difference between the degree distributions during steady period (Fig. 5(a),(b)) and during crisis (Fig. 5 (c),(d)). These results further verify the superiority of network based on Granger causality versus network based on correlation coefficients.

*4.2.4. Degree centrality*

The research on the topological structure of networks has important theoretical and practical significance to understand the formation mechanism of complex financial networks. In order to further reveal the topological characteristics in the network of financial system, we compute the degree centrality (Eq. (23)) of 44 stocks and exhibit the results during the crisis period and steady period in Fig. 6 and Fig. 7 respectively.

Through the comparison of degree centrality graphs, there are obvious distinctions in the degree centrality measure between the crisis period (Fig. 6) and the steady period (Fig. 7). The degree centrality values decay at a slower rate in the crisis period, while the values decay rapidly during the steady period. This indicates that more stocks have the bigger degree centrality value during the crisis, while more stocks have the smaller degree centrality value during steady period. The topology structures are different between bull and bearish periods, and higher centrality value also implies critical role in the systemic. The roles of nodes are also time-varying. It indicates that these network based connectedness measures can identify and quantify the financial crisis. According to the centrality results, financial institutions with the highest val-

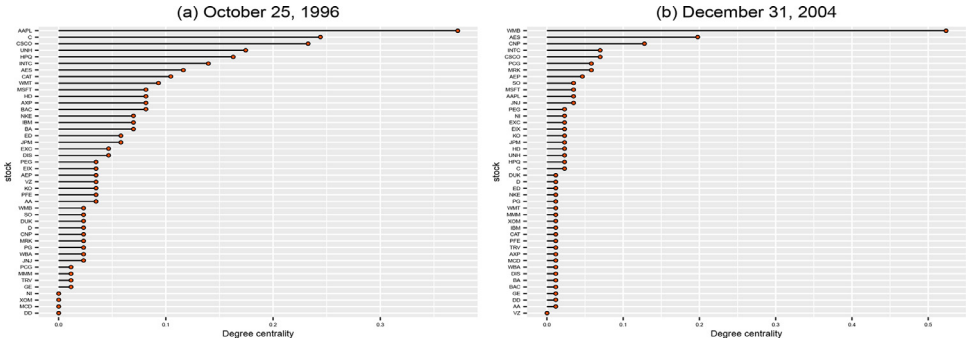



**Fig. 7.** Degree Centrality of nodes during steady period.

ues are heavily connected, and hence suffered the most during crisis. The results confirms that the impact of each financial crisis on the industry is significantly different (Fig. 4 in [44]). Furthermore, we can analyze the evolutionary characteristics of individual stocks based on the time-varying Gragner causality networks.

Interconnections can amplify financial shocks and exacerbate contagion of losses. The time-varying Granger causality graphs of financial institutions can examine the interconnectedness and evolution of structural features. The above analysis shows that the crisis period exhibited the most connectivity in financial companies. Specially, from the Figs. 1(a)-4, Figs. 6-7, we can conclude that the value of dynamic causality index indicates strong causality relationships within financial institutions in crisis period. Time-varying Granger causality graphs can measure the evolution of financial system through global measures such as DGC, as well as the variety of each stocks through individual measures such as degree centrality, distributions of in and out degrees.

## 5. Conclusion and future work

Detecting the evolution characteristics of causal network for high dimensional systems is a key issue in many applied domains. In this paper, we propose time-varying group Lasso Granger causality graphs to model and analyse the time-varying directed relationships underlying multivariate time series. A kernel reweighted group lasso method is proposed to recover the time-varying Granger causality structures from observations. The lagged temporal variables are grouped according to the time series they belong to, which makes it possible to efficiently model causality for high dimensional VAR(p) models with time-varying coefficients. We show that the time-varying group Lasso Granger causality graphs can efficiently identify the evolution of dependence relations features in high dimensional time series with $p$ order ($p \geq 1$) lag relations. Numerical experiments through two simulation settings(time series models with smoothly varying structure and change points respectively) demonstrate the effectiveness of the proposed method. The method is applied to construct time-varying networks for New York Stock Exchange(NYSE) dataset. The results show that the time-varying group Lasso Granger causality graphs can reveal the evolution features of causal relations for the time-varying networks.

Despite these benefits of the proposed method, several issues require further investigation. First, there are a number of alternative implementations that may be pursued. For example, our work applies the group structure of the lagged temporal variables to improve the estimation. An important problems that remain to be investigated is how to incorporation of domain knowledge of the variables as another class of group structure into the model. The network structure and parameter are assumed to change smoothly overtime. Pyne et al. [45] proposed an algorithm that does not impose any structural constraints for reconstructing time-varying gene regulatory networks with shortlisted candidate regulators (denoted TGS). In future work, we plan to relax the assumption about the time-varying structure, for example, considering the combination with TGS. In addition, combining with the neural Granger causality methods [37] can provide developments about time-varying nonlinear causality inference.

Second is about the application. In the instance of NYSE dataset, we analyze the evolution of dynamic Granger causality networks and the distribution of in degree and out degree during steady period and crisis period. The goal is to verify the advantage of the proposed method for recovering the dependence structure from observations. For future study, measures to characterize the Granger causality network structure should be defined (such as entropy measures in [32,34,39]) to capture interconnectedness and detect crisis. On the other hand, sometimes we are more interested in the characteristic of an individual node than the global behavior of the whole network. In such cases, there are many indicators such as degree centrality and node degree to characterize the centrality of nodes. Then how to depict the relative importance of the vertex in the network and determine whether it is core or peripheral is another future research work. The method can also be applied to many kinds of scientific fields and directly facilitate new discoveries.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant no. 11601404), the Youth Innovation Team of Shaanxi Universities, and the Yanta Scholars Foundation of Xian University of Finance and Economics.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2022.108789

## References

[1] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans, Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[2] M. Wang, J. Hu, H.A. Abbass, Brainprint: EEG biometric identification based on analyzing brain connectivity graphs, Pattern Recognit. 105 (2020) 107381.
[3] H. Lim, D. Kim, Pairwise dependence-based unsupervised feature selection, Pattern Recognit. 111 (2021) 107663.
[4] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica. 37 (3) (1969) 424–438.

[5] Y. Sun, J. Li, J. Liu, et al., Using causal discovery for feature selection in multivariate numerical time series, Mach. Learn. 101 (1–3) (2015) 377–395.
[6] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel Granger causality and the analysis of dynamical networks, Phys. Rev. E 77 (2008) 056215.
[7] P. Schwab, D. Miladinovic, W. Karlen, W. Karlen, Granger-causal attentive mixtures of experts: learning important features with neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Hawaii, 2019, pp. 4846–4853.
[8] G. Bhattacharya, K. Ghosh, A.S. Chowdhury, Granger causality driven AHP for feature weighted kNN, Pattern Recognit 66 (2017) 425–436.
[9] M. Eichler, Graphical modelling of multivariate time series, Probab. Theory Rel. 153 (2012) 233–268.
[10] X. Gong, X. Liu, X. Xiong, et al., Financial systemic risk measurement based on causal network connectedness analysis, Int. Rev. Econ. Finance 64 (2019) 290–307.
[11] Y. Chen, J. Hu, W. Zhang, Too connected to fail? Evidence from a Chinese financial risk spillover network, China & World Economy, 28 (6) (2018) 78–100.
[12] Z. Zhang, Y. Tian, L. Bai, J. Xiahou, E.R. Hancock, High-order covariate interacted lasso for feature selection, Pattern Recognit, Lett. 87 (2017) 139–146.
[13] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat Soc. B. 58 (1996) 267–288.
[14] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat, Soc. B. 68 (2006) 49–67.
[15] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. B. 67 (5) (2015) 301–320.
[16] L. Cui, L. Bai, Y. Wang, X. Jin, E.R. Hancock, Internet financing credit risk evaluation using multiple structural interacting elastic net feature selection, Pattern Recognit. 114 (2021) 107835.
[17] L. Cui, L. Bai, Z. Zhang, Y. Wang, E.R. Hancock, Identifying the most informative features using a structurally interacting elastic net, Neurocomputing 336 (2019) 13–26.
[18] L. Cui, L. Bai, Y. Wang, X. Jin, E.R. Hancock, Fused lasso for feature selection using structural information, Pattern Recognit. 119 (2021) 108058.
[19] A. Arnold, Y. Liu, N. Abe, Temporal causal modeling with graphical Granger methods, Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2007) 66–75.
[20] A.C. Lozano, N. Abe, Y. Liu, S. Rosset, Grouped graphical Granger modeling for gene expression regulatory networks discovery, Bioinformatics. 25 (2009) 110–118.
[21] A. Bolstad, B.D.V. Veen, R. Nowak, Causal network inference via group sparse regularization, Signal Processing: A publication of the IEEE Signal Processing Society. 59 (6) (2011) 2628–2641.
[22] S. Basu, A. Shojaie, G. Michailidis, Network granger causality with inherent grouping structure, J. Mach. Learn. Res. 16 (2015) 417–453.
[23] G. Yang, L. Wang, X. Wang, Reconstruction of complex directional networks with group lasso nonlinear conditional granger causality, Sci. Rep.UK. 7 (2017) 1–13.
[24] V. Jethava, C. Bhattacharyya, D. Dubhashi, Computational approaches for reconstruction of time-varying biological networks from omics data, IEE P. Syst. Biol. 1 (2013) 209–239.
[25] S. Cekic, D. Grandjean, O. Renaud, Time, frequency and time-varying Granger–causality measures in neuroscience, Stat. Med. 37 (2018) 1910–1931.
[26] Y. Li, M.Y. Lei, Y.Z. Guo, Z.Y. Hu, H.L. Wei, Time-varying nonlinear causality detection using regularized orthogonal least squares and multi-wavelets with applications to EEG, IEEE Access. 6 (2018) 17826–17840.
[27] Z. Wang, E.E.K. Lu, X. Yang, Y. Xu, T.S. Huang, Time varying dynamic Bayesian network for nonstationary events modeling and online inference, IEEE T. Signal Proces. 59 (4) (2011) 1553–1568.
[28] M. Iacopini, L. Rossini, Bayesian nonparametric graphical models for time-varying parameters VAR, SSRN Electronic Journal (2019) 1906.02140.
[29] L. Song, M. Kolar, E. Xing, Time-varying dynamic Bayesian networks, In: the 22th annual conference of Advances in Neural Information Processing Systems (2009).
[30] M.M. Guo, Y.J. Wang, G.Z. Xu, G. Milsap, N.V. Thakor, N. Crone, Time-varying dynamic bayesian network model and its application to brain connectivity using electrocortico graph, Acta Phys. Sin.CH. 65 (3) (2016). 038702-1-11
[31] Y. Yin, P. Shang, J. Xia, Compositional segmentation of time series in the financial markets, Appl. Math. Comput. 268 (2015) 399–412.
[32] C. Ye, C.H. Comin, T.K. Peron, F.N. Silva, F.A. Rodrigues, L.F. Costa, A. Torsello, E.R. Hancock, Thermodynamic characterization of networks using graph polynomials, Phys. Rev. E. 92 (3) (2015) 032810.
[33] L. Bai, L. Rossi, L. Cui, J. Cheng, E.R. Hancock, A quantum-inspired similarity measure for the analysis of complete weighted graphs, IEEE T. on Cybernetics 50 (3) (2020) 1264–1277, doi:10.1109/TCYB.2019.2913038.
[34] L. Bai, L. Cui, L. Xu, Y. Wang, Z. Zhang, E.R. Hancock, Entropic dynamic time warping kernels for co-evolving financial time series analysis, IEEE T. Neur. Net. Lear. 99 (2020) 1–15, doi:10.1109/TNNLS.2020.3006738.
[35] S. Guo, C. Ladroue, Granger causality: theory and applications, Frontiers in Computational and Systems Biology. Comput. Biol. 15 (2010) 83–111. Springer, London
[36] S. Guo, Y. Wang, Q. Yao, High dimensional and banded vector autoregression, Biometrika. 103 (2016) 889–903.
[37] A.S. Chivukula, J. Li, W. Liu, Discovering Granger-causal features from deep learning networks, Advances in Artificial Intelligence. Lecture Notes in Computer Science, 11320, Springer, Cham, 2018, doi:10.1007/978-3-030-03991-262.
[38] F.B. Lu, Y.M. Hong, S.Y. Wang, K.K. Lai, J. Liu, Time-varying Granger causality tests for applications in global crude oil markets, Energy Econ. 42 (1) (2014) 289–298.
[39] C. Ye, R.C. Wilson, E.R. Hancock, Network analysis using entropy component analysis, Journal of Complex Networks. 6 (3) (2018) 404–429, doi:10.1093/comnet/cnx045.
[40] J. Liu, J. Ye, Moreau-yosida regularization for grouped tree structure learning, In: the 23th annual conference of Advances in Neural Information Processing Systems, pp.1459-1467, 2010.
[41] J. Liu, S. Ji, J. Ye, SLEP: Sparse learning with efficient projections, 2011. http://www.public.asu.edu/~jye02/Software/SLEP.
[42] M. Hutchison, T. Womelsdorf, J. Gati, S. Everling, R. Menon, Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques, Hum. Brain Mapp 34 (2013) 2154–2177.
[43] A. Samal, H.K. Pharasi, S.J. Ramaia, H. Kannan, E. Saucan, J. Jost, A. Chakraborti, Network geometry and market instability, R. Soc. Open Sci. 8 (2021) 201734.
[44] L. Qiu, T.M. Jia, H.J. Yang, Differitial network investigated influences of financial crises on industries, Acta Phys. Sin., 65 (19) (2016) 198901.
[45] S. Pyne, A.R.R. Kumar, A. Anand, Rapid reconstruction of time-varying gene regulatory networks, IEEE ACM T. Comput. Bi. 17 (1) (2020) 278–291, doi:10.1109/TCBB.2018.2861698.

**Wei Gao** is currently a professor in the School of Statistics, Xi'an University of Finance and Economics in Xi'an, China. She received the Ph.D., M.S. degree in applied mathematics from Northwestern Polytechnical University, Xi'an, China. Her research interest includes time series analysis and graphical models and the applications to economics.

**Haizhong Yang** is currently a lecturer in the School of Statistics, Xi'an University of Finance and Economics in Xi'an, China. He received the M.S. degree in statistics and the Ph.D. degree in systems engineering from Northwestern Polytechnical University. His research interest includes asymptotic analysis and quantitative risk management.