



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/212538/>

Version: Published Version

Article:

Chen, Jia, Li, Degui, Li, Yu-Ning et al. (2025) Estimating Time-Varying Networks for High-Dimensional Time Series. *Journal of Econometrics*. 105941. ISSN: 0304-4076

<https://doi.org/10.1016/j.jeconom.2024.105941>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Estimating time-varying networks for high-dimensional time series

Jia Chen ^{a,b}, Degui Li ^{c,d}, Yu-Ning Li ^e, Oliver Linton ^{f,*}

^a Department of Economics, University of Macau, China

^b Department of Economics and Related Studies, University of York, UK

^c Faculty of Business Administration, University of Macau, China

^d Department of Mathematics, University of York, UK

^e School for Business and Society, University of York, UK

^f Faculty of Economics, University of Cambridge, UK



ARTICLE INFO

JEL classification:

C13
C32
C38
C55

Keywords:

Factor model
Granger causality
Partial correlation
Time-varying network
VAR

ABSTRACT

We explore time-varying networks for high-dimensional locally stationary time series, using the large VAR model framework with both the transition and (error) precision matrices evolving smoothly over time. Two types of time-varying graphs are investigated: one containing directed edges of Granger causality linkages, and the other containing undirected edges of partial correlation linkages. Under the sparse structural assumption, we propose a penalised local linear method with time-varying weighted group LASSO to jointly estimate the transition matrices and identify their significant entries, and a time-varying CLIME method to estimate the precision matrices. The estimated transition and precision matrices are then used to determine the time-varying network structures. Under some mild conditions, we derive the theoretical properties of the proposed estimates including the consistency and oracle properties. In addition, we extend the methodology and theory to cover highly-correlated large-scale time series, for which the sparsity assumption becomes invalid and we allow for common factors before estimating the factor-adjusted time-varying networks. We provide extensive simulation studies and an empirical application to a large U.S. macroeconomic dataset to illustrate the finite-sample performance of our methods.

1. Introduction

In recent years, the network analysis has become an effective tool to explore inter-connections among a large number of variables, with applications to various disciplines such as epidemiology, economics, finance, and social networks (e.g., [Newman, 2002](#); [Burt et al., 2013](#); [Diebold and Yilmaz, 2014, 2015](#); [Hautsch et al., 2014](#); [Scott, 2017](#); [Barigozzi and Brownlees, 2019](#); [Zhu et al., 2019](#)). The so-called graphical model is commonly used in the network analysis to visualise the connectedness of a large panel with vertices representing variables in the panel and the presence of an edge indicating appropriate (conditional) dependence between the variables. In the past decades, most of the existing literature on statistical estimation and inference of network data limits attention to the *static* network, which is assumed to be invariant over time (e.g., [Yuan and Lin, 2007](#); [Fan et al., 2009](#); [Loh and Wainwright, 2013](#); [Basu et al., 2015](#); [Zhao et al., 2022](#)). However, such an assumption may be too restrictive and often fails in practical applications where the underlying data generating mechanism is dynamic. There have been some attempts in the recent literature to relax the static network assumption, allowing the connectivity structure to exhibit time-varying features. For example, [Kolar et al. \(2010\)](#) and [Zhou et al. \(2010\)](#) study dynamic network models with smooth time-varying structural changes; whereas [Wang et al. \(2021\)](#)

* Corresponding author.

E-mail addresses: chenjia@um.edu.mo (J. Chen), degui@um.edu.mo (D. Li), yuning.li@york.ac.uk (Y.-N. Li), obl20@cam.ac.uk (O. Linton).

consider change-point detection and estimation in dynamic networks. However, most of the aforementioned literature typically assumes that the network data are independent, which often becomes invalid in practice. We aim to relax this restrictive assumption and model large-scale network data under a general temporal dependence structure.

Vector autoregression (VAR) is a fundamental modelling tool for multivariate time series data (e.g., Lütkepohl, 2006). In recent years, there has been increasing interest in extending the finite-dimensional VAR to the high-dimensional setting. Under appropriate sparsity restrictions on the transition (or autoregressive coefficient) matrices, various regularised methods have been proposed to estimate high-dimensional VAR models and identify non-zero entries in the transition matrices (e.g., Basu and Michailidis, 2015; Han et al., 2015; Kock and Callot, 2015; Davis et al., 2016). Zhu et al. (2017) introduce a network VAR model by incorporating the adjacency matrix to capture the network effect and estimate the model via ordinary least squares. More recently, Chen et al. (2023) and Miao et al. (2023) further study high-dimensional VAR and network VAR with latent common factors, allowing strong cross-sectional dependence in large panel time series. The methodology and theory developed in these papers heavily rely on the stationarity assumption with both transition and volatility matrices being time-invariant.

The stable VAR model cannot capture smooth structural changes and breaks in the underlying data generating process, two typical dynamic features in time series data collected over a long time span. To address this problem, Ding et al. (2017) consider a time-varying VAR model for high-dimensional time series (allowing the number of variables to diverge at a sub-exponential rate of the sample size), and estimate the time-varying transition matrices by combining the kernel smoothing with ℓ_1 -regularisation, whereas Safikhani and Shojaie (2022) simultaneously detect breaks and estimate transition matrices in high-dimensional VAR via a three-stage procedure using the total variation penalty. Xu et al. (2020) detect structural breaks and estimate smooth changes (between breaks) in the covariance and precision matrices of high-dimensional time series (covering VAR as a special case). In the present paper, we aim to estimate the time-varying transition and precision matrices in the high-dimensional VAR under the local stationarity framework. Motivated by the stable network time series analysis in Barigozzi and Brownlees (2019), we use the estimated transition and precision matrices to further construct two time-varying networks: one containing directed edges of Granger causality linkages, and the other containing undirected edges of partial correlation linkages.

The proposed time-varying network via VAR is naturally connected to the locally stationary models, which have been systematically studied in the literature for low-dimensional time series. Dahlhaus (1997) is among the first to introduce a locally stationary time series model via a time-varying spectral representation. Dahlhaus and Subba Rao (2006) study a time-varying ARCH model and propose a kernel-weighted quasi-maximum likelihood estimation method. Hafner and Linton (2010) further consider a time-varying version of GARCH model and introduce a semiparametric method to estimate both the parametric and nonparametric components involved. Vogt (2012) and Zhang and Wu (2012) study nonparametric kernel-based estimation and inference in a general class of locally stationary time series. Koo and Linton (2012) extend the locally stationary model framework to the diffusion process. Yan et al. (2020) develop a kernel estimation method and theory for time-varying vector moving average models. The present paper complements the locally stationary time series literature by further exploring the high-dimensional dynamic network structure.

We study the time-varying VAR and network models for large-scale time series, allowing the number of variables to be much larger than the time series length. Under the sparsity assumption on the transition and precision matrices with smooth structural changes, we introduce a three-stage estimation procedure: (i) preliminary local linear estimation of the transition matrices and their derivatives with time-varying LASSO; (ii) joint local linear estimation and feature selection of the time-varying transition matrices with weighted group LASSO; (iii) estimation of the precision matrix via time-varying CLIME. To guarantee the oracle property, the weights of LASSO in the second estimation stage are constructed via a local linear approximation to the SCAD penalty (e.g., Zou and Li, 2008) using the consistent preliminary estimates obtained in the first stage. Our penalised estimation methodology for the time-varying transition matrices is connected to various nonparametric screening and shrinkage methods developed for high-dimensional functional-coefficient models (e.g., Wang and Xia, 2009; Lian, 2012; Fan et al., 2014a; Liu et al., 2014; Li et al., 2015), whereas the time-varying CLIME is a natural extension of the conventional CLIME for static precision matrix estimation (e.g., Cai et al., 2011). The theoretical properties of the techniques developed in the aforementioned literature (such as the oracle property and minimax optimal convergence rates) rely on the independent data assumption. Extension of the methodology and theory to the high-dimensional locally stationary time series is non-trivial, requiring new technical tools such as the concentration inequality for time-varying VAR. Under some regularity conditions, we show that the proposed local linear estimates with weighted group LASSO equal to the infeasible oracle estimates with prior information on the significant entries of time-varying transition matrices, and the precision matrix estimate with time-varying CLIME is uniformly consistent with sensible convergence rates under various matrix norms. The estimated transition matrices are used to consistently estimate the uniform network structure with directed Granger causality linkages, whereas the estimated precision matrix is used to construct the network structure with undirected partial correlation linkages.

We further consider highly-correlated large-scale time series, for which the sparsity model assumption is no longer valid and the methodology and theory need to be substantially modified. The approximate factor model (e.g., Chamberlain and Rothschild, 1983) or its time-varying version (e.g., Su and Wang, 2017) is employed to accommodate the strong cross-sectional dependence among a large number of time series. In particular, we assume that the high-dimensional idiosyncratic error process in the approximate factor model satisfies the time-varying VAR structure with the sparsity restriction imposed on its transition and precision matrices. The latent common and idiosyncratic components need to be estimated consistently. With the approximated idiosyncratic error vectors, the penalised local linear estimation method with weighted group LASSO and time-varying CLIME are applied to estimate the time-varying transition and precision matrices. Subsequently, the factor-adjusted time-varying network estimates with directed Granger causality and undirected partial correlation linkages are obtained. By considering both time-varying transition matrices and

time-varying precision matrices in the VAR structure, our paper extends the recent work on the factor-adjusted stable VAR model estimation (e.g., [Krampe and Margaritella, 2022](#); [Barigozzi et al., 2024](#); [Fan et al., 2023](#)).

Our simulation studies demonstrate that the proposed methodology can accurately estimate the time-varying Granger and partial correlation networks when the number of time series variables is comparable to the sample size. In particular, for the time-varying transition matrix estimation, the penalised local linear method with weighted group LASSO outperforms the conventional local linear method (which often fails in the high-dimensional time series setting) and produces numerical results similar to those of the oracle estimation. For the time-varying error precision matrix estimation, the numerical performance of the proposed time-varying CLIME is comparable to that of the time-varying graphical LASSO. We further apply the developed methodology to the FRED-MD macroeconomic dataset and estimate both the Granger causality and partial correlation networks via the proposed time-varying VAR model. Through additional analysis of the time-varying coefficients during the pre- and post-global financial crisis periods, we demonstrate that the proposed method has the potential to enhance our understanding of the time-varying structure of the macroeconomy.

The rest of the paper is organised as follows. Section 2 introduces the time-varying VAR and network model structures. Section 3 presents the estimation procedures for the time-varying transition and precision matrices and Section 4 gives the asymptotic properties of the developed estimates. Section 5 considers the factor-adjusted time-varying VAR model and network estimation. Sections 6 and 7 report simulation studies and an empirical application, respectively. Section 8 concludes the paper. A supplemental document contains proofs of the main theorems, some technical lemmas with proofs, verification of a key assumption, discussions on tuning parameter selection, and additional simulation and empirical results. Throughout the paper, we let $|\cdot|_0$, $|\cdot|_1$, $\|\cdot\|$ and $|\cdot|_{\max}$ denote the L_0 , L_1 , L_2 (Euclidean) and maximum norms of a vector, respectively. Let \mathbf{I}_d and $\mathbf{O}_{d \times d}$ be a $d \times d$ identity matrix and null matrix, respectively. For a $d \times d$ matrix $\mathbf{W} = (w_{ij})_{d \times d}$, we let $\|\mathbf{W}\| = \lambda_{\max}^{1/2}(\mathbf{W}^T \mathbf{W})$ be the operator norm, $\|\mathbf{W}\|_F = [\text{Tr}(\mathbf{W}^T \mathbf{W})]^{1/2}$ the Frobenius norm, $\|\mathbf{W}\|_1 = \max_{1 \leq j \leq d} \sum_{i=1}^d |w_{ij}|$, $\|\mathbf{W}\|_{\max} = \max_{1 \leq i \leq d} \max_{1 \leq j \leq d} |w_{ij}|$, and $|\mathbf{W}|_1 = \sum_{i=1}^d \sum_{j=1}^d |w_{ij}|$, where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of a matrix and $\text{Tr}(\cdot)$ is the trace. Denote the determinant of a square matrix as $\det(\cdot)$. Let $a_n \sim b_n$, $a_n \propto b_n$ and $a_n \gg b_n$ denote that $a_n/b_n \rightarrow 1$, $0 < \underline{c} \leq a_n/b_n \leq \bar{c} < \infty$ and $b_n/a_n \rightarrow 0$, respectively.

2. Time-varying VAR and network models

In this section, we first introduce a locally stationary VAR model with time-varying transition and precision matrices, and then define two types of time-varying network structures with Granger causality and partial correlation linkages, respectively. Section 5 will further generalise them to factor-adjusted time-varying VAR and network settings.

2.1. Time-varying VAR models

Suppose that $(X_t : t = 1, \dots, n)$ with $X_t = (x_{t,1}, \dots, x_{t,d})^T$ is a sequence of d -dimensional random vectors generated by a time-varying VAR model of order p :

$$X_t = \sum_{k=1}^p \mathbf{A}_{t,k} X_{t-k} + e_t \quad \text{with} \quad e_t = \Sigma_t^{1/2} \varepsilon_t, \quad t = 1, \dots, n, \tag{2.1}$$

where $\mathbf{A}_{t,k} = \mathbf{A}_k(t/n)$, $k = 1, \dots, p$, are $d \times d$ time-varying transition matrices with each entry being a smooth deterministic function of scaled times, $\Sigma_t = \Sigma(t/n)$ is a $d \times d$ time-varying volatility matrix, and (ε_t) is a sequence of independent and identically distributed (i.i.d.) d -dimensional random vectors with zero mean and identity covariance matrix. Define $\Omega_t = \Omega(t/n)$ as the inverse of Σ_t , the time-varying precision matrix. We consider the ultra large time series setting, i.e., the dimension d is allowed to diverge at an exponential rate of the sample size n . The time-varying VAR model (2.1) is a natural extension of the finite-dimensional time-varying VAR to high-dimensional time series. If Σ_t is replaced by a time-invariant covariance matrix, (2.1) becomes the same model as that considered by [Ding et al. \(2017\)](#). Furthermore, when both $\mathbf{A}_{t,k}$, $k = 1, \dots, p$, and Σ_t are time-invariant constant matrices, (2.1) becomes the high-dimensional stable VAR:

$$X_t = \sum_{k=1}^p \mathbf{A}_k X_{t-k} + \Sigma^{1/2} \varepsilon_t, \tag{2.2}$$

which has been extensively studied in the recent literature (e.g., [Basu and Michailidis, 2015](#); [Han et al., 2015](#); [Kock and Callot, 2015](#); [Barigozzi and Brownlees, 2019](#); [Liu and Zhang, 2021](#)). Throughout the paper, we assume that the following conditions are satisfied.

Assumption 1. (i) Uniformly over $\tau \in [0, 1]$, it holds that $\det(\mathbf{I}_d - \sum_{k=1}^p \mathbf{A}_k(\tau) z^k) \neq 0$ for any $z \in \mathbb{C}$ with modulus no larger than one, where \mathbb{C} denotes the set of complex numbers. Each entry in $\mathbf{A}_k(\cdot)$ is second-order continuously differentiable over $[0, 1]$.

(ii) The precision matrix $\Omega(\tau)$ is positive definite uniformly over $\tau \in [0, 1]$, and the operator norm of $\Sigma(\tau)$ is uniformly bounded over $\tau \in [0, 1]$. Furthermore, each entry in $\Sigma(\tau)$ and $\Omega(\tau)$ is second-order continuously differentiable over $[0, 1]$.

(iii) For any d -dimensional vector u satisfying $\|u\| = 1$, $E[\exp\{t_1(u^T \varepsilon_t)^2\}] \leq C_0 < \infty$, where t_1 and C_0 are positive constants.

The first condition in [Assumption 1](#)(i) is a natural extension of the stability assumption imposed on the constant transition matrices (e.g., [Lütkepohl, 2006](#)), indicating that the time-varying VAR process is locally stationary/stable. Without loss of generality, we assume the following Wold representation:

$$X_t = \sum_{k=0}^{\infty} \Phi_{t,k} e_{t-k}, \tag{2.3}$$

with the coefficient matrices $\Phi_{t,k}$ satisfying that, for k sufficiently large,

$$\max_{0 \leq t \leq n} \|\Phi_{t,k}\| \leq C_1 \rho^k, \tag{2.4}$$

where C_1 is a positive constant and $0 < \rho < 1$. A similar assumption can be found in Ding et al. (2017). In some special model settings, (2.4) may be violated, and we refer the interested readers to the discussions in Basu and Michailidis (2015) and Liu and Zhang (2021). In fact, the condition (2.4) may be removed by imposing some high-level conditions (e.g., the sub-Gaussian condition on $x_{t,i}$ proved in Lemma B.1). The smoothness conditions in Assumption 1(i)(ii) are common in kernel-based local estimation method and theory. The sub-Gaussian moment condition in Assumption 1(iii) is not uncommon in the literature of high-dimensional feature selection and covariance/precision matrix estimation (e.g., Wainwright, 2019), and is weaker than the Gaussian assumption frequently used in the high-dimensional VAR literature (e.g., Basu and Michailidis, 2015; Kock and Callot, 2015).

2.2. Time-varying network structures

Write $\mathbf{A}_{t,k} = (a_{k,ij|t})_{d \times d}$, $\mathbf{\Omega}_t = (\omega_{ij|t})_{d \times d}$, $\mathbf{A}_k(\tau) = (a_{k,ij}(\tau))_{d \times d}$ and $\mathbf{\Omega}(\tau) = (\omega_{ij}(\tau))_{d \times d}$, where $1 \leq t \leq n$ and $0 \leq \tau \leq 1$. We define the network structure via a time-varying graph $\mathbb{G}_t = (\mathbb{V}, \mathbb{E}_t)$, where $\mathbb{V} = \{1, 2, \dots, d\}$ denotes a set of vertices, and $\mathbb{E}_t = \{(i, j) \in \mathbb{V} \times \mathbb{V} : c_{ij|t} \neq 0, i \neq j\}$ denotes a time-varying set of edges. The choice of $c_{ij|t}$ is determined by the definition of linkage. The construction of \mathbb{G}_t is similar to that in Kolar et al. (2010) and Zhou et al. (2010) for independent network data. Following the stable network analysis in Barigozzi and Brownlees (2019) and Barigozzi et al. (2024), we next consider two types of time-varying linkages: the directed Granger causality linkage and undirected partial correlation linkage.

The definition of Granger causality is first introduced by Granger (1969) to investigate the causal relations in small economic time series systems. In the context of stable VAR (with order p), we say that $x_{t,j}$ Granger causes $x_{t,i}$ if there exists $k \in \{1, 2, \dots, p\}$ such that $x_{t-k,j}$ improves predictability of $x_{t,i}$ by reducing the forecasting error. It is a natural idea to use the stable transition matrices $\mathbf{A}_k = (a_{k,ij})_{d \times d}$ in (2.2) to determine the Granger causality structure, i.e., if there exists at least one k such that $a_{k,ij} \neq 0$, then $x_{t,j}$ Granger causes $x_{t,i}$. We may extend the stable Granger causality structure to a more general time-varying version using (2.1). At a given time point t , we say that lags of $x_{t,j}$ Granger cause $x_{t,i}$ if there exists at least one k such that $a_{k,ij|t} \neq 0$. Hence, for given $\tau \in (0, 1)$, we define the time-varying local graph $\mathbb{G}_\tau^G = (\mathbb{V}, \mathbb{E}_\tau^G)$ with

$$\mathbb{E}_\tau^G = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \exists k \in \{1, 2, \dots, p\}, a_{k,ij}(\tau) \neq 0\}. \tag{2.5}$$

The partial correlation is a commonly-used conditional dependence measure for network time series. We next extend it to the time-varying setting using $\mathbf{\Omega}_t = \mathbf{\Omega}(t/n)$ in (2.1). Let $\rho_{ij|t} = \text{cor}(e_{t,i}, e_{t,j} | e_{t,k}, k \neq i, j)$ be the time-varying (contemporaneous) partial correlation between the innovations $e_{t,i}$ and $e_{t,j}$, where $e_{t,i}$ is the i th element of e_t . Following Dempster (1972), we may show that $\rho_{ij|t} \neq 0$ is equivalent to $\omega_{ij|t} \neq 0$ for $i \neq j$. Hence, we can construct the set of edges by collecting the index pairs of the non-zero entries in the time-varying precision matrix. For $\tau \in (0, 1)$, define the local graph $\mathbb{G}_\tau^P = (\mathbb{V}, \mathbb{E}_\tau^P)$ with

$$\mathbb{E}_\tau^P = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \omega_{ij}(\tau) \neq 0, i \neq j\}. \tag{2.6}$$

In practice, the primary interest often lies in the full network structures over the entire time interval. This requires the construction of a uniform version of \mathbb{G}_τ^G and \mathbb{G}_τ^P . Denote the uniform graphs by $\mathbb{G}^G = (\mathbb{V}, \mathbb{E}^G)$ and $\mathbb{G}^P = (\mathbb{V}, \mathbb{E}^P)$, with

$$\mathbb{E}^G = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \exists k \in \{1, 2, \dots, p\} \text{ and } \tau \in (0, 1), a_{k,ij}(\tau) \neq 0\} \tag{2.7}$$

and

$$\mathbb{E}^P = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \exists \tau \in (0, 1), \omega_{ij}(\tau) \neq 0, i \neq j\}. \tag{2.8}$$

It is easy to verify that $\mathbb{E}_\tau^G \subset \mathbb{E}^G$ and $\mathbb{E}_\tau^P \subset \mathbb{E}^P$ for any $\tau \in (0, 1)$. Section 3.4 below defines the discrete versions of the above uniform networks and provides their estimates.

3. Methodology

Let $A_{k,i}^\top(\cdot)$ and $C_i^\top(\cdot)$ be the i th row of $\mathbf{A}_k(\cdot)$ and $\mathbf{\Omega}^{-1/2}(\cdot)$, respectively,

$$\alpha_{i,\cdot}(\cdot) = \left[A_{1,i}^\top(\cdot), \dots, A_{p,i}^\top(\cdot) \right]^\top, \quad \mathbf{X}_t = \left(X_t^1, \dots, X_{t-p+1}^1 \right)^\top, \tag{3.1}$$

and $\tau_t = t/n$. The time-varying VAR model (2.1) can be equivalently written as

$$x_{t,i} = \alpha_{i,\cdot}^\top(\tau_t) \mathbf{X}_{t-1} + e_{t,i} \quad \text{with} \quad e_{t,i} = C_i^\top(\tau_t) \varepsilon_t, \quad i = 1, \dots, d, \tag{3.2}$$

which is a high-dimensional time-varying coefficient autoregressive model with a scalar response and pd candidate predictors for each i . As the dimension of the predictors is allowed to be ultra large, we need to impose an appropriate sparsity restriction on the vector of time-varying parameters $\alpha_{i,\cdot}(\cdot)$ to limit the number of its significant elements. High-dimensional varying-coefficient models have been systematically studied in the literature and various nonparametric screening and shrinkage methods have been proposed to select the significant covariates, estimate the coefficient functions and identify the model structure under the independent data assumption (e.g., Wang et al., 2008; Wang and Xia, 2009; Lian, 2012; Cheng et al., 2014; Fan et al., 2014a; Liu et al., 2014; Li et al.,

2015). In this section, under the high-dimensional locally stationary time series framework, we propose a three-stage procedure to estimate the Granger causality and partial correlation network structures: (i) first obtain preliminary local linear estimates of $\alpha_{i,\cdot}(\cdot)$ (and its derivatives) using time-varying LASSO, which serves as a first-stage screening of the predictors in \mathbf{X}_{t-1} ; (ii) conduct local linear estimation and feature selection using weighted group LASSO, where the weights are constructed via a local linear approximation to the SCAD penalty using the preliminary estimates of $\alpha_{i,\cdot}(\cdot)$ from Stage (i); (iii) estimate the error precision matrix $\Omega(\cdot)$ via the time-varying CLIME method. The estimated transition and precision matrices are finally used to construct the uniform network structures.

3.1. Preliminary time-varying LASSO estimation

For $\tau \in (0, 1)$, under the smoothness condition on the transition matrices in Assumption 1(i), we have the following local linear approximation to $\alpha_{i,\cdot}(\tau_t)$:

$$\alpha_{i,\cdot}(\tau_t) \approx \alpha_{i,\cdot}(\tau) + \alpha'_{i,\cdot}(\tau)(\tau_t - \tau), \quad i = 1, \dots, d,$$

when τ_t falls within a small neighbourhood of τ , where $\alpha'_{i,\cdot}(\cdot)$ is a (pd) -dimensional vector of the first-order derivatives of the elements in $\alpha_{i,\cdot}(\cdot)$. Hence, for each $i \in \{1, 2, \dots, d\}$ and a given $\tau \in (0, 1)$, we define the following local linear objective function (e.g., Fan and Gijbels, 1996):

$$\mathcal{L}_i(\alpha, \beta \mid \tau) = \frac{1}{n} \sum_{i=1}^n \left\{ x_{t,i} - [\alpha + \beta(\tau_t - \tau)]^\top \mathbf{X}_{t-1} \right\}^2 K_h(\tau_t - \tau), \tag{3.3}$$

where $K_h(\cdot) = \frac{1}{h} K(\cdot/h)$ with $K(\cdot)$ being a kernel function and h being a bandwidth or smoothing parameter. The estimates of $\alpha_{i,\cdot}(\tau)$ and $\alpha'_{i,\cdot}(\tau)$ are obtained by minimising $\mathcal{L}_i(\alpha, \beta \mid \tau)$ with respect to α and β . However, this local linear estimation is only feasible when the dimension of the predictors is fixed or significantly smaller than the sample size n (e.g., Cai, 2007; Li et al., 2011). In our high-dimensional setting, as the number of predictors may exceed n , it is challenging to obtain satisfactory estimation by directly minimising $\mathcal{L}_i(\alpha, \beta \mid \tau)$. To address this issue, we assume that the number of significant components in $\alpha_{i,\cdot}(\tau)$ is much smaller than n and then incorporate a LASSO penalty term in the local linear objective function (3.3).

The LASSO estimation was first introduced by Tibshirani (1996) in the context of linear regression and has become one of the most commonly-used tools in high-dimensional variable and feature selection. We next adopt a time-varying version of the LASSO estimation. Define

$$\mathcal{L}_i^*(\alpha, \beta \mid \tau) = \mathcal{L}_i(\alpha, \beta \mid \tau) + \lambda_1 (|\alpha|_1 + h|\beta|_1), \tag{3.4}$$

where λ_1 is a tuning parameter. Let $\tilde{\alpha}_{i,\cdot}(\tau)$ and $\tilde{\alpha}'_{i,\cdot}(\tau)$ be the solution to the minimisation of $\mathcal{L}_i^*(\alpha, \beta \mid \tau)$ with respect to α and β . We call them the preliminary time-varying LASSO estimates. This LASSO estimation may not accurately identify the true significant predictors, but can remove a large number of irrelevant predictors and hence, serves as a preliminary screening step. Furthermore, the first-stage estimates will be used to construct weights in the weighted group LASSO in the second stage to more precisely estimate the time-varying parameters and accurately select the significant predictors.

3.2. Penalised local linear estimation with weighted group LASSO

In order to estimate the uniform Granger causality network, we next introduce a global penalised method to simultaneously estimate the time-varying parameters at τ_t , $t = 1, \dots, n$, and identify the non-zero index sets $\mathcal{J}_i = \bigcup_{t=1}^n \mathcal{J}_i(\tau_t)$ and $\mathcal{J}'_i = \bigcup_{t=1}^n \mathcal{J}'_i(\tau_t)$, where

$$\mathcal{J}_i(\tau) = \{1 \leq j \leq pd : \alpha_{i,j}(\tau) \neq 0\} \quad \text{and} \quad \mathcal{J}'_i(\tau) = \{1 \leq j \leq pd : \alpha'_{i,j}(\tau) \neq 0\}$$

with $\alpha_{i,j}(\cdot)$ and $\alpha'_{i,j}(\cdot)$ being the j th element of $\alpha_{i,\cdot}(\cdot)$ and $\alpha'_{i,\cdot}(\cdot)$, respectively. For each i , note that identifying the zero elements in $\alpha'_{i,\cdot}(\tau_t)$ (uniformly over t) is equivalent to identifying the indices j , $1 \leq j \leq pd$, such that $D_{i,j} = 0$, where

$$D_{i,j}^2 = \sum_{t=1}^n \left[\alpha_{i,j}(\tau_t) - \frac{1}{n} \sum_{s=1}^n \alpha_{i,j}(\tau_s) \right]^2.$$

In practice, $D_{i,j}^2$ can be approximated by

$$\tilde{D}_{i,j}^2 = \sum_{t=1}^n \left[\tilde{\alpha}_{i,j}(\tau_t) - \frac{1}{n} \sum_{s=1}^n \tilde{\alpha}_{i,j}(\tau_s) \right]^2,$$

using the preliminary time-varying LASSO estimates $\tilde{\alpha}_{i,j}(\tau_t)$, $t = 1, \dots, n$. Let $\mathbf{A} = (\alpha_{\cdot,1}, \dots, \alpha_{\cdot,n})^\top$ with $\alpha_{\cdot,t} = (\alpha_{1|t}, \dots, \alpha_{pd|t})^\top$, and $\mathbf{B} = (\beta_{\cdot,1}, \dots, \beta_{\cdot,n})^\top$ with $\beta_{\cdot,t} = (\beta_{1|t}, \dots, \beta_{pd|t})^\top$. We define a global version of the penalised objective function with weighted group LASSO:

$$\mathcal{Q}_i(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^n \mathcal{L}_i(\alpha_{\cdot,t}, \beta_{\cdot,t} \mid \tau_t) + \sum_{j=1}^{pd} p'_{\lambda_2} \left(\|\tilde{\alpha}_{i,j}\| \right) \|\alpha_j\| + \sum_{j=1}^{pd} p'_{\lambda_2} \left(\tilde{D}_{i,j} \right) \|h\beta_j\|, \tag{3.5}$$

where

$$\tilde{\alpha}_{i,j} = [\tilde{\alpha}_{i,j}(\tau_1), \dots, \tilde{\alpha}_{i,j}(\tau_n)]^\top, \quad \alpha_j = (\alpha_{j1}, \dots, \alpha_{jn})^\top, \quad \beta_j = (\beta_{j1}, \dots, \beta_{jn})^\top,$$

while λ_2 is a tuning parameter and $p'_\lambda(\cdot)$ is the derivative of the SCAD penalty function:

$$p'_\lambda(z) = \lambda \left[I(z \leq \lambda) + \frac{(a_0 \lambda - z)_+}{(a_0 - 1)\lambda} I(z > \lambda) \right],$$

with $a_0 = 3.7$ as suggested in Fan and Li (2001) and $I(\cdot)$ being the indicator function. The penalty terms in (3.5) are motivated by the local linear approximation to the SCAD penalty function (Zou and Li, 2008). The terms $p'_{\lambda_2}(\|\tilde{\alpha}_{i,j}\|)$ and $p'_{\lambda_2}(\|\tilde{D}_{i,j}\|)$ in (3.5) serve as the weights for the group LASSO, and their values are determined by the preliminary estimates in Section 3.1, i.e., the corresponding weight is heavy when $\|\tilde{\alpha}_{i,j}\|$ or $\|\tilde{D}_{i,j}\|$ is close to zero, whereas it is light or equal to zero when $\|\tilde{\alpha}_{i,j}\|$ or $\|\tilde{D}_{i,j}\|$ is large.

An advantage of using $\tilde{D}_{i,j}$ in the second penalty term over the L_2 -norm of $\tilde{\alpha}'_j = [\tilde{\alpha}'_{i,j}(\tau_1), \dots, \tilde{\alpha}'_{i,j}(\tau_n)]^\top$ is that the estimates of the time-varying parameters involved in $\tilde{D}_{i,j}$ often perform more stably than their derivative counterparts.

Let $\hat{\mathbf{A}}_i$ and $\hat{\mathbf{B}}_i$ be the minimiser of $\Omega_i(\mathbf{A}, \mathbf{B})$ with respect to \mathbf{A} and \mathbf{B} , where

$$\begin{aligned} \hat{\mathbf{A}}_i &= (\hat{\alpha}_{i,1}, \dots, \hat{\alpha}_{i,pd}) \quad \text{with} \quad \hat{\alpha}_{i,j} = [\hat{\alpha}_{i,j}(\tau_1), \dots, \hat{\alpha}_{i,j}(\tau_n)]^\top, \\ \hat{\mathbf{B}}_i &= (\hat{\alpha}'_{i,1}, \dots, \hat{\alpha}'_{i,pd}) \quad \text{with} \quad \hat{\alpha}'_{i,j} = [\hat{\alpha}'_{i,j}(\tau_1), \dots, \hat{\alpha}'_{i,j}(\tau_n)]^\top. \end{aligned}$$

The index set \mathcal{J}_i is estimated by $\hat{\mathcal{J}}_i = \{j : \hat{\alpha}_{i,j} \neq \mathbf{0}_n\}$, and \mathcal{J}'_i is estimated by $\hat{\mathcal{J}}'_i = \{j : \hat{\alpha}'_{i,j} \neq \mathbf{0}_n\}$, where $\mathbf{0}_k$ is a k -dimensional vector of zeros. A similar shrinkage estimation method is used by Li et al. (2015) and Chen et al. (2021) to identify a high-dimensional semi-varying coefficient model structure for independent data. So far as we know, there is no work on such a penalised technique and its relevant theory for high-dimensional locally stationary time series data.

3.3. Estimation of the time-varying precision matrix

In this section, we study the estimation of $\Omega(\cdot)$ in model (2.1), which is crucial to uncover the time-varying and uniform network structures based on partial correlations. Estimation of large static precision matrices has been extensively studied under the sparsity assumption, and various estimation techniques, such as the penalised likelihood, graphical Danzig selector and CLIME, have been proposed in the literature (e.g., Lam and Fan, 2009; Yuan, 2010; Cai et al., 2011). Xu et al. (2020) further introduce a time-varying CLIME method for high-dimensional locally stationary time series which are observable. Note that in this paper, $\Omega(\cdot)$ is the time-varying precision matrix for the high-dimensional unobservable error vector e_t and hence, its estimation requires substantial modification of the time-varying CLIME methodology and theory.

With $\hat{\alpha}_{i,\cdot}(\cdot)$, $i = 1, \dots, d$, from Section 3.2, we can then extract estimates of the time-varying transition matrices, denoted by $\hat{\mathbf{A}}_k(\tau)$, $t = 1, \dots, n$, $k = 1, \dots, p$, and approximate e_t by

$$\hat{e}_t = (\hat{e}_{t,1}, \dots, \hat{e}_{t,d})^\top = X_t - \sum_{k=1}^p \hat{\mathbf{A}}_k(\tau) X_{t-k}, \quad t = 1, \dots, n. \tag{3.6}$$

The approximation accuracy depends on the uniform prediction rates of the time-varying weighted group LASSO estimates. In order to apply the time-varying CLIME, we assume that $\Omega(\cdot)$ satisfies a uniform sparsity assumption, a natural extension of the classic sparsity assumption to the locally stationary time series setting. Specifically, we assume $\{\Omega(\tau) : 0 \leq \tau \leq 1\} \in \mathcal{S}(q, \xi_d)$, where

$$\mathcal{S}(q, \xi_d) = \left\{ \mathbf{W}(\tau) = [w_{ij}(\tau)]_{d \times d}, 0 \leq \tau \leq 1 : \mathbf{W}(\tau) > 0, \sup_{0 \leq \tau \leq 1} \|\mathbf{W}(\tau)\|_1 \leq C_2, \sup_{0 \leq \tau \leq 1} \max_{1 \leq i \leq d} \sum_{j=1}^d |w_{ij}(\tau)|^q \leq \xi_d \right\}, \tag{3.7}$$

where $0 \leq q < 1$, “ $\mathbf{W} > 0$ ” denotes that \mathbf{W} is positive definite, and C_2 is a bounded positive constant. Specially, we adopt the definition $0^0 = 0$. Define

$$\hat{\Sigma}(\tau) = [\hat{\sigma}_{ij}(\tau)]_{d \times d} \quad \text{with} \quad \hat{\sigma}_{ij}(\tau) = \sum_{t=1}^n \varpi_{n,t}(\tau) \hat{e}_{t,i} \hat{e}_{t,j} / \sum_{t=1}^n \varpi_{n,t}(\tau), \tag{3.8}$$

where the weight function $\varpi_{n,t}(\cdot)$ is constructed via the local linear smoothing:

$$\varpi_{n,t}(\tau) = K\left(\frac{\tau_t - \tau}{b}\right) s_{n,2}(\tau) - K_1\left(\frac{\tau_t - \tau}{b}\right) s_{n,1}(\tau),$$

in which $s_{n,j}(\tau) = \sum_{t=1}^n K_j\left(\frac{\tau_t - \tau}{b}\right)$, $K_j(x) = x^j K(x)$, and b is a bandwidth. With the uniform sparsity assumption (3.7), we estimate $\Omega(\tau)$ via the time-varying CLIME method:

$$\hat{\Omega}(\tau) = [\hat{\omega}_{ij}(\tau)]_{d \times d} = \arg \min_{\Omega} \|\Omega\|_1 \quad \text{subject to} \quad \|\hat{\Sigma}(\tau)\Omega - \mathbf{I}_d\|_{\max} \leq \lambda_3, \tag{3.9}$$

where λ_3 is a tuning parameter. As the underlying time-varying precision matrix is symmetric, the matrix estimate obtained from (3.9) needs to be symmetrised to obtain the final estimate, denoted as $\hat{\tilde{\Omega}}(\tau) = [\hat{\tilde{\omega}}_{ij}(\tau)]_{d \times d}$, where

$$\hat{\tilde{\omega}}_{ij}(\tau) = \hat{\omega}_{ij}(\tau) = \tilde{\omega}_{ij}(\tau) I(|\tilde{\omega}_{ij}(\tau)| \leq |\tilde{\omega}_{ji}(\tau)|) + \tilde{\omega}_{ji}(\tau) I(|\tilde{\omega}_{ij}(\tau)| > |\tilde{\omega}_{ji}(\tau)|). \tag{3.10}$$

3.4. Estimation of uniform time-varying networks

In practice, when the sample size n is sufficiently large, it is often sensible to approximate the uniform edge sets, \mathbb{E}^G and \mathbb{E}^P , by the following discrete versions:

$$\mathbb{E}_n^G = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \exists k \in \{1, 2, \dots, p\} \text{ and } t \in \{1, \dots, n\}, a_{k,ij}(\tau_t) \neq 0\} \quad (3.11)$$

and

$$\mathbb{E}_n^P = \{(i, j) \in \mathbb{V} \times \mathbb{V} : \exists t \in \{1, \dots, n\}, \omega_{ij}(\tau_t) \neq 0, i \neq j\}. \quad (3.12)$$

Hence, we next estimate \mathbb{E}_n^G and \mathbb{E}_n^P instead of \mathbb{E}^G and \mathbb{E}^P . With the time-varying transition and precision matrix estimates in Sections 3.2 and 3.3, we can estimate \mathbb{E}_n^G by

$$\widehat{\mathbb{E}}_n^G = \left\{ (i, j) \in \mathbb{V} \times \mathbb{V} : \exists k \in \{1, 2, \dots, p\}, \sum_{t=1}^n \widehat{a}_{k,ij}^2(\tau_t) > 0 \right\}, \quad (3.13)$$

where $\widehat{a}_{k,ij}(\tau_t)$ denotes the (i, j) -entry of $\widehat{\mathbf{A}}_k(\tau_t)$, and estimate \mathbb{E}_n^P by

$$\widehat{\mathbb{E}}_n^P = \left\{ (i, j) \in \mathbb{V} \times \mathbb{V} : \exists t \in \{1, \dots, n\}, |\widehat{\omega}_{ij}(\tau_t)| \geq \lambda_3, i \neq j \right\}, \quad (3.14)$$

where λ_3 is the tuning parameter used in the time-varying CLIME.

4. Main theoretical results

To ease the notational burden, throughout this section, we focus on the time-varying VAR(1) model:

$$X_t = \mathbf{A}(\tau_t)X_{t-1} + \Sigma_t^{1/2}\varepsilon_t, \quad (4.1)$$

where $\mathbf{A}(\tau) = [\alpha_{ij}(\tau)]_{d \times d}$. For a general time-varying VAR(p) model (2.1), it can be equivalently re-written as a (pd) -dimensional VAR(1) model as follows:

$$\mathbf{X}_t = \mathbf{A}_t^* \mathbf{X}_{t-1} + \mathbf{e}_t,$$

where \mathbf{X}_t is defined in (3.1), $\mathbf{e}_t = (e_t^\top, 0_d^\top, \dots, 0_d^\top)^\top$, and \mathbf{A}_t^* is a $(pd) \times (pd)$ time-varying transition matrix:

$$\mathbf{A}_t^* = \begin{pmatrix} \mathbf{A}_{t,1} & \mathbf{A}_{t,2} & \dots & \mathbf{A}_{t,p-1} & \mathbf{A}_{t,p} \\ \mathbf{I}_d & \mathbf{O}_{d \times d} & \dots & \mathbf{O}_{d \times d} & \mathbf{O}_{d \times d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{O}_{d \times d} & \mathbf{O}_{d \times d} & \dots & \mathbf{I}_d & \mathbf{O}_{d \times d} \end{pmatrix}.$$

4.1. Uniform consistency of the time-varying LASSO estimates

Define

$$\boldsymbol{\Psi}(\tau) = \begin{bmatrix} \boldsymbol{\Psi}_0(\tau) & \boldsymbol{\Psi}_1(\tau) \\ \boldsymbol{\Psi}_1(\tau) & \boldsymbol{\Psi}_2(\tau) \end{bmatrix} \text{ with } \boldsymbol{\Psi}_k(\tau) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\tau_t - \tau}{h} \right)^k X_{t-1} X_{t-1}^\top K_h(\tau_t - \tau), \quad k = 0, 1, 2, \quad (4.2)$$

and

$$\mathcal{B}_i(\tau) = \left\{ \left(u_1^\top, u_2^\top \right)^\top : \|u_1\|^2 + \|u_2\|^2 = 1, \sum_{j=1}^d (|u_{1,j}| + |u_{2,j}|) \leq 3 \left(\sum_{j \in \mathcal{J}_i(\tau)} |u_{1,j}| + \sum_{j \in \mathcal{J}'_i(\tau)} |u_{2,j}| \right) \right\},$$

where $u_1 = (u_{1,1}, \dots, u_{1,d})^\top$ and $u_2 = (u_{2,1}, \dots, u_{2,d})^\top$ are two d -dimensional vectors and $\mathcal{J}_i(\tau)$ and $\mathcal{J}'_i(\tau)$ are defined as in Section 3.2 but with $p = 1$. To derive the uniform consistency property of the preliminary time-varying LASSO estimates defined in Section 3.1, we need the following assumptions, some of which may be weakened at the cost of lengthier proofs.

Assumption 2. (i) The kernel $K(\cdot)$ is a bounded, continuous and symmetric probability density function with a compact support $[-1, 1]$.

(ii) The bandwidth h satisfies

$$nh / \log^2(n \vee d) \rightarrow \infty \text{ and } sh^2 \log(n \vee d) \rightarrow 0,$$

where $s = \max_{1 \leq i \leq d} s_i$ with s_i being the cardinality of the index set \mathcal{J}_i .

Assumption 3. (i) The tuning parameter λ_1 satisfies

$$\zeta_{n,d} := \log(n \vee d) [(nh)^{-1/2} + sh^2] = o(\lambda_1) \text{ and } \sqrt{s} \lambda_1 / h \rightarrow 0.$$

(ii) There exists a positive constant κ_0 such that, with probability approaching one (w.p.a.1),

$$\min_{1 \leq i \leq d} \min_{1 \leq t \leq n} \inf_{u \in \mathcal{B}_i(\tau_t)} u^\top \Psi(\tau_t) u \geq \kappa_0. \tag{4.3}$$

Assumption 2(i) is a mild restriction which can be satisfied by some commonly-used kernels such as the uniform kernel and the Epanechnikov kernel. The compact support assumption on the kernel function is not essential and can be replaced by appropriate tail conditions. The bandwidth conditions in **Assumption 2**(ii) are crucial for deriving the uniform convergence properties of the kernel-based quantities. When s is bounded and d diverges at a polynomial rate of n , the conditions can be simplified to $nh/\log^2 n \rightarrow \infty$ and $h^2 \log n \rightarrow 0$. **Assumption 3**(ii) can be seen as a uniform version of the so-called restricted eigenvalue condition widely used in high-dimensional linear regression models (e.g., Bickel et al., 2009; Basu and Michailidis, 2015). Appendix D in the supplement provides sufficient conditions for the high-dimensional locally stationary Gaussian time series to satisfy **Assumption 3**(ii). Furthermore, with the Hanson-Wright inequality for time-varying (non-Gaussian) VAR processes (e.g., Proposition 6.2 in Zhang and Wu, 2021), we may show that $\max_{1 \leq i \leq n} \|\Psi(\tau_i) - E[\Psi(\tau_i)]\|_{\max} = O_p\left(\sqrt{\log(n \vee d)/(nh)}\right)$. Then, using Lemma D.1 in Appendix A and assuming $s\sqrt{\log(n \vee d)/(nh)} = o(1)$, a sufficient condition for (4.3) is

$$\min_{1 \leq i \leq d} \min_{1 \leq t \leq n} \inf_{u \in \mathcal{B}_i(\tau_t)} u^\top E[\Psi(\tau_t)] u \geq \kappa_0.$$

Theorem 4.1. Suppose that **Assumptions 1–3** are satisfied. Then we have

$$\max_{1 \leq i \leq d} \max_{1 \leq t \leq n} \|\tilde{\alpha}_{i,\cdot}(\tau_t) - \alpha_{i,\cdot}(\tau_t)\| = O_p\left(\sqrt{s\lambda_1}\right). \tag{4.4}$$

Theorem 4.1 shows that the preliminary time-varying LASSO estimates of the transition matrices are uniformly consistent with the convergence rate relying on s and λ_1 . Although the dimension d is allowed to diverge at an exponential rate of n , the number of significant elements in $\alpha_{i,\cdot}(\cdot)$ cannot diverge too fast in order to guarantee the consistency property. Furthermore, the uniform convergence result (4.4) can be strengthened to

$$\max_{1 \leq i \leq d} \sup_{0 \leq \tau \leq 1} \|\tilde{\alpha}_{i,\cdot}(\tau) - \alpha_{i,\cdot}(\tau)\| = O_p\left(\sqrt{s\lambda_1}\right). \tag{4.5}$$

A similar uniform convergence property holds for the first-order derivative function estimates, see (A.1) in the proof of **Theorem 4.1**.

4.2. The oracle property of the weighted group LASSO estimates

Denote the complement of \mathcal{J}_i and \mathcal{J}'_i as $\bar{\mathcal{J}}_i$ and $\bar{\mathcal{J}}'_i$, respectively, i.e., $\bar{\mathcal{J}}_i = \bigcap_{l=1}^n \{j : \alpha_{i,j}(\tau_l) = 0\}$ and $\bar{\mathcal{J}}'_i = \bigcap_{l=1}^n \{j : \alpha'_{i,j}(\tau_l) = 0\}$. Let $\mathbf{A}^o = (\alpha_{\cdot 1}^o, \dots, \alpha_{\cdot n}^o)^\top$ and $\mathbf{B}^o = (\beta_{\cdot 1}^o, \dots, \beta_{\cdot n}^o)^\top$, where $\alpha_{\cdot j}^o = (\alpha_{1j}^o, \dots, \alpha_{dj}^o)^\top$ with $\alpha_{jl}^o = 0$ for $j \in \bar{\mathcal{J}}_i$ and $\beta_{\cdot j}^o = (\beta_{1j}^o, \dots, \beta_{dj}^o)^\top$ with $\beta_{jl}^o = 0$ for $j \in \bar{\mathcal{J}}'_i$. Define the (infeasible) oracle estimates:

$$\hat{\mathbf{A}}_i^o = (\hat{\alpha}_{i,1}^o, \dots, \hat{\alpha}_{i,d}^o) \quad \text{with} \quad \hat{\alpha}_{i,j}^o = [\hat{\alpha}_{i,j}^o(\tau_1), \dots, \hat{\alpha}_{i,j}^o(\tau_n)]^\top, \tag{4.6}$$

$$\hat{\mathbf{B}}_i^o = (\hat{\alpha}'_{i,1}^o, \dots, \hat{\alpha}'_{i,d}^o) \quad \text{with} \quad \hat{\alpha}'_{i,j}^o = [\hat{\alpha}'_{i,j}^o(\tau_1), \dots, \hat{\alpha}'_{i,j}^o(\tau_n)]^\top, \tag{4.7}$$

as the values of \mathbf{A}^o and \mathbf{B}^o that minimise $\mathcal{Q}_i(\mathbf{A}^o, \mathbf{B}^o)$. We need to impose the following condition on the tuning parameter λ_2 and the lower bounds for the significant time-varying coefficients in the transition matrix.

Assumption 4. (i) The tuning parameter λ_2 satisfies

$$\sqrt{ns} \log(n \vee d) \zeta_{n,d} + \sqrt{ns} \lambda_1 = o(\lambda_2),$$

where $\zeta_{n,d}$ is defined in **Assumption 3**(i).

(ii) It holds that

$$\min_{1 \leq i \leq d} \min_{j \in \mathcal{J}_i} \left(\sum_{l=1}^n \alpha_{i,j}^2(\tau_l) \right)^{\frac{1}{2}} \geq (a_0 + 1)\lambda_2 \quad \text{and} \quad \min_{1 \leq i \leq d} \min_{j \in \mathcal{J}'_i} D_{i,j} \geq (a_0 + 1)\lambda_2,$$

where $a_0 = 3.7$ is defined in the SCAD penalty.

When s is a fixed positive integer, $h \propto n^{-1/5}$, $\lambda_1 \propto n^{-2/5+\eta_0}$ with $0 < \eta_0 < 1/5$, and $d \sim \exp\{n^{\eta_1}\}$ with $0 < \eta_1 < \eta_0$, it is easy to verify **Assumption 4**(i) by setting $\lambda_2 \propto n^{1/2-\eta_2}$ with $0 < \eta_2 < 2/5 - [\eta_0 \vee (2\eta_1)]$. **Assumption 4**(ii) imposes restrictions on the lower bounds for the time-varying coefficient functions and their deviations from the means. These restrictions are weaker than **Assumption 6**(ii) in Li et al. (2015) and **Assumption 8** in Chen et al. (2021), and they ensure that the significant coefficient functions and derivatives can be detected w.p.a.1.

Theorem 4.2. Suppose that **Assumptions 1–4** are satisfied. The minimiser to the objective function of the weighted group LASSO, $\mathcal{Q}_i(\mathbf{A}, \mathbf{B})$, exists and equals the oracle estimates defined in (4.6) and (4.7) w.p.a.1. In addition, we have the following mean squared convergence

result:

$$\max_{1 \leq i \leq d} \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d [\hat{\alpha}_{ij}(\tau_t) - \alpha_{ij}(\tau_t)]^2 = O_P \left(s \zeta_{n,d}^2 \right), \tag{4.8}$$

where s is defined in Assumption 2(ii) and $\zeta_{n,d}$ is defined in Assumption 3(i).

Since the penalised local linear estimates are identical to the infeasible oracle estimates defined in (4.6) and (4.7) *w.p.a.1*, the sparsity property holds for the global model selection procedures proposed in Section 3.2, i.e., the zero elements in the time-varying transition matrix can be estimated exactly as zeros. When s is a fixed positive integer, $h \propto n^{-1/5}$, $\lambda_1 \propto n^{-2/5+\eta_0}$ with $0 < \eta_0 < 1/5$, and $d \sim \exp \{n^{\eta_1}\}$ with $0 < \eta_1 < \eta_0$, it is easy to show that the convergence rate in Theorem 4.2 is $O(n^{-4/5+2\eta_1})$. Following the proof of Theorem 4.2, we may verify properties (i)–(iv) for the folded concave penalty function discussed in Fan et al. (2014b) *w.p.a.1*. Hence, Theorem 4.2 may be regarded as a generalisation of Theorem 1 in Fan et al. (2014b) and Theorem 3.1 in Li et al. (2015) to high-dimensional locally stationary time series.

With the oracle property in Theorem 4.2, it is straightforward to derive the following consistency property of the network estimates for the directed edges of Granger causality linkages.

Corollary 4.1. *Under the assumptions of Theorem 4.2, we have*

$$P \left(\hat{\mathbb{E}}_n^G = \mathbb{E}_n^G \right) \rightarrow 1. \tag{4.9}$$

4.3. Uniform consistency of the time-varying CLIME estimates

To derive the uniform consistency property of the time-varying CLIME estimates, we need the following conditions on the tuning parameters b and λ_3 .

Assumption 5. (i) The bandwidth b satisfies

$$b \rightarrow 0 \quad \text{and} \quad nb / [\log(n \vee d)]^3 \rightarrow \infty.$$

In addition, $s \zeta_{n,d} \sqrt{\log(n \vee d)} \rightarrow 0$, where $\zeta_{n,d}$ is defined in Assumption 3(i).

(ii) There exists a sufficiently large constant C_3 such that $\lambda_3 = C_3 \left(v_{n,d}^\circ + v_{n,d}^* \right)$, where

$$v_{n,d}^\circ = \left[\frac{\log(n \vee d)}{nb} \right]^{1/2} + b^2 \quad \text{and} \quad v_{n,d}^* = s \zeta_{n,d} \sqrt{\log(n \vee d)}.$$

The following theorem gives the uniform convergence rates of the time-varying precision matrix estimate $\hat{\mathbf{\Omega}}(\tau)$ under various matrix norms.

Theorem 4.3. *Suppose Assumptions 1–5 are satisfied and $\{\mathbf{\Omega}(\tau) : 0 \leq \tau \leq 1\} \in \mathcal{S}(q, \xi_d)$. Then we have*

$$\sup_{0 \leq \tau \leq 1} \left\| \hat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau) \right\|_{\max} = O_P \left(v_{n,d}^\circ + v_{n,d}^* \right), \tag{4.10}$$

$$\sup_{0 \leq \tau \leq 1} \left\| \hat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau) \right\| = O_P \left(\xi_d (v_{n,d}^\circ + v_{n,d}^*)^{1-q} \right), \tag{4.11}$$

$$\sup_{0 \leq \tau \leq 1} \frac{1}{d} \left\| \hat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau) \right\|_F^2 = O_P \left(\xi_d (v_{n,d}^\circ + v_{n,d}^*)^{2-q} \right), \tag{4.12}$$

where ξ_d is defined in (3.7), $v_{n,d}^\circ$ and $v_{n,d}^*$ are defined in Assumption 5(ii).

The uniform convergence rates in Theorem 4.3 rely on $v_{n,d}^\circ$ and $v_{n,d}^*$. The first rate $v_{n,d}^\circ$ is the conventional uniform convergence rate for nonparametric kernel-based quantities, whereas the second rate $v_{n,d}^*$ is from the approximation errors of \hat{e}_t to the latent VAR errors e_t . Note that the dimension d affects the uniform convergence rates via ξ_d and $\log(n \vee d)$, and the uniform consistency property holds in the ultra-high dimensional setting when d diverges at an exponential rate of n . For example, when s is a fixed positive integer, $h \propto n^{-1/5}$, $b \propto n^{-1/5}$, and $d \sim \exp \{n^{\eta_1}\}$ with $0 < \eta_1 < \eta_0$, it is easy to verify that $v_{n,d}^\circ + v_{n,d}^* \propto n^{-2/5+3\eta_1/2}$. Theorem 4.3 can be seen as an extension of Theorem 1 in Cai et al. (2011) to the high-dimensional locally stationary time series setting.

From Theorem 4.3, we readily have the following consistency property for the network estimates of the undirected edges of partial correlation linkages.

Corollary 4.2. *Under the assumptions of Theorem 4.3, if $\min_{(i,j) \in \mathbb{E}^P} \min_{1 \leq t \leq n} |\omega_{ij}(\tau_t)| \gg \lambda_3$, we have*

$$P \left(\hat{\mathbb{E}}_n^P = \mathbb{E}_n^P \right) \rightarrow 1. \tag{4.13}$$

5. Factor-adjusted time-varying VAR and networks

In this section, we let $(Z_t : t = 1, \dots, n)$ with $Z_t = (z_{t,1}, \dots, z_{t,d})^\top$ be an observed sequence of d -dimensional random vectors. To accommodate strong cross-sectional dependence which is not uncommon for large-scale time series collected in practice, we assume that Z_t is generated by an approximate factor model:

$$Z_t = \Lambda F_t + X_t, \quad t = 1, \dots, n, \tag{5.1}$$

where $\Lambda = (\Lambda_1, \dots, \Lambda_d)^\top$ is a $d \times k$ matrix of factor loadings, F_t is a k -dimensional vector of latent factors and (X_t) is assumed to satisfy the time-varying VAR model (2.1). More generally, we may assume the following time-varying factor model structure:

$$Z_t = \Lambda_t F_t + X_t, \quad t = 1, \dots, n, \tag{5.2}$$

where $\Lambda_t = \Lambda(t/n)$ is a time-varying factor loading matrix with each entry being a smooth function of scaled times. The approximate factor model and its time-varying generalisation have been extensively studied in the literature (e.g., Chamberlain and Rothschild, 1983; Bai and Ng, 2002; Stock and Watson, 2002; Motta et al., 2011; Su and Wang, 2017). Model (5.2) can be seen as a special case of the time-varying dynamic factor model (e.g., Eichler et al., 2011; Barigozzi et al., 2021). The primary interest of this section is to estimate the time-varying networks for the idiosyncratic error vector X_t . Even though the components of Z_t may be highly correlated, those of X_t are often only weakly correlated. Hence, it is sensible to impose the sparsity assumption on the time-varying transition and precision matrices of the idiosyncratic error process, making it possible to apply the estimation methodology proposed in Section 3. However, this is non-trivial as neither the common components (ΛF_t or $\Lambda_t F_t$) nor the idiosyncratic error components are observable. Motivated by recent work on bridging factor and sparse models for high-dimensional data (e.g., Krampe and Margaritella, 2022; Fan et al., 2023), we next use the principal component analysis (PCA) or its localised version to remove the common components driven by latent factors in the observed time series data.

Let $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, $\mathbf{F} = (F_1, \dots, F_n)^\top$ and $\mathbf{X} = (X_1, \dots, X_n)^\top$. For the conventional factor model (5.1), we conduct an eigenanalysis on the $n \times n$ matrix $\mathbf{Z}\mathbf{Z}^\top$. The estimate of \mathbf{F} , denoted as $\hat{\mathbf{F}} = (\hat{F}_1, \dots, \hat{F}_n)^\top$, is obtained as the $n \times k$ matrix consisting of the eigenvectors (multiplied by \sqrt{n}) corresponding to the k largest eigenvalues of $\mathbf{Z}\mathbf{Z}^\top$. The factor loading matrix is estimated by $\hat{\Lambda} = (\hat{\Lambda}_1, \dots, \hat{\Lambda}_d)^\top = \mathbf{Z}^\top \hat{\mathbf{F}}/n$. Consequently, the common component ΛF_t is estimated by $\hat{\Lambda} \hat{F}_t$ and the idiosyncratic error component X_t is estimated by

$$\hat{X}_t = Z_t - \hat{\Lambda} \hat{F}_t, \quad t = 1, \dots, n. \tag{5.3}$$

For the time-varying factor model (5.2), the above PCA estimation procedure needs some amendments. Specifically, let

$$K_{t,h_*}(\tau) = \frac{K_{h_*}(\tau_t - \tau)}{\sum_{s=1}^n K_{h_*}(\tau_s - \tau)}, \quad 0 < \tau < 1,$$

where h_* is a bandwidth and $K_{h_*}(\cdot)$ is defined as in Section 3.1, and define the localised data matrix:

$$\mathbf{Z}(\tau) = [Z_1(\tau), \dots, Z_n(\tau)]^\top \quad \text{with} \quad Z_t(\tau) = Z_t K_{t,h_*}^{1/2}(\tau).$$

Through an eigenanalysis on the matrix $\mathbf{Z}(\tau)\mathbf{Z}^\top(\tau)$, we can obtain the local PCA estimates of the factors and factor-loading matrix, denoted by $\hat{\mathbf{F}}(\tau) = [\hat{F}_1(\tau), \dots, \hat{F}_n(\tau)]^\top$ and $\hat{\Lambda}(\tau)$, respectively. Then, the idiosyncratic error vector X_t is approximated by

$$\hat{X}_t = Z_t - \hat{\Lambda}(\tau_t) \hat{F}_t(\tau_t), \quad t = 1, \dots, n, \tag{5.4}$$

where we keep the same notation \hat{X}_t as in (5.3) to avoid notational burden.

As in Section 4, we only consider the time-varying VAR(1) model for the idiosyncratic error vector. With the approximation \hat{X}_t , we can apply the three-stage estimation procedure proposed in Section 3. Denote the preliminary time-varying LASSO estimate as $\hat{\alpha}_{ij}^\dagger(\cdot)$, the second-stage weighted group LASSO estimate as $\hat{\alpha}_{ij}^\ddagger(\cdot)$, and the factor-adjusted time-varying precision matrix estimate as $\hat{\Omega}^\ddagger(\cdot) = [\hat{\omega}_{ij}^\ddagger(\cdot)]_{d \times d}$. Subsequently, we may construct the uniform network estimates $\hat{\mathbb{E}}_n^{G,\ddagger}$ and $\hat{\mathbb{E}}_n^{P,\ddagger}$, defined similarly to $\hat{\mathbb{E}}_n^G$ and $\hat{\mathbb{E}}_n^P$ in (3.13) and (3.14), but with $\hat{\alpha}_{ij}(\cdot)$ and $\hat{\omega}_{ij}(\cdot)$ replaced by $\hat{\alpha}_{ij}^\ddagger(\cdot)$ and $\hat{\omega}_{ij}^\ddagger(\cdot)$, respectively. To derive the convergence properties of these factor-adjusted estimates, we need the following assumption, which modifies Assumptions 3–5 to incorporate the approximation error of the idiosyncratic error components.

Assumption 6. (i) Denote $\delta_X = \max_{1 \leq t \leq n} \|\hat{X}_t - X_t\|_{\max}$. It holds that $[\log(n \vee d)]^{1/2} s \delta_X = o_p(1)$.

(ii) Assumption 3(i) holds when $\zeta_{n,d}$ is replaced by $\zeta_{n,d}^\ddagger = \zeta_{n,d} + [\log(n \vee d)]^{1/2} s \delta_X$.

(iii) Assumption 4(i) holds when $\zeta_{n,d}$ is replaced by $\zeta_{n,d}^\ddagger$.

(iv) Assumption 5 holds when $\zeta_{n,d}$ and $v_{n,d}^*$ are replaced by $\zeta_{n,d}^\ddagger$ and $v_{n,d}^\ddagger = s \zeta_{n,d}^\ddagger \sqrt{\log(n \vee d)}$, respectively.

Assumption 6(i) imposes a high-level condition on the approximation of latent X_t in the factor model, i.e., the approximation error δ_X uniformly converges to zero with a rate faster than $s^{-1}[\log(n \vee d)]^{-1/2}$. By Corollary 1 in Fan et al. (2013), a typical rate for the approximation error from PCA estimation of the conventional factor model (5.1) is

$$\delta_X = O_p \left((\log n)^{1/2} \left[(\log d)^{1/2} n^{-1/2} + n^{1/\nu} d^{-1/2} \right] \right), \tag{5.5}$$

where $\nu > 2$ is a positive number related to moment restrictions. From Theorem 3.5 in Su and Wang (2017), we may obtain the typical uniform rate for δ_X under the time-varying factor model (5.2) when the local PCA estimation is used. In Assumption 6(ii)–(iv), we amend Assumptions 3(i), 4(i) and 5(ii) to incorporate the approximation error δ_X . However, if we further assume that $h \propto n^{-1/5}$ and d diverges at a polynomial rate of n satisfying $d \gg n^{1+2/\nu}$, then the rate in (5.5) can be simplified to $\delta_X = O_P((\log d)n^{-1/2}) = o_P(h^2)$ and thus $\zeta_{n,d} \propto \zeta_{n,d}^\dagger$. Consequently, we may remove Assumption 6(ii)–(iv) and δ_X would not be involved in the estimation convergence rates under model (5.1).

The following two propositions extend the theoretical results in Section 4 to the factor-adjusted time-varying VAR and networks.

Proposition 5.1. *Suppose that the factor model (5.1) or (5.2), and Assumptions 1, 2 and 3(ii) are satisfied.*

(i) *Under Assumption 6(i)(ii), we have*

$$\max_{1 \leq i \leq d} \max_{1 \leq t \leq n} \sum_{j=1}^d \left[\hat{\alpha}_{ij}^\dagger(\tau_t) - \alpha_{ij}(\tau_t) \right]^2 = O_P(s\lambda_1^2). \tag{5.6}$$

(ii) *Under Assumption 6(i)–(iii), the oracle property holds for the second-stage weighted group LASSO estimates and furthermore,*

$$\max_{1 \leq i \leq d} \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d \left[\hat{\alpha}_{ij}^\dagger(\tau_t) - \alpha_{ij}(\tau_t) \right]^2 = O_P\left(s\left(\zeta_{n,d}^\dagger\right)^2\right). \tag{5.7}$$

(iii) *Under Assumption 6 and the sparsity condition that $\{\Omega(\tau) : 0 \leq \tau \leq 1\} \in \mathcal{S}(q, \xi_d)$, we have*

$$\sup_{0 \leq \tau \leq 1} \left\| \hat{\Omega}^\dagger(\tau) - \Omega(\tau) \right\|_{\max} = O_P\left(v_{n,d}^\circ + v_{n,d}^\dagger\right), \tag{5.8}$$

$$\sup_{0 \leq \tau \leq 1} \left\| \hat{\Omega}^\dagger(\tau) - \Omega(\tau) \right\| = O_P\left(\xi_d(v_{n,d}^\circ + v_{n,d}^\dagger)^{1-q}\right), \tag{5.9}$$

$$\sup_{0 \leq \tau \leq 1} \frac{1}{d} \left\| \hat{\Omega}^\dagger(\tau) - \Omega(\tau) \right\|_F^2 = O_P\left(\xi_d(v_{n,d}^\circ + v_{n,d}^\dagger)^{2-q}\right). \tag{5.10}$$

Proposition 5.2. (i) *Under the assumptions of Proposition 5.1(ii), we have*

$$P\left(\hat{\mathbb{E}}_n^{G,\dagger} = \mathbb{E}_n^G\right) \rightarrow 1. \tag{5.11}$$

(ii) *Under the assumptions of Proposition 5.1(iii) and $\min_{(i,j) \in \mathbb{E}^P} \min_{1 \leq t \leq n} |\omega_{ij}(\tau_t)| \gg \lambda_3$, we have*

$$P\left(\hat{\mathbb{E}}_n^{P,\dagger} = \mathbb{E}_n^P\right) \rightarrow 1. \tag{5.12}$$

6. Monte-Carlo simulation

In this section, we provide four simulated examples to examine the finite-sample numerical performance of the proposed high-dimensional time-varying VAR and network estimates. Throughout this section, we denote the proposed time-varying weighted group LASSO method as tv-wgLASSO and the time-varying CLIME method as tv-CLIME. We compare the performance of the tv-wgLASSO with the time-varying oracle estimator, denoted as tv-Oracle, which estimates only the true significant coefficient functions (assuming the zero coefficient functions were known), and the full time-varying estimator, denoted as tv-Full, which estimates all the coefficient functions without penalisation. We compare the performance of tv-CLIME with the time-varying graphical LASSO method, denoted as tv-GLASSO, which is implemented using the R package “glassoFast” on the VAR residuals. In addition, to investigate the loss of estimation accuracy due to the VAR model error approximation, we also report results from the infeasible tv-CLIME, which directly uses the true VAR errors (rather than residuals) in the estimation of precision matrices.

In the simulation, we use the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ with bandwidth $h = b = 0.75[\log(d)/n]^{1/5}$ as in Li et al. (2015). The bandwidth for the local PCA is set as $h_* = (2.35/\sqrt{12})[\sqrt{d}/n]^{1/5}$ as in Su and Wang (2017). The sample size n is 200 and 400, and the dimension d is 50 and 100. Although such dimensions are smaller than the sample size, when $n = 200$ and $d = 100$, the “effective sample size” used in each local linear estimation in (3.3) is approximately $2nh \approx 140$, which is smaller than the combined number of unknown coefficient functions and their derivatives, $2d = 200$, to be estimated from (3.3). Consequently, in this case we fail to implement the naive tv-Full estimation. There are three tuning parameters in the proposed estimation procedure: λ_1 in the first stage of preliminary time-varying LASSO estimation, λ_2 in the second stage of time-varying weighted group LASSO, and λ_3 in the third stage of time-varying CLIME. They are selected by the Bayesian information criterion (BIC), the generalised information criterion (GIC), and the extended Bayesian information criterion (EBIC), respectively. Appendix A in the supplement gives definitions of these information criteria.

To evaluate whether the time-varying network structure is accurately uncovered, we report the false positive (FP), the false negative (FN), the true positive rate (TPR), the true negative rate (TNR), the positive predictive value (PPV), the negative predictive value (NPV), the F1 score (F1), and the Matthews correlation coefficient (MCC). Definitions of these measures are provided in Appendix A of the supplement. To evaluate the performance of the functional coefficient estimators, we report the average R squared

(average R^2) of the regressions in (3.2), the average scaled Frobenius norm of estimation errors of functional coefficients (EE_A), and the root-mean-squared error of the VAR residuals towards the errors ($RMSE_e$). Taking our proposed tv-wgLASSO estimator for time-varying VAR(1) as an example, EE_A and $RMSE_e$ are defined, respectively, as

$$EE_A = \frac{1}{n\sqrt{d}} \sum_{i=1}^n \left\| \hat{A}_1(\tau_i) - A_1(\tau_i) \right\|_F \quad \text{and} \quad RMSE_e = \sqrt{\frac{1}{nd} \sum_{i=1}^d \sum_{i=1}^n (\hat{e}_{i,i} - e_{i,i})^2}. \tag{6.1}$$

To evaluate the performance of the precision matrix estimators, we report the average scaled Frobenius norm of estimation errors (EE_{Ω}), defined as

$$EE_{\Omega} = \frac{1}{n\sqrt{d}} \sum_{i=1}^n \left\| \hat{\Omega}(\tau_i) - \Omega(\tau_i) \right\|_F. \tag{6.2}$$

All the above measures are calculated for each Monte Carlo replication and then averaged over 100 replications. Except for exact values of 0's and 1's, the FP and FN values are rounded to 2 decimal places, while the others are rounded to 3 decimal places.

Example 1. The data is generated from a time-varying VAR(1) model with $A_1(\tau)$ being a diagonal matrix for all $\tau \in [0, 1]$. Each diagonal entry of $A_1(\tau)$ independently takes the value of either $0.64\Phi(5(\tau - 1/2))$ or $0.64 - 0.64\Phi(5(\tau - 1/2))$ with an equal probability of 0.5, where $\Phi(\cdot)$ is the standard normal distribution function. We set $\Omega(\tau)$ to be a block diagonal matrix: $\Omega(\tau) = \mathbf{I}_{d/2} \otimes \Omega_*(\tau)$, where $\Omega_*(\tau) = [\omega_{ij,*}(\tau)]_{2 \times 2}$ with $\omega_{11,*}(\tau) = \omega_{22,*}(\tau) \equiv 1$, and $\omega_{12,*}(\tau) = \omega_{21,*}(\tau) = 1.4\Phi(5(\tau - 1/2)) - 0.7$. The diagonal structure of $A_1(\tau)$ implies that no Granger causality exists between variables, whereas the block diagonal structure of $\Omega(\tau)$ results in weak cross-sectional dependence between the components of X_t .

Table 1 reports the estimation results of the time-varying transition matrices and Granger networks. For the proposed tv-wgLASSO, the FP and FN values are very small compared with d^2 , which is the total number of potential directed Granger causality linkages or entries of the transition matrix. This leads to large values of the TPR, TNR, PPV, NPV, F1 and MCC measures, all of which are close to 1. We can also see that the FP and FN values double when d increases from 50 to 100, but decrease substantially when n grows from 200 to 400. These results clearly show that tv-wgLASSO can accurately uncover the time-varying Granger network as long as the sample size is moderately large. The average R^2 of tv-wgLASSO is close to that of tv-Oracle, but the naive tv-Full method tends to have large R^2 due to model over-fitting. Although the EE_A values of tv-wgLASSO are larger than those of tv-Oracle when $n = 200$, they drop significantly and are even slightly smaller than those of tv-Oracle when $n = 400$. A similar pattern can be observed in $RMSE_e$, indicating that the proposed tv-wgLASSO is capable of providing good approximations to VAR errors, which are used in the subsequent time-varying precision matrix estimation. Unsurprisingly, the tv-Full method fails to estimate the time-varying transition matrix when $d = 100$ and $n = 200$.

Table 2 reports the estimation results for the time-varying precision matrices and partial correlation networks. When $n = 200$, both tv-CLIME and tv-GLASSO have zero FP values, whereas tv-CLIME has smaller FN than tv-GLASSO. Hence, the proposed tv-CLIME performs better than tv-GLASSO in terms of the F1 and MCC measures. When $n = 400$, both tv-CLIME and tv-GLASSO correctly uncover the time-varying partial correlation networks. In terms of the precision matrix estimation accuracy (EE_{Ω}), tv-GLASSO performs slightly better than tv-CLIME. In addition, by comparing the tv-CLIME and the infeasible tv-CLIME, we may conclude that the VAR error approximation has negligible impact on the precision matrix and partial correlation network estimation.

Example 2. The data is generated from a time-varying VAR(1) model with $A_1(\tau)$ being an upper triangular matrix for all $\tau \in [0, 1]$. Each diagonal entry of $A_1(\tau)$ takes the value of $0.7\Phi(5(\tau - 1/2))$, each super-diagonal entry takes the value of $0.7 - 0.7\Phi(5(\tau - 1/2))$, and the remaining entries take the value of 0. We set $\Omega(\tau) = [\omega_{ij}(\tau)]_{d \times d}$ to be a banded symmetric matrix for all $\tau \in [0, 1]$ with $\omega_{ii}(\tau) \equiv 1$, $\omega_{i,i+1}(\tau) = 0.7\Phi(5(\tau - 1/2)) - 0.7$, $\omega_{i,i+2}(\tau) = 0.7 - 0.7\Phi(5(\tau - 1/2))$, and $\omega_{i,j}(\tau) \equiv 0$ if $|i - j| > 2$.

Table 3 reports the estimation results for the time-varying transition matrices and Granger networks. Note that the time series variables in this example are more correlated to each other than those in Example 1, which affects the network estimation accuracy. When $d = 100$ and $n = 200$, the FP and FN values of tv-wgLASSO reach their maximum at 20.73 and 37.55, respectively, whereas the F1 and MCC values are around 0.85. As in Example 1, the F1 and MCC values increase when n increases from 200 to 400, and again the average R^2 of tv-wgLASSO is close to that of tv-Oracle. However, tv-wgLASSO has much larger EE_A and $RMSE_e$ than tv-Oracle.

Table 4 reports the estimation results for the time-varying precision matrices and partial correlation networks. It follows from the EE_A and $RMSE_e$ results in Table 3 that the VAR error approximation is poorer than that in Example 1. Consequently the proposed tv-CLIME performs worse than the infeasible tv-CLIME using the true VAR errors in the estimation. In particular, FN of the tv-CLIME is much larger than that of the infeasible tv-CLIME when $n = 200$. Due to the same reason, the infeasible tv-CLIME also outperforms the tv-GLASSO. In addition, we find that the tv-CLIME is better than the tv-GLASSO in uncovering the time-varying precision network when $n = 200$, and they perform equally well when $n = 400$.

Example 3. The data is generated from a VAR(1) model with $A_1(\tau) = [a_{ij}(\tau)]_{d \times d}$ being a Toeplitz matrix and $a_{ij}(\tau) = (0.4 - 0.1\tau)^{|i-j|+1}$. The precision matrix $\Omega(\tau) = [\omega_{ij}(\tau)]_{d \times d}$ is also a Toeplitz matrix with $\omega_{ij}(\tau) = (0.8 - 0.1\tau)^{|i-j|}$. In this example, both the transition and precision matrices are non-sparse so that we can examine how our proposed methods perform when the (exact) sparsity assumption fails.

As all the entries of the transition and precision matrices are nonzero, all possible linkages between any two nodes in the Granger and partial correlation networks exist. Hence, it is not as useful to calculate network discovery measures such as FP and TNR as

Table 1

Transition matrix and Granger network estimation in Example 1: a comparison of the methods of tv-wgLASSO, tv-Oracle, and tv-Full. The top panel presents results for Granger network estimation using measures of FP, FN, TPR, TNR, PPV, NPV, F1, and MCC*. The middle panel reports the average R^2 of the regressions in (3.2). The bottom panel reports the estimation error for the transition matrix $A_1(\cdot)$, EE_A , and the root-mean-squared error, $RMSE_e$, of the residuals for estimating VAR errors.

Measure	Dimension	tv-wgLASSO		tv-Oracle		tv-Full	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0.97	0.04	0	0	2450	2450
	$d = 100$	1.73	0.08	0	0	–	9900
FN	$d = 50$	3.53	0.08	0	0	0	0
	$d = 100$	8.55	0.15	0	0	–	0
TPR	$d = 50$	0.929	0.998	1	1	1	1
	$d = 100$	0.915	0.999	1	1	–	1
TNR	$d = 50$	1.000	1.000	1	1	0	0
	$d = 100$	1.000	1.000	1	1	–	0
PPV	$d = 50$	0.980	0.999	1	1	0.02	0.02
	$d = 100$	0.982	0.999	1	1	–	0.01
NPV	$d = 50$	0.999	1.000	1	1	1	1
	$d = 100$	0.999	1.000	1	1	–	1
F1	$d = 50$	0.953	0.999	1	1	0.039	0.039
	$d = 100$	0.947	0.999	1	1	–	0.020
MCC	$d = 50$	0.953	0.999	1	1	0	0
	$d = 100$	0.947	0.999	1	1	–	0
Average R^2	$d = 50$	0.289	0.296	0.296	0.297	0.933	0.721
	$d = 100$	0.296	0.306	0.305	0.307	–	0.959
EE_A	$d = 50$	0.214	0.160	0.185	0.163	54.29	1.410
	$d = 100$	0.224	0.163	0.189	0.166	–	112.8
$RMSE_e$	$d = 50$	0.203	0.115	0.162	0.120	1.119	0.876
	$d = 100$	0.213	0.113	0.159	0.119	–	1.145

* The FP, FN, TPR, TNR, PPV, NPV, F1, and MCC stand for, respectively, the false positive, the false negative, the true positive rate, the true negative rate, the positive predictive value, the negative predictive value, the F1 score, and the Matthews correlation coefficient, whose definitions can be found in Appendix A.

Table 2

Precision matrix and partial correlation network estimation in Example 1: a comparison of the methods of tv-CLIME, infeasible tv-CLIME, and tv-GLASSO. The top panel presents results for partial correlation network estimation using measures of FP, FN, TPR, TNR, PPV, NPV, F1, and MCC. The bottom panel reports the estimation error for the precision matrix $\Omega(\cdot)$, EE_Ω .

Measure	Dimension	tv-CLIME		infeasible tv-CLIME		tv-GLASSO	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0	0.02	0	0.02	0	0
	$d = 100$	0	0.03	0	0.01	0	0
FN	$d = 50$	5.06	0	3.49	0	9.24	0
	$d = 100$	13.25	0	9.01	0	28.31	0
TPR	$d = 50$	0.798	1	0.860	1	0.630	0
	$d = 100$	0.735	1	0.820	1	0.434	0
TNR	$d = 50$	1	1.000	1	1.000	1	1
	$d = 100$	1	1.000	1	1.000	1	1
PPV	$d = 50$	1	0.999	1	0.999	1	1
	$d = 100$	1	0.999	1	1.000	1	1
NPV	$d = 50$	0.996	1	0.097	1	0.992	1
	$d = 100$	0.997	1	0.998	1	0.994	1
F1	$d = 50$	0.884	1.000	0.922	1.000	0.768	1
	$d = 100$	0.845	1.000	0.899	1.000	0.600	1
MCC	$d = 50$	0.889	1.000	0.925	1.000	0.788	1
	$d = 100$	0.855	1.000	0.904	1.000	0.653	1
EE_Ω	$d = 50$	0.510	0.436	0.503	0.435	0.451	0.407
	$d = 100$	0.481	0.421	0.473	0.419	0.433	0.397

in Examples 1 and 2. Instead, we report only the Average R^2 , EE_A , $RMSE_e$, and EE_Ω for this example in Table 5. The tv-Oracle is equivalent to tv-Full and both have a large number of parameters to be estimated relative to the effective sample size in the local linear estimation (see (3.3)) for the time-varying transition matrices, especially when $d = 100$ and $n = 200$. Consequently, the EE_A and $RMSE_e$ of the tv-wgLASSO are much smaller than those of the tv-Oracle. The EE_Ω results of the tv-CLIME are very close to those of the infeasible tv-CLIME, suggesting that the VAR error approximation has little impact on the tv-CLIME performance as discussed in Example 1. In addition, the EE_Ω results of the tv-CLIME and infeasible tv-CLIME are generally close to those of tv-GLASSO. This simulation shows that the proposed tv-wgLASSO and tv-CLIME perform reasonably well when the sparsity assumptions on transition and precision matrices are not satisfied.

Table 3

Transition matrix and Granger network estimation in Example 2: a comparison of the methods of tv-wgLASSO, tv-Oracle, and tv-Full. The top panel presents results for Granger network estimation using measures of FP, FN, TPR, TNR, PPV, NPV, F1, and MCC. The middle panel reports the average R^2 of the regressions in (3.2). The bottom panel reports the estimation error for the transition matrix $A_1(\cdot)$, EE_A , and the root-mean-squared error, $RMSE_e$, of the residuals for estimating VAR errors.

Measure	Dimension	tv-wgLASSO		tv-Oracle		tv-Full	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	13.53	12.75	0	0	2401	2401
	$d = 100$	20.73	7.73	0	0	–	9801
FN	$d = 50$	18.56	11.11	0	0	0	0
	$d = 100$	37.55	13.90	0	0	–	0
TPR	$d = 50$	0.813	0.888	1	1	1	1
	$d = 100$	0.811	0.930	1	1	–	1
TNR	$d = 50$	0.994	0.995	1	1	0	0
	$d = 100$	0.998	0.999	1	1	–	0
PPV	$d = 50$	0.859	0.875	1	1	0.040	0.040
	$d = 100$	0.888	0.960	1	1	–	0.020
NPV	$d = 50$	0.992	0.995	1	1	0	0
	$d = 100$	0.996	0.999	1	1	–	0
F1	$d = 50$	0.834	0.881	1	1	0.076	0.076
	$d = 100$	0.847	0.945	1	1	–	0.039
MCC	$d = 50$	0.828	0.876	1	1	0	0
	$d = 100$	0.846	0.943	1	1	–	0
Average R^2	$d = 50$	0.465	0.448	0.477	0.462	0.963	0.829
	$d = 100$	0.473	0.467	0.483	0.471	–	0.978
EE_A	$d = 50$	0.328	0.250	0.171	0.122	58.44	1.510
	$d = 100$	0.323	0.204	0.168	0.122	–	82.60
$RMSE_e$	$d = 50$	0.631	0.476	0.417	0.305	1.673	1.414
	$d = 100$	0.613	0.390	0.414	0.309	–	1.720

Table 4

Precision matrix and partial correlation network estimation in Example 2: a comparison of the methods of tv-CLIME, infeasible tv-CLIME, and tv-GLASSO. The top panel presents results for partial correlation network estimation using measures of FP, FN, TPR, TNR, PPV, NPV, F1, and MCC. The bottom panel reports the estimation error for the precision matrix $\Omega(\cdot)$, EE_{Ω} .

Measure	Dimension	tv-CLIME		infeasible tv-CLIME		tv-GLASSO	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0.03	0.04	0.02	0.03	0	0.01
	$d = 100$	0.01	0	0	0.01	0	0.01
FN	$d = 50$	12.62	0.82	2.34	0	20.84	0.06
	$d = 100$	24.71	0.23	6.21	0.01	49.73	0.43
TPR	$d = 50$	0.742	0.983	0.952	1	0.575	0.997
	$d = 100$	0.750	0.998	0.937	1.000	0.498	0.996
TNR	$d = 50$	1.000	1.000	1.000	1.000	1	1.000
	$d = 100$	1.000	1	1	1.000	1	1.000
PPV	$d = 50$	0.999	0.999	1.000	0.999	1	1.000
	$d = 100$	1.000	1	1	1.000	1	1.000
NPV	$d = 50$	0.989	0.999	0.998	1	0.983	1.000
	$d = 100$	0.995	1.000	0.999	1.000	0.990	1.000
F1	$d = 50$	0.850	0.991	0.975	1.000	0.725	0.998
	$d = 100$	0.857	0.999	0.967	1.000	0.662	0.998
MCC	$d = 50$	0.856	0.991	0.975	1.000	0.749	0.998
	$d = 100$	0.864	0.999	0.967	1.000	0.701	0.998
EE_{Ω}	$d = 50$	0.598	0.533	0.526	0.485	0.560	0.514
	$d = 100$	0.560	0.489	0.486	0.458	0.536	0.496

Example 4. The data is generated from a factor-adjusted time-varying VAR model in the form of (5.2). The idiosyncratic errors X_t are generated from the VAR(1) model in Example 2. There are two factors, i.e., $F_t = (F_{t,1}, F_{t,2})^T$, generated from two univariate AR(1) processes: $F_{t,1} = 0.6F_{t-1,1} + \sqrt{1 - 0.6^2}u_{t,1}^F$ and $F_{t,2} = 0.3F_{t-1,2} + \sqrt{1 - 0.3^2}u_{t,2}^F$, where $u_{t,1}^F$ and $u_{t,2}^F$ are independently drawn from a standard normal distribution. The $d \times 2$ factor loading matrix is generated as $A_t = (A_{t,1}, A_{t,2})$, where $A_{t,1} \equiv A_1$ is a time-invariant vector drawn from a standard d -dimensional normal distribution and $A_{t,2} = (A_{1t,2}, \dots, A_{dt,2})^T$ with $A_{it,2} = 2/(1 + \exp\{-2[10(t/n) - 5(i/d) - 2]\})$ for $i = 1, \dots, d$.

As the error vectors X_t in this factor-adjusted VAR model are generated from the same DGP as Example 2, in which we have compared the different methods, we report results only for our tv-wgLASSO and tv-CLIME to assess how the initial factor estimation affects their estimation accuracy. Table 6 reports the results for the time-varying transition matrices and Granger networks for the idiosyncratic errors, and the time-varying precision matrices and partial correlation networks. Comparing results with Tables 3 and 4

Table 5

Transition and precision matrix estimation in Example 3. The top panel presents the average R^2 , EE_A , and $RMSE_e$ from the three methods of tv-wgLASSO, tv-Oracle, and tv-Full for the estimation of the transition matrix. The bottom panel presents the estimation error, EE_{Ω} , for the precision matrix from the methods of tv-CLIME, infeasible tv-CLIME, and tv-GLASSO.

Measure	Dimension	tv-wgLASSO		tv-Oracle		tv-Full	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
Average R^2	$d = 50$	0.009	0.029	0.891	0.588	0.891	0.588
	$d = 100$	0.005	0.020	–	0.930	–	0.930
EE_A	$d = 50$	0.383	0.348	56.66	1.927	56.66	1.927
	$d = 100$	0.388	0.364	–	97.60	–	97.60
$RMSE_e$	$d = 50$	0.515	0.463	1.716	1.300	1.716	1.300
	$d = 100$	0.523	0.486	–	1.776	–	1.776
		tv-CLIME		infeasible tv-CLIME		tv-GLASSO	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
EE_{Ω}	$d = 50$	1.669	1.601	1.613	1.572	1.584	1.570
	$d = 100$	1.674	1.615	1.616	1.580	1.587	1.588

Table 6

Factor-adjusted transition matrix and Granger network estimation using tv-wgLASSO and factor-adjusted precision matrix and partial correlation network estimation using tv-CLIME in Example 4. The top panel presents results for Granger network and partial correlation network estimation using measures of FP, FN, TPR, TNR, PPV, NPV, F1, and MCC. The middle panel reports the average R^2 of the regressions in (3.2) estimated using tv-wgLASSO. The bottom panel reports the estimation error for the factor-adjusted transition matrix $A_1(\cdot)$, EE_A , the root-mean-squared error, $RMSE_e$, of the factor-adjusted VAR residuals for estimating true errors, and the estimation error for the factor-adjusted precision matrix $\Omega(\cdot)$, EE_{Ω} .

Measure	Dimension	tv-wgLASSO		tv-CLIME	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	11.35	10.60	0.01	0.01
	$d = 100$	20.40	10.41	0	0.02
FN	$d = 50$	35.97	14.77	38.22	5.36
	$d = 100$	65.45	20.68	65.99	2.21
TPR	$d = 50$	0.637	0.851	0.220	0.891
	$d = 100$	0.671	0.896	0.333	0.978
TNR	$d = 50$	0.995	0.996	1.000	1.000
	$d = 100$	0.998	0.999	1	1.000
PPV	$d = 50$	0.852	0.890	0.999	1.000
	$d = 100$	0.869	0.945	1	1.000
NPV	$d = 50$	0.985	0.994	0.969	0.995
	$d = 100$	0.993	0.998	0.987	1.000
F1	$d = 50$	0.725	0.869	0.349	0.941
	$d = 100$	0.756	0.920	0.496	0.989
MCC	$d = 50$	0.725	0.865	0.448	0.941
	$d = 100$	0.759	0.919	0.570	0.988
Average R^2	$d = 50$	0.298	0.350	–	–
	$d = 100$	0.339	0.389	–	–
EE_A	$d = 50$	0.413	0.283	–	–
	$d = 100$	0.396	0.241	–	–
$RMSE_e$	$d = 50$	1.319	1.025	–	–
	$d = 100$	1.230	0.856	–	–
EE_{Ω}	$d = 50$	–	–	0.670	0.585
	$d = 100$	–	–	0.628	0.534

for Example 2, we can observe that the factor-adjusted estimation introduces additional estimation errors, leading to smaller values of F1 and MCC. The impact is more marked when $n = 200$ but reduces substantially when $n = 400$. As in the previous examples, the F1 and MCC values increase when n increases from 200 to 400. Thus we may conclude that, although the factor model estimation errors are passed on to the subsequent three-stage estimation procedure, their impact on the estimation of networks is not significant when the sample size is moderately large ($n = 400$).

7. Application

In this section, we employ the proposed methods to estimate the Granger causality and partial correlation networks for variables from the FRED-MD macroeconomic dataset. Our method holds particular relevance for economists and decision-makers in three aspects: (i) it provides insight into the dynamic relationships among a large number of economic variables in terms of which variables would change due to a past change in a specific variable and by how much. This information will enable policy makers to make more effective policies by taking a broader view of how economic variables interact dynamically; (ii) it provides information about the

contemporary pairwise dependencies among a large set of variables in which the effects of all other variables have been controlled for; (iii) it improves our understanding about how the dynamic relationships and contemporary dependencies among economic variables evolve over time, which may aid timely policy adjustment in response to changing economic conditions.

Accessible from the Fred-MD website,¹ the dataset comprises 127 U.S. macroeconomic variables observed monthly from January 1959, making it an important data source for both academic and policy studies. These macroeconomic variables are classified into eight groups: consumption, orders and inventories; housing; interest and exchange rates; labour market; money and credit; output and income; prices; and the stock market. More detailed description can be found in [McCracken and Ng \(2016\)](#).

7.1. Estimation of Granger causality and partial correlation networks

Our sample spans January 1959–July 2022 with a total of 763 observations. We follow [McCracken and Ng \(2016, 2020\)](#) to remove outliers, fill missing values, and standardise each variable so that they have zero mean and unit variance. We consider the two factor modelling methods in Section 5 to accommodate strong cross-sectional dependence: the approximate factor model (5.1) with constant factor loadings and the time-varying factor model (5.2) with dynamic factor loadings. The information criteria proposed by [Bai and Ng \(2002\)](#) and [Su and Wang \(2017\)](#) are used to determine the number of factors in these two models (see Appendix A in the supplement for a description of the criteria). Seven factors are selected for the factor model with constant loadings, whereas only four are selected for the time-varying factor model. Since the latter provides a more parsimonious model specification, we hereafter report network estimation results only for this model. The estimated idiosyncratic errors from the factor model, denoted as $\hat{\varepsilon}_{t,i}$, $i = 1, \dots, 127$, $t = 1, \dots, 763$, are then used for our VAR modelling and network analysis. [Miao et al. \(2023\)](#) propose to determine the optimal order of a high-dimensional VAR model via a ratio criterion, which compares the Frobenius norms of the estimated transition matrices over different lags. We extend their criterion to our time-varying VAR model setting (see Appendix A in the supplement for detail) and subsequently select the time-varying VAR(1) model.

[Fig. 1](#) plots the estimated Granger networks from the factor-adjusted static VAR(1) and time-varying VAR(1) models. From the estimated time-varying transition matrices, we uncover 190 directed linkages in the Granger causality network, among which 78 are self-linkages and 143 are linkages within the same category. In particular, the self-linkages, which correspond to the significant diagonal entries of the transition matrices, indicate that the macroeconomic variables in the following four categories: consumption, orders and inventories; interest and exchange rates; money and credit; and prices, are more persistent than the others, even though all the variables have been transformed into stationary series in the preliminary analysis. By contrast, we find 155 directed linkages for the Granger network estimated via static VAR(1) and hence, our time-varying VAR(1) model is able to uncover more linkages in the network, some existing at only some time points but not all. If no factor-adjustment is undertaken, unsurprisingly more linkages in the Granger causality network are uncovered (see Appendix A in the supplement for detail).

We further explore the smooth dynamic evolvement of the VAR(1) coefficients and Granger causality network. Taking the logarithmic growth rate of S&P PE ratio (S&P PE ratio)² as an example, there are four directed linkages to this variable from the acceleration of the logarithmic monetary base (BOGMBASE), the logarithmic return of S&P 500 index (S&P 500), the logarithmic return of S&P 500 industrials index (S&P: indust), and the S&P PE ratio (which gives a self-linkage). We re-estimate the corresponding functional coefficients using the autoregressive model with only the four predictors, and draw their 90% confidence bands using the R package “tvReg”. [Fig. 2](#) shows the estimated curves and confidence bands. We can find that S&P PE ratio is persistent and in general positively correlated with BOGMBASE in the preceding month. The estimated time-varying coefficient of S&P: indust is close to zero and significant at some time periods. It is thus unsurprising that the static VAR(1) model with classic LASSO penalisation does not detect the Granger causality linkage from this variable. In fact, LASSO tends to select only one variable in a group of highly-correlated predictors. Due to high correlation between S&P: indust and S&P 500, only the linkage from S&P 500 is selected in the static VAR(1) based network. By contrast, the proposed preliminary time-varying LASSO selects both series at different time periods, and the second-stage weighted group LASSO aggregates the information over time and selects both series.

We plot the estimated partial correlation networks in [Fig. 3](#). It can be seen that the partial correlation networks are more sparse. Using the factor-adjusted time-varying CLIME, 234 undirected linkages are detected, among which 205 are linkages within the same category. In contrast, the estimated network without factor adjustment contains 236 linkages with 211 within the same category. Unlike the Granger network, it seems that whether to make factor adjustment has little impact on the resulting partial correlation network.

To examine the time-varying pattern of the partial correlation network, we consider, as an example, the partial correlation linkages between S&P PE ratio and the four variables: S&P 500, S&P: indust, the increment of S&P composite common stock: dividend yield (S&P div yield), and the spread between Moody’s seasoned baa corporate bond and effective federal funds rate (BAAFFM). We re-estimate the relevant elements in the time-varying precision matrix with a 200-month moving window ([Jankova and van de Geer, 2015](#)), and draw 90% confidence bands using R package “SILGGM”. These are presented in [Fig. 4](#). Note that each partial correlation has the opposite sign to the corresponding entry in the precision matrix. Hence, [Fig. 4](#) suggests that S&P PE ratio is positively (partially) correlated with S&P 500 and S&P: indust, whilst negatively (partially) correlated with S&P div yield. Partial correlation linkage with BAAFFM is insignificant in most time periods except the years between 1995–2010. The time evolving patterns observed in [Fig. 4](#) suggest that a time-varying model can better describe the network structure among the FRED-MD variables.

¹ <https://research.stlouisfed.org/econ/mccracken/fred-databases/>

² We show in the parentheses the variable names used in the FRED-MD dataset. Variable transformation is conducted following the guideline for the dataset.

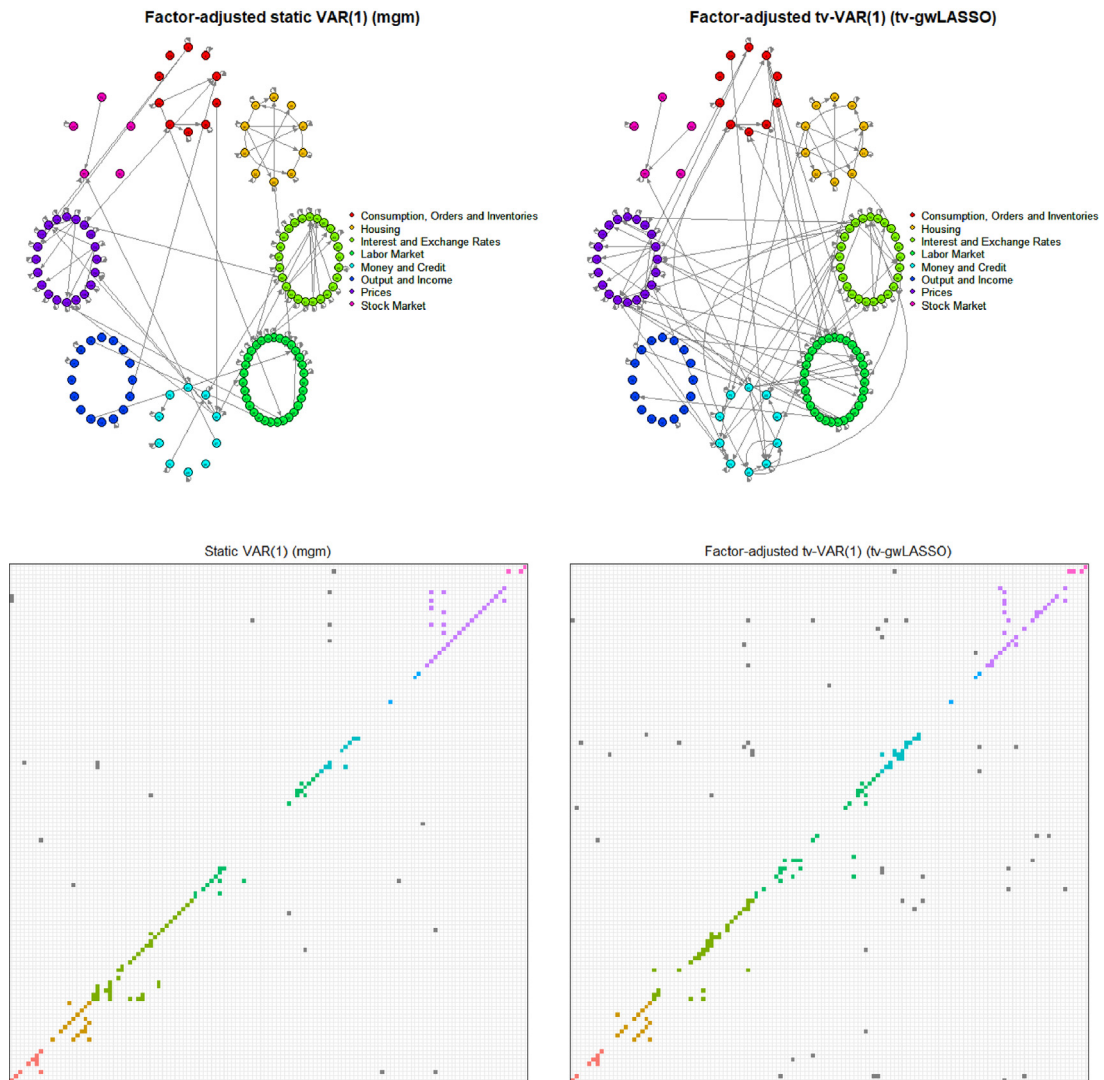


Fig. 1. Estimated Granger causality networks using the factor-adjusted static VAR(1) model (top left) and time-varying VAR(1) model (top right) along with their corresponding adjacency matrices (bottom left and right). Each node in the network represents one of the 127 U.S. macroeconomic variables, which are categorised into eight groups shown as the eight circles in different colours. In the adjacency matrices, linkages within each group are colour-coded to match the group colour, while linkages across groups are shown in grey.

7.2. Analysis of time-varying networks for pre- and post-global financial crisis

In accordance with the sample period and break location identified by Duan et al. (2023), we examine the sub-sample period spanning December 2001–January 2013, which is further split into the pre-global financial crisis (pre-GFC) period, spanning December 2001–July 2007, and the post-global financial crisis (post-GFC) period, spanning August 2007–January 2013. The number of observations for the two periods are 68 and 66, respectively.

Table 7 highlights the five most substantial changes in the coefficients of the transition matrix of the time-varying VAR(1) model before and after the global financial crisis (GFC), as measured by the difference in the rescaled mean coefficients (RMCs) in the two periods, where for a sub-period $[t_1, t_2]$ of $[1, n]$, the RMC of an estimated coefficient $\hat{\alpha}_{ij}^\dagger(\cdot)$ is defined as

$$RMC(\hat{\alpha}_{ij}^\dagger)_{[t_1, t_2]} = \frac{1}{\max_{1 \leq t \leq n} |\hat{\alpha}_{ij}^\dagger(\tau_t)|} \cdot \frac{1}{(t_2 - t_1 + 1)} \sum_{\tau=t_1}^{t_2} \hat{\alpha}_{ij}^\dagger(\tau_t).$$

We see the biggest change (i.e., increase) in the sensitivity of AAA bond yields (AAA) to changes in the effective federal funds rate (FEDFUNDS) after GFC. The second biggest change is in the autoregressive coefficient of housing starts in the northeastern U.S. (HOUSTNE), showing a big drop in the persistency of the variable after GFC. Table 7 also reveals a drop to negative value

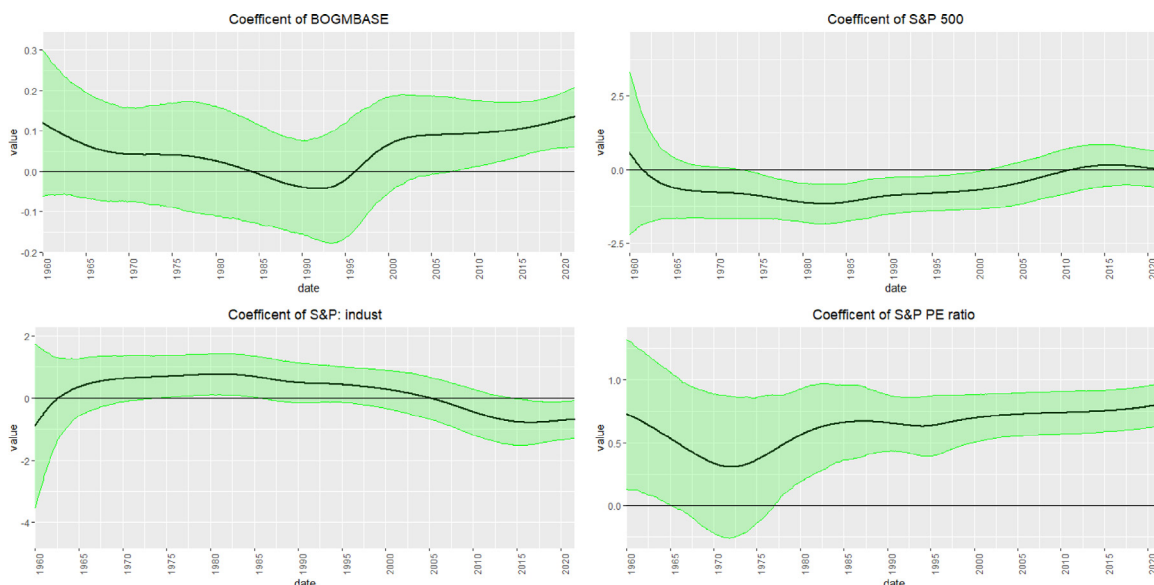


Fig. 2. Estimated time-varying VAR(1) coefficients linked to S&P PE ratio with 90% confident bands.

Table 7

The five coefficients in the transition matrix of the factor-adjusted time-varying VAR(1) model that have undergone the most substantial changes from the pre-GFC (December 2001–July 2007) period to the post-GFC (August 2007–January 2013) period, as measured by the absolute difference between the rescaled mean coefficients (RMCs) in the two periods. Each coefficient corresponds to a directed linkage in the Granger causality network from each of the variables in the “From” column to the variables in the “To” column.

Rank	From	To	RMC (Pre-GFC)	RMC (Post-GFC)	Difference
1	FEDFUNDS	AAA	0.120	0.603	0.482
2	HOUSTNE	HOUSTNE	0.618	0.194	−0.424
3	CES3000000008	CMRMTSPLx	0.034	−0.361	−0.395
4	M1SL	TOTRESNS	0.948	0.600	−0.348
5	BOGMBASE	NONBORRES	0.355	0.014	−0.340

Table 8

The five pairwise partial correlations that have undergone the most substantial changes from the pre-GFC period (December 2001 – July 2007) to the post-GFC (August 2007 – January 2013) period, as measured by the absolute difference between the rescaled mean coefficients (RMCs) of the partial correlations in the two periods. Each partial correlation corresponds to an undirected linkage in the partial correlation network between “Variable 1” and “Variable 2”.

Rank	Variable 1	Variable 2	RMC (Pre-GFC)	RMC (Post-GFC)	Difference
1	CUMFNS	T1YFFM	−0.601	−0.012	0.589
2	INDPRO	IPBUSEQ	0.595	0.019	−0.576
3	TB6MS	TB6SMFF	0.617	0.084	−0.533
4	CUSR0000SA0L5	PCEPI	0.698	0.204	−0.494
5	CPIULFSL	PCEPI	0.646	0.160	−0.486

in the relationship between labour earnings in the manufacturing sector (CES3000000008) and the subsequent sales performance of manufacturing and trade industries (CMRMTSPLx). The last two rows of Table 7 show the diminishing effects of M1 money stock (M1SL) on the subsequent total reserves of depository institutions (TOTRESNS) and of the monetary base (BOGMBASE) on the subsequent non-borrowed reserves of depository institutions (NONBORRES) after GFC.

We also consider changes in the pairwise partial correlations between the FRED-MD variables before and after GFC. Table 8 highlights the five most substantial changes in the partial correlations, as measured by the absolute difference in the RMCs of the partial correlations in the pre-GFC and post-GFC periods. Given the relation between partial correlations and elements of the precision matrix (i.e., $\rho_{ij} = -\omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}}$), the RMC of the estimated (i,j)th partial correlation $\hat{\rho}_{ij}^{\dagger}(\cdot)$ in a sub-period $[t_1, t_2]$ of $[1, n]$, is calculated as

$$RMC(\hat{\rho}_{ij}^{\dagger})_{[t_1, t_2]} = \frac{1}{\max_{1 \leq t \leq n} \left[\left| \hat{\omega}_{ij}^{\dagger}(\tau_t) \right| / \sqrt{\hat{\omega}_{ii}^{\dagger}(\tau_t) \hat{\omega}_{jj}^{\dagger}(\tau_t)} \right]} \cdot \frac{1}{(t_2 - t_1 + 1)} \sum_{t=t_1}^{t_2} \frac{-\hat{\omega}_{ij}^{\dagger}(\tau_t)}{\sqrt{\hat{\omega}_{ii}^{\dagger}(\tau_t) \hat{\omega}_{jj}^{\dagger}(\tau_t)}}$$

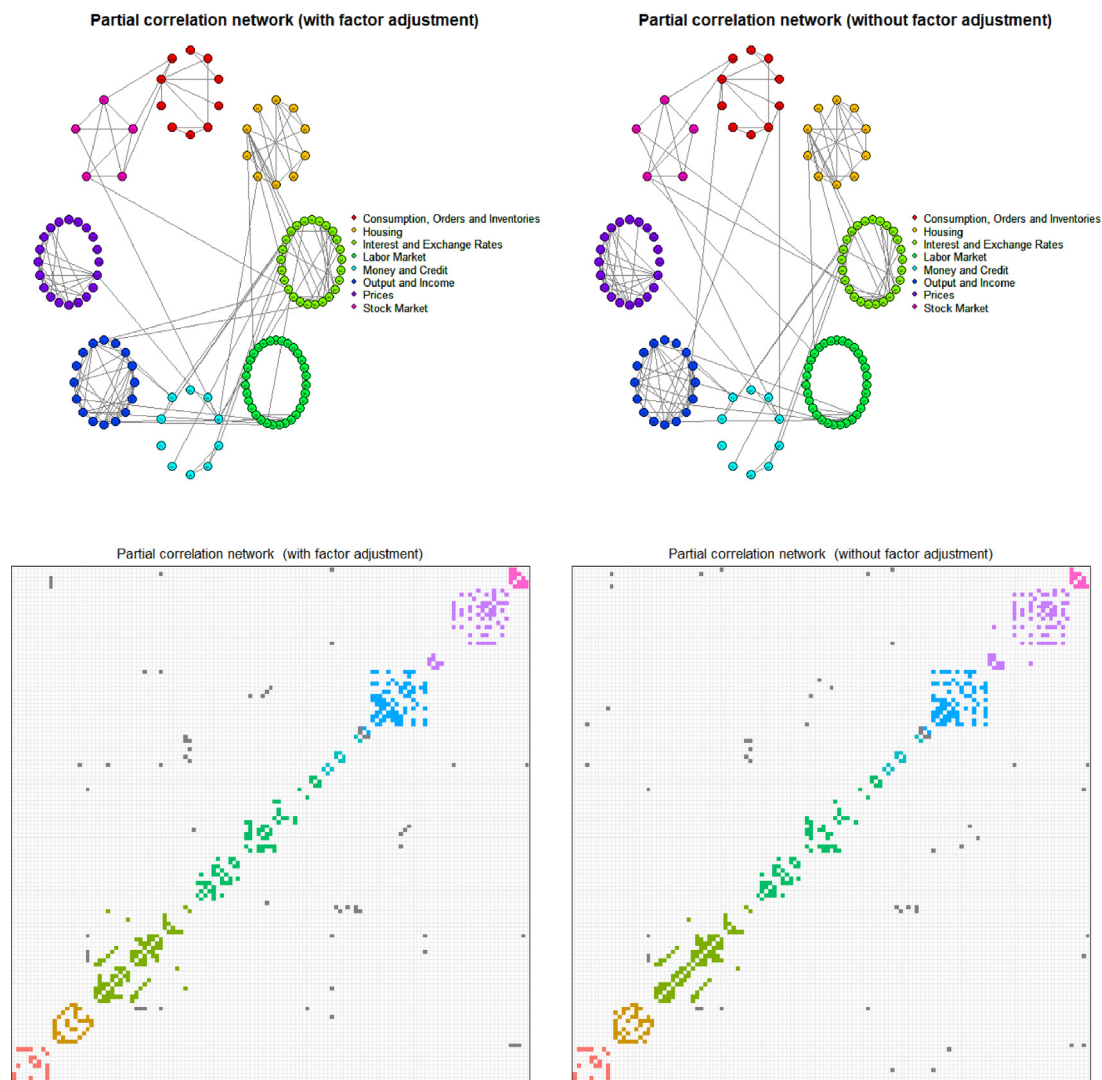


Fig. 3. Estimated partial correlation networks with (top left) and without (top right) factor adjustment along with their corresponding adjacency matrices (bottom left and right). Each node in the network represents one of the 127 U.S. macroeconomic variables, which are categorised into eight groups shown as the eight circles in different colours. In the adjacency matrices, linkages within each group are colour-coded to match the group colour, while linkages across groups are shown in grey.

The biggest change is observed in the negative partial correlation between capacity utilisation in manufacturing (CUMFNS) and the spread of 1-year treasury bond (T1YFFM), which becomes much weaker after GFC. The next four biggest changes are in the positive partial correlations between the industrial production index (IP) and the specific segment of industrial production related to business equipment (IPBUSEQ), the 6-month treasury bill rate (TB6MS) and the spread of 6-month treasury bill (TB6SMFFM), the consumer price index for all items less medical care (CUSR000SA0L5) and the personal consumption expenditure price index (PCEPI), and the consumer price index for all items less food (CPIULFSL) and the personal consumption expenditure price index (PCEPI), all becoming weaker after GFC although still being positive.

8. Conclusion

In this paper we consider a general time-varying VAR model for high-dimensional locally stationary time series, which allows for smooth structural changes over time. A three-stage estimation procedure combining time-varying LASSO, weighted group LASSO and time-varying CLIME is developed to estimate both the transition and error precision matrices. The estimated transition and precision matrices are then used to construct networks with directed Granger causality linkages and undirected partial correlation

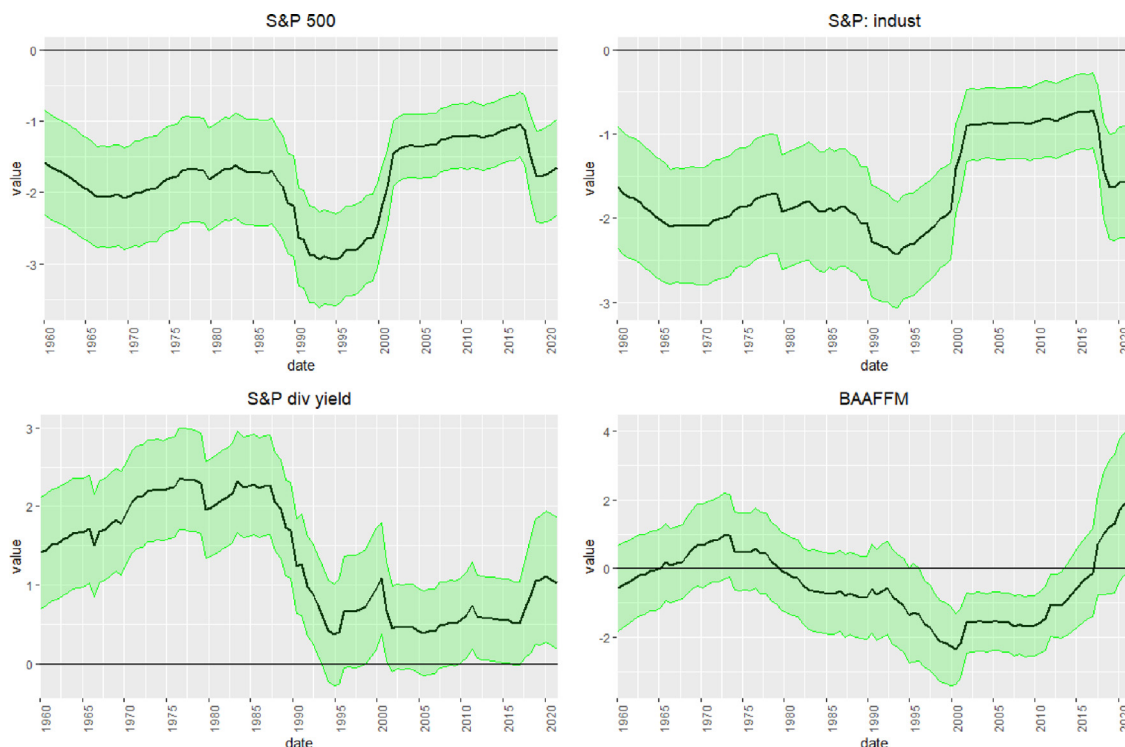


Fig. 4. Estimated time-varying elements in the precision matrix linked to S&P PE ratio with 90% confident bands..

linkages, respectively. Under the sparse structural assumption and other technical conditions, we derive the uniform consistency and oracle properties of the proposed estimators. For large-scale time series with high cross-sectional dependence, which are likely to violate the sparsity assumption, we extend our methodology and theory to a more general factor-adjusted time-varying VAR and networks. The simulation study shows that the developed methodology has reliable finite-sample performance. The empirical application showcases the potential of our network model and method in enhancing understanding of the time evolving relationships among economic variables on a large scale.

Acknowledgments

We would like to thank the editor, a guest associate editor, and two anonymous referees for valuable comments and suggestions. J. Chen was partly supported by the Economic and Social Research Council in the UK (ES/T01573X/1). D. Li was partly supported by the Leverhulme Research Fellowship, UK (RF-2023-396).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jeconom.2024.105941>.

References

- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 90, 191–221.
- Barigozzi, M., Brownlees, C., 2019. NETS: Network estimation for time series. *J. Appl. Econometrics* 34, 347–364.
- Barigozzi, M., Cho, H., Owens, D., 2024. FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series. *J. Bus. Econom. Statist.* 42, 890–902.
- Barigozzi, M., Hallin, M., Soccorsi, S., von Sachs, R., 2021. Time-varying general dynamic factor models and the measurement of financial connectedness. *J. Econometrics* 222, 324–343.
- Basu, S., Michailidis, G., 2015. Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* 43, 1535–1567.
- Basu, S., Shojaie, A., Michailidis, G., 2015. Network granger causality with inherent grouping structure. *J. Mach. Learn. Res.* 16, 417–453.
- Bickel, P., Ritov, Y., Tsybakov, A., 2009. Simultaneous analysis of lasso and dantzig selector. *Ann. Statist.* 37, 1705–1732.
- Burt, R.S., Kilduff, M., Tasselli, S., 2013. Social network analysis: foundations and frontiers on advantage. *Annu. Rev. Psychol.* 64, 527–547.
- Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* 136, 163–188.
- Cai, T.T., Liu, W., Luo, X., 2011. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* 106, 594–607.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean–variance analysis in large asset markets. *Econometrica* 51, 1305–1324.
- Chen, E., Fan, J., Zhu, X., 2023. Community network autoregression for high-dimensional time series. *J. Econometrics* 235, 1239–1256.

- Chen, J., Li, D., Wei, L., Zhang, W., 2021. Nonparametric homogeneity pursuit in functional-coefficient models. *J. Nonparametr. Stat.* 33, 387–416.
- Cheng, M., Honda, T., Li, J., Peng, H., 2014. Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *Ann. Statist.* 42, 1819–1849.
- Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. *Ann. Statist.* 25, 1–37.
- Dahlhaus, R., Subba Rao, S., 2006. Statistical inference for time-varying ARCH processes. *Ann. Statist.* 34, 1075–1114.
- Davis, R., Zang, P., Zheng, T., 2016. Sparse vector autoregressive modeling. *J. Comput. Graph. Statist.* 25, 1077–1096.
- Dempster, A.P., 1972. Covariance selection. *Biometrics* 28, 157–175.
- Diebold, F., Yilmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econometrics* 182, 119–134.
- Diebold, F., Yilmaz, K., 2015. *Financial and Macroeconomic Connectedness: A Network Approach To Measurement and Monitoring*. Oxford University Press.
- Ding, X., Qiu, Z., Chen, X., 2017. Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electron. J. Stat.* 11, 3871–3902.
- Duan, J., Bai, J., Han, X., 2023. Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *J. Econometrics* 233, 209–236.
- Eichler, M., Motta, G., von Sachs, R., 2011. Fitting dynamic factor models to non-stationary time series. *J. Econometrics* 163, 51–70.
- Fan, J., Feng, Y., Wu, Y., 2009. Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* 3, 521–541.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall.
- Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96, 1348–1360.
- Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussion). *J. R. Stat. Soc. Ser. B* 75, 603–680.
- Fan, J., Ma, Y., Dai, W., 2014a. Nonparametric independence screening in sparse ultra-high dimensional varying coefficient models. *J. Amer. Statist. Assoc.* 109, 1270–1284.
- Fan, J., Masini, R., Medeiros, M., 2023. Bridging factor and sparse models. *Ann. Statist.* 51, 1692–1717.
- Fan, J., Xue, L., Zou, H., 2014b. Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* 42, 819–849.
- Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Hafner, C., Linton, O., 2010. Efficient estimation of a multivariate multiplicative volatility model. *J. Econometrics* 159, 55–73.
- Han, F., Lu, H., Liu, H., 2015. A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* 16, 3115–3150.
- Hautsch, N., Schaumburg, J., Schienle, M., 2014. Forecasting systemic impact in financial networks. *Int. J. Forecast.* 30, 781–794.
- Jankova, J., van de Geer, S., 2015. Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* 9, 1205–1229.
- Kock, A.B., Callot, L., 2015. Oracle inequalities for high dimensional vector autoregressions. *J. Econometrics* 186, 325–344.
- Kolar, M., Song, L., Ahmed, A., Xing, E., 2010. Estimating time-varying networks. *Ann. Appl. Stat.* 4, 94–123.
- Koo, B., Linton, O., 2012. Estimation of semiparametric locally stationary diffusion models. *J. Econometrics* 170, 210–233.
- Krampe, J., Margaritella, L., 2022. Factor models with sparse VAR idiosyncratic components. Working paper available at <https://arxiv.org/pdf/2112.07149.pdf>.
- Lam, C., Fan, J., 2009. Sparsity and rates of convergence in large covariance matrix estimation. *Ann. Statist.* 37, 4254–4278.
- Li, D., Chen, J., Gao, J., 2011. Nonparametric time-varying coefficient panel data models with fixed effects. *Econom. J.* 14, 387–408.
- Li, D., Ke, Y., Zhang, W., 2015. Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. *Ann. Statist.* 43, 2676–2705.
- Lian, H., 2012. Variable selection for high-dimensional generalized varying-coefficient models. *Statist. Sinica* 22, 1563–1588.
- Liu, J., Li, R., Wu, R., 2014. Feature selection for varying coefficient models with ultrahigh dimensional covariates. *J. Amer. Statist. Assoc.* 109, 266–274.
- Liu, L., Zhang, D., 2021. Robust estimation of high-dimensional vector autoregressive models. Working paper available at <https://arxiv.org/abs/2109.10354>.
- Loh, P., Wainwright, M., 2013. Structural estimation for discrete graphical models: Generalized covariance matrices and their inverse. *Ann. Statist.* 41, 3022–3049.
- Lütkepohl, H., 2006. *New Introduction to Multiple Time Series Analysis*. Springer.
- McCracken, M.W., Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. *J. Bus. Econom. Statist.* 34, 574–589.
- McCracken, M.W., Ng, S., 2020. FRED-QD: A quarterly database for macroeconomic research. Working paper available at <https://www.nber.org/papers/w26872>.
- Miao, K., Phillips, P.C.B., Su, L., 2023. High-dimensional VARs with common factors. *J. Econometrics* 233, 155–183.
- Motta, G., Hafner, C., von Sachs, R., 2011. Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* 27, 1279–1319.
- Newman, M.E.J., 2002. Spread of epidemic disease on networks. *Phys. Rev. Ser. E* 66, 016128.
- Safikhani, A., Shojaie, A., 2022. Joint structural break detection and parameter estimation in high-dimensional non-stationary VAR models. *J. Amer. Statist. Assoc.* 117, 251–264.
- Scott, J., 2017. *Social Network Analysis*, fourth ed. Sage, London.
- Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* 97, 1167–1179.
- Su, L., Wang, X., 2017. On time-varying factor models: estimation and testing. *J. Econometrics* 198, 84–101.
- Tibshirani, R.J., 1996. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B* 58, 267–288.
- Vogt, M., 2012. Nonparametric regression for locally stationary time series. *Ann. Statist.* 40, 2601–2633.
- Wainwright, M.J., 2019. High-dimensional statistics: A non-asymptotic viewpoint. In: *Cambridge Series in Statistical and Probabilistic Mathematics*.
- Wang, L., Li, H., Huang, J., 2008. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.* 103, 1556–1569.
- Wang, H., Xia, Y., 2009. Shrinkage estimation of the varying-coefficient model. *J. Amer. Statist. Assoc.* 104, 747–757.
- Wang, D., Yu, Y., Rinaldo, A., 2021. Optimal change point detection and localization in sparse dynamic networks. *Ann. Statist.* 49, 203–232.
- Xu, M., Chen, X., Wu, W., 2020. Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* 22 (55).
- Yan, Y., Gao, J., Peng, B., 2020. A class of time-varying vector moving average (∞) models. Working paper available at <https://arxiv.org/abs/2010.01492>.
- Yuan, M., 2010. High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* 11, 2261–2286.
- Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* 94, 19–35.
- Zhang, T., Wu, W.B., 2012. Inference of time varying regression models. *Ann. Statist.* 40, 1376–1402.
- Zhang, D., Wu, W.B., 2021. Convergence of covariance and spectral density estimators for high-dimensional locally stationary processes. *Ann. Statist.* 49, 233–254.
- Zhao, J., Liu, X., Wang, H., Leng, C., 2022. Dimension reduction for covariates in network data. *Biometrika* 109, 85–102.
- Zhou, S., Lafferty, J., Wasserman, L., 2010. Time varying undirected graphs. *Mach. Learn.* 80, 295–319.
- Zhu, X., Chang, X., Li, R., Wang, H., 2019. Portal nodes screening for large scale social networks. *J. Econometrics* 209, 145–157.
- Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H., 2017. Network vector autoregression. *Ann. Statist.* 45, 1096–1123.
- Zou, H., Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *Ann. Statist.* 36, 1509–1566.