UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Estimating Time-Varying Networks for High-Dimensional Time Series

Jia Chen, University of York
Degui Li, University of York
Yuning Li, University of York
Oliver Linton, University of Cambridge

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.

Reference Details

2273 Cambridge Working Papers in Economics
2231 Janeway Institute Working Paper Series

Published 14 December 2022

Key Words CLIME, factor model, Granger causality, LASSO, local linear smoothing, partial correlation, time-varying network, VAR
JEL-codes C13, C14, C32, C38

Websites www.econ.cam.ac.uk/cwpe
www.janeway.econ.cam.ac.uk/working-papers

Estimating Time-Varying Networks for High-Dimensional Time Series

Jia Chen*, Degui Li†, Yuning Li‡, Oliver Linton§

University of York and University of Cambridge

Version: December 13, 2022

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.

Keywords: CLIME, factor model, Granger causality, LASSO, local linear smoothing, partial correlation, time-varying network, VAR.

JEL Classification: C13, C14, C32, C38.

*Department of Economics and Related Studies, University of York, UK. Jia Chen's research was partially supported by the ESRC (Grant Reference: ES/T01573X/1).

†Department of Mathematics, University of York, UK.

‡Department of Economics and Related Studies, University of York, UK.

§Faculty of Economics, University of Cambridge, Cambridge, UK.

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, Kilduff and Tasselli, 2013; Diebold and Ylmaz, 2014, 2015; Hautsch, Schaumburg and Schienle, 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, Feng and Wu, 2009; Loh and Wainwright, 2013; Basu, Shojaie and Michailidis, 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, Lafferty and Wasserman (2010) study dynamic network models with smooth time-varying structural changes; whereas Wang, Yu and Rinaldo (2021) 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, Lu and Liu, 2015; Kock and Callot, 2015; Davis, Zang and Zheng, 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, Fan and Zhu (2020) and Miao, Phillips and Su (2022) 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, Qiu and Chen (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, Chen and Wu (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 jointly estimate the time-varying transition and precision matrices in the high-dimensional sparse 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, Gao and Peng (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, Ma and Dai, 2014; Liu, Li and Wu, 2014; Li, Ke and Zhang, 2015), whereas the time-varying CLIME is a natural extension of the conventional CLIME for static precision matrix estimation (e.g., Cai, Liu and Luo, 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 in which case 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. Our paper thus substantially extends the recent work on the factor-adjusted stable VAR model estimation (e.g., Fan, Masini and Medeiros, 2021; Barigozzi, Cho and Owens, 2022; Krampe and Margaritella, 2022).

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.

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 and discussions on tuning parameter selection. 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 $\|\mathbf{W}\| = \lambda_{\max}^{1/2}\left(\mathbf{W}^\intercal\mathbf{W}\right)$ be the operator norm, $\|\mathbf{W}\|_F = \left[\mathsf{Tr}\left(\mathbf{W}^\intercal\mathbf{W}\right)\right]^{1/2}$ the Frobenius norm, $\|\mathbf{W}\|_1 = \max_{1\leqslant j\leqslant d}\sum_{i=1}^{d}|w_{ij}|$, $\|\mathbf{W}\|_{\max} = \max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}|w_{ij}|$, 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 $\mathsf{Tr}(\cdot)$ is the trace. Denote the determinant of a square matrix as $\mathsf{det}(\cdot)$. Let $a_n \sim b_n$, $a_n \propto b_n$ and $a_n \gg b_n$ denote that $a_n/b_n \to 1$, $0 < \underline{c} \leqslant a_n/b_n \leqslant \bar{c} < \infty$ and $b_n/a_n \to 0$, 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 the factor-adjusted time-varying VAR and network setting.

2.1 Time-varying VAR models

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

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

where $\mathbf{A}_{t,k} = \mathbf{A}_k(t/n)$, $k = 1, \ldots, p$, are $d \times d$ time-varying transition matrices with each entry being a smooth deterministic function of scaled times, $\mathbf{\Sigma}_t = \mathbf{\Sigma}(t/n)$ 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 $\mathbf{\Omega}_t = \mathbf{\Omega}(t/n)$ as the inverse of $\mathbf{\Sigma}_t$, 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 $\mathbf{\Sigma}_t$ is replaced by a time-invariant covariance matrix, (2.1) becomes the same model as that considered by Ding, Qiu and Chen (2017). Furthermore, when both $\mathbf{A}_{t,k}$, $k = 1, \ldots, p$, and $\mathbf{\Sigma}_t$ are time-invariant constant matrices, (2.1) becomes the high-dimensional stable VAR:

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

which has been extensively studied in the recent literature (e.g., Basu and Michailidis, 2015; Han, Lu and Liu, 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 $\det\left(\mathbf{I}_d - \sum_{k=1}^{p} \mathbf{A}_k(\tau) z^k\right) \neq 0$ 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 $\mathbf{\Omega}(\tau)$ is positive definite uniformly over $\tau \in [0,1]$, and the operator norm of $\mathbf{\Sigma}(\tau)$ is uniformly bounded over $\tau \in [0,1]$. Furthermore, each entry in $\mathbf{\Sigma}(\tau)$ and $\mathbf{\Omega}(\tau)$ is second-order continuously differentiable over $[0,1]$.

(iii) For any d-dimensional vector u satisfying $\|u\| = 1$, $\mathsf{E}\left[\exp\left\{\iota_1 (u^\tau \varepsilon_t)^2\right\}\right] \leqslant C_0 < \infty$, where ι_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 and leading to the following Wold representation

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

with the coefficient matrices $\mathbf{\Phi}_{t,k}$ being absolutely summable (in appropriate matrix norm). For example, when $p = 1$, we have $\mathbf{\Phi}_{t,0} = \mathbf{I}_d$ and $\mathbf{\Phi}_{t,k} = \Pi_{j=1}^{k} \mathbf{A}_{t-j+1,1}$ for $k \geqslant 1$. Assume that, for k sufficiently large,

$$\max_{0 \leqslant t \leqslant n} \|\mathbf{\Phi}_{t,k}\| \leqslant C_1 \rho^k, \tag{2.4}$$

where C_1 is a positive constant and $0 < \rho < 1$. A similar assumption can be found in Ding, Qiu

and Chen (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 $\mathbf{A}_{t,k} = \left(a_{k,ij|t}\right)_{d\times d}$, $\mathbf{\Omega}_t = \left(\omega_{ij|t}\right)_{d\times d}$, $\mathbf{A}_k(\tau) = \left(a_{k,ij}(\tau)\right)_{d\times d}$ and $\mathbf{\Omega}(\tau) = \left(\omega_{ij}(\tau)\right)_{d\times d}$, where $1 \leqslant t \leqslant n$ and $0 \leqslant \tau \leqslant 1$. We define the network structure via a time-varying graph $\mathbb{G}_t = (\mathbb{V}, \mathbb{E}_t)$, where $\mathbb{V} = \{1, 2, \ldots, d\}$ denotes a set of vertices, and $\mathbb{E}_t = \left\{(i,j) \in \mathbb{V}\times\mathbb{V} : \ c_{ij|t} \neq 0, \ i \neq j\right\}$ 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, Lafferty and Wasserman (2010) for independent network data. Following the stable network analysis in Barigozzi and Brownlees (2019) and Barigozzi, Cho and Owens (2022), 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 $k \in \{1, 2, \ldots, p\}$ 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 $\mathbf{A}_k = \left(a_{k,ij}\right)_{d\times d}$ 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 $\mathbb{G}_\tau^G = \left(\mathbb{V}, \mathbb{E}_\tau^G\right)$ with

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

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 $\rho_{ij|t} = \mathsf{cor}(e_{t,i}, e_{t,j}|e_{t,k}, k \neq i,j)$ 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 $\mathbb{G}_\tau^P = \left(\mathbb{V}, \mathbb{E}_\tau^P\right)$ 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 $\mathbb{G}_\tau^G$ and $\mathbb{G}_\tau^P$. Denote the uniform graphs by $\mathbb{G}^G = \left(\mathbb{V}, \mathbb{E}^G\right)$ and $\mathbb{G}^P = \left(\mathbb{V}, \mathbb{E}^P\right)$, with

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

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 provide their estimates.

3 Methodology

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

$$\boldsymbol{\alpha}_{i\bullet}(\cdot) = \left[A_{1,i}^\intercal(\cdot), \ldots, A_{p,i}^\intercal(\cdot)\right]^\intercal, \ \ \mathbf{X}_t = \left(X_t^\intercal, \ldots, X_{t-p+1}^\intercal\right)^\intercal, \tag{3.1}$$

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

$$x_{t,i} = \boldsymbol{\alpha}_{i\bullet}^\intercal(\tau_t)\mathbf{X}_{t-1} + e_{t,i} \ \text{ with } \ e_{t,i} = C_i^\intercal(\tau_t)\varepsilon_t, \ \ i = 1, \ldots, d, \tag{3.2}$$

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 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 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, Li and Huang, 2008; Wang and Xia, 2009; Lian, 2012; Cheng *et al.*, 2014; Fan, Ma and Dai, 2014; Liu, Li and Wu, 2014; Li, Ke and Zhang, 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 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ (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 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ from Stage (i); (iii) estimate the error precision matrix $\boldsymbol{\Omega}(\cdot)$ 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 $\boldsymbol{\alpha}_{i\bullet}(\tau_t)$:

$$\boldsymbol{\alpha}_{i\bullet}(\tau_t) \approx \boldsymbol{\alpha}_{i\bullet}(\tau) + \boldsymbol{\alpha}'_{i\bullet}(\tau)(\tau_t - \tau), \quad i = 1, \ldots, d,$$

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

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

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 $\boldsymbol{\alpha}_{i\bullet}(\tau)$ and $\boldsymbol{\alpha}'_{i\bullet}(\tau)$ are obtained by minimising $\mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. 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, Chen and Gao, 2011). In our high-dimensional setting, as the number of predictors may exceed n, it is challenging to obtain satisfactory estimation by directly minimising $\mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$. To address this issue, we assume that the number of significant components in $\boldsymbol{\alpha}_{i\bullet}(\tau)$ 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

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

where λ_1 is a tuning parameter. Let $\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau)$ and $\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau)$ be the solution to the minimisation of $\mathcal{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. 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, $t = 1, \ldots, n$, 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

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

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

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

In practice, $D^2_{i,j}$ can be estimated by

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

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

$$\mathcal{Q}_i(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^{n} \mathcal{L}_i(\boldsymbol{\alpha}_{\bullet t}, \boldsymbol{\beta}_{\bullet t} \mid \tau_t) + \sum_{j=1}^{pd} p'_{\lambda_2}\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \|\boldsymbol{\alpha}_j\| + \sum_{j=1}^{pd} p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right) \|h\boldsymbol{\beta}_j\|, \tag{3.5}$$

where

$$\widetilde{\boldsymbol{\alpha}}_{i,j} = [\widetilde{\alpha}_{i,j}(\tau_1), \ldots, \widetilde{\alpha}_{i,j}(\tau_n)]^\intercal, \ \ \boldsymbol{\alpha}_j = \left(\alpha_{j|1}, \ldots, \alpha_{j|n}\right)^\intercal, \ \ \boldsymbol{\beta}_j = \left(\beta_{j|1}, \ldots, \beta_{j|n}\right)^\intercal,$$

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

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

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 $p'_{\lambda_2}(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|)$ and $p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right)$ 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 $\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|$ or $\widetilde{D}_{i,j}$ is close to zero, whereas it is light or equal to zero when $\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|$ or $\widetilde{D}_{i,j}$ is large. An advantage of using $\widetilde{D}_{i,j}$ in the second penalty term over the L_2-norm of $\widetilde{\boldsymbol{\alpha}}'_j = \left[\widetilde{\alpha}'_{i,j}(\tau_1), \ldots, \widetilde{\alpha}'_{i,j}(\tau_n)\right]^\intercal$ is that the estimates of the time-varying parameters involved in $\widetilde{D}_{i,j}$ often perform more stably than their derivative counterparts.

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

$$\widehat{\mathbf{A}}_i = (\widehat{\boldsymbol{\alpha}}_{i,1}, \ldots, \widehat{\boldsymbol{\alpha}}_{i,pd}) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}_{i,j} = [\widehat{\alpha}_{i,j}(\tau_1), \ldots, \widehat{\alpha}_{i,j}(\tau_n)]^\intercal,$$
$$\widehat{\mathbf{B}}_i = (\widehat{\boldsymbol{\alpha}}'_{i,1}, \ldots, \widehat{\boldsymbol{\alpha}}'_{i,pd}) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}'_{i,j} = \left[\widehat{\alpha}'_{i,j}(\tau_1), \ldots, \widehat{\alpha}'_{i,j}(\tau_n)\right]^\intercal.$$

The index set $\mathcal{J}_i$ is estimated by $\widehat{\mathcal{J}}_i = \{j : \widehat{\boldsymbol{\alpha}}_{i,j} \neq \mathbf{0}_n\}$, and $\mathcal{J}'_i$ is estimated by $\widehat{\mathcal{J}}'_i = \left\{j : \widehat{\boldsymbol{\alpha}}'_{i,j} \neq \mathbf{0}_n\right\}$, where $\mathbf{0}_k$ is a k-dimensional vector of zeros. A similar shrinkage estimation method is used by Li, Ke and Zhang (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 $\boldsymbol{\Omega}(\cdot)$ in model (2.1), which is crucial to uncover the time-varying and uniform network structures of 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, Liu and Luo, 2011). Xu, Chen and Wu (2020) further introduce a time-varying CLIME method for high-dimensional locally stationary time series which are observable. Note that in this paper, $\boldsymbol{\Omega}(\cdot)$ 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 $\widehat{\boldsymbol{\alpha}}_{i\bullet}(\cdot)$, $i = 1, \ldots, d$, from Section 3.2, we can then extract estimates of the time-varying transition matrices, denoted by $\widehat{\mathbf{A}}_k(\tau_t)$, $t = 1, \ldots, n$, $k = 1, \ldots, p$, and approximate e_t by

$$\widehat{e}_t = (\widehat{e}_{t,1}, \ldots, \widehat{e}_{t,d})^\intercal = X_t - \sum_{k=1}^{p} \widehat{\mathbf{A}}_k(\tau_t) X_{t-k}, \quad t = 1, \ldots, n. \tag{3.6}$$

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 $\mathbf{\Omega}(\cdot)$ satisfies a uniform sparsity assumption, a natural extension of the classic sparsity assumption to the locally stationary time series setting. Specifically, we assume $\{\mathbf{\Omega}(\tau) : 0 \leqslant \tau \leqslant 1\} \in \mathcal{S}(q, \xi_d)$, where

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

where $0 \leqslant q < 1$, "$\mathbf{W} \succ 0$" denotes that $\mathbf{W}$ is positive definite, and C_2 is a bounded positive constant. Define

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

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 $\mathbf{\Omega}(\tau)$ via the time-varying CLIME method:

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

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 $\widehat{\mathbf{\Omega}}(\tau) = [\widehat{\omega}_{ij}(\tau)]_{d\times d}$, where

$$\widehat{\omega}_{ij}(\tau) = \widehat{\omega}_{ji}(\tau) = \widetilde{\omega}_{ij}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| \leqslant |\widetilde{\omega}_{ji}(\tau)|\right) + \widetilde{\omega}_{ji}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| > |\widetilde{\omega}_{ji}(\tau)|\right). \tag{3.10}$$

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:

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

and

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

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

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

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

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

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:

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

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), $\mathbf{e}_t = \left(e_t^\intercal, 0_d^\intercal, \ldots, 0_d^\intercal\right)^\intercal$, and $\mathbf{A}_t^*$ is a $(pd) \times (pd)$ time-varying transition matrix:

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

4.1 Uniform consistency of the time-varying LASSO estimates

Define

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

and

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

where $\mathcal{J}_i(\tau)$ 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*

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

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

Assumption 3. *(i) The tuning parameter* λ_1 *satisfies*

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

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

$$\min_{1 \leqslant i \leqslant d} \min_{1 \leqslant t \leqslant n} \inf_{u \in \mathcal{B}_i(\tau_t)} u^\intercal \boldsymbol{\Psi}(\tau_t) u \geqslant \kappa_0. \tag{4.3}$$

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 $nh/\log^2 n \to \infty$ and $h^2 \log n \to 0$. 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, Ritov and Tsybakov, 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 $\max_{1 \leqslant t \leqslant n} \|\boldsymbol{\Psi}(\tau_t) - \mathsf{E}[\boldsymbol{\Psi}(\tau_t)]\|_{\max} = O_P\left(\sqrt{\log(n \vee d)/(nh)}\right)$. Then, using Lemma D.1 in Appendix D and assuming $s\sqrt{\log(n \vee d)/(nh)} = o(1)$, a sufficient condition

for (4.3) is

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathsf{E}\left[\Psi(\tau_t)\right] u \geqslant \kappa_0.$$

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

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{4.4}$$

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

$$\max_{1\leqslant i\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau)-\boldsymbol{\alpha}_{i\bullet}(\tau)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{4.5}$$

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 $\overline{\mathcal{J}}_i$ and $\overline{\mathcal{J}}_i'$, respectively, i.e., $\overline{\mathcal{J}}_i = \bigcap_{t=1}^n \{j:\ \alpha_{i,j}(\tau_t)=0\}$ and $\overline{\mathcal{J}}_i' = \bigcap_{t=1}^n \{j:\ \alpha_{i,j}'(\tau_t)=0\}$. Let $\mathbf{A}^o = (\boldsymbol{\alpha}_{\bullet 1}^o,\ldots,\boldsymbol{\alpha}_{\bullet n}^o)^\intercal$ and $\mathbf{B}^o = (\boldsymbol{\beta}_{\bullet 1}^o,\ldots,\boldsymbol{\beta}_{\bullet n}^o)^\intercal$, where $\boldsymbol{\alpha}_{\bullet t}^o = (\alpha_{1|t}^o,\ldots,\alpha_{d|t}^o)^\intercal$ with $\alpha_{j|t}^o = 0$ for $j\in\overline{\mathcal{J}}_i$ and $\boldsymbol{\beta}_{\bullet t}^o = (\beta_{1|t}^o,\ldots,\beta_{d|t}^o)^\intercal$ with $\beta_{j|t}^o = 0$ for $j\in\overline{\mathcal{J}}_i'$. Define the (infeasible) oracle estimates:

$$\widehat{\mathbf{A}}_i^o = \left(\widehat{\boldsymbol{\alpha}}_{i,1}^o,\ldots,\widehat{\boldsymbol{\alpha}}_{i,d}^o\right) \ \text{ with } \ \widehat{\boldsymbol{\alpha}}_{i,j}^o = \left[\widehat{\alpha}_{i,j}^o(\tau_1),\ldots,\widehat{\alpha}_{i,j}^o(\tau_n)\right]^\intercal, \tag{4.6}$$

$$\widehat{\mathbf{B}}_i^o = \left(\widehat{\boldsymbol{\alpha}}_{i,1}'^o,\ldots,\widehat{\boldsymbol{\alpha}}_{i,d}'^o\right) \ \text{ with } \ \widehat{\boldsymbol{\alpha}}_{i,j}'^o = \left[\widehat{\alpha}_{i,j}'^o(\tau_1),\ldots,\widehat{\alpha}_{i,j}'^o(\tau_n)\right]^\intercal, \tag{4.7}$$

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*

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

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

(ii) It holds that

$$\min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}_i}\left(\sum_{t=1}^{n}\alpha_{i,j}^2(\tau_t)\right)^{\frac{1}{2}}\geqslant(a_0+1)\lambda_2 \text{ and } \min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}_i'}D_{i,j}\geqslant(a_0+1)\lambda_2,$$

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, Ke and Zhang (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,* $Q_i(\mathbf{A}, \mathbf{B})$*, 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:*

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

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. Following the proof of Theorem 4.2, we may verify properties (i)–(iv) for the folded concave penalty function discussed in Fan, Xue and Zou (2014) *w.p.a.1*. Hence, Theorem 4.2 may be regarded as a generalisation of Theorem 1 in Fan, Xue and Zou (2014) and Theorem 3.1 in Li, Ke and Zhang (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*

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

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*

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

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

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

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

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

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

$$\sup_{0\leqslant\tau\leqslant 1}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{\max} = O_P\left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right), \tag{4.10}$$

$$\sup_{0\leqslant\tau\leqslant 1}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\| = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d})^{1-q}\right), \tag{4.11}$$

$$\sup_{0\leqslant\tau\leqslant 1}\frac{1}{d}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_F^2 = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d})^{2-q}\right), \tag{4.12}$$

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

The uniform convergence rates in Theorem 4.3 rely on $\nu^{\diamond}_{n,d}$ and $\nu^{*}_{n,d}$. The first rate $\nu^{\diamond}_{n,d}$ is the conventional uniform convergence rate for nonparametric kernel-based quantities, whereas the second rate $\nu^{*}_{n,d}$ is from the approximation errors of $\widehat{e}_t$ 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. Theorem 4.3 can be seen as an extension of Theorem 1 in Cai, Liu and Luo (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* $\min_{(i,j)\in\mathbb{E}^P} \min_{1\leqslant t\leqslant n} |\omega_{ij}(\tau_t)| \gg \lambda_3$, *we have*

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^P = \mathbb{E}_n^P\right) \to 1. \tag{4.13}$$

5 Factor-adjusted time-varying VAR and networks

In this section, we let $(Z_t : t = 1,\ldots,n)$ with $Z_t = (z_{t,1},\ldots,z_{t,d})^\intercal$ 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:

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

where $\boldsymbol{\Lambda} = (\Lambda_1,\ldots,\Lambda_d)^\intercal$ 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:

$$Z_t = \boldsymbol{\Lambda}_t F_t + X_t, \quad t = 1,\ldots,n, \tag{5.2}$$

where $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}(t/n)$ is a time-varying factor loading matrix with each entry being a smooth function of scaled time. 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, Hafner and von Sachs, 2011; Su and Wang, 2017). 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 ($\boldsymbol{\Lambda} F_t$ or $\boldsymbol{\Lambda}_t F_t$) nor the idiosyncratic error components are observable. Motivated by recent work on bridging factor and sparse models for high-dimensional data (e.g., Fan, Masini and Medeiros, 2021; Krampe and Margaritella, 2022), 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 $\mathbf{Z} = (Z_1,\ldots,Z_n)^\intercal$, $\mathbf{F} = (F_1,\ldots,F_n)^\intercal$ and $\mathbf{X} = (X_1,\ldots,X_n)^\intercal$. For the conventional factor model (5.1), we conduct an eigenanalysis on the $n \times n$ matrix $\mathbf{Z}\mathbf{Z}^\intercal$. The estimate of $\mathbf{F}$, denoted as $\widehat{\mathbf{F}} = \left(\widehat{F}_1,\ldots,\widehat{F}_n\right)^\intercal$, is obtained as the $n \times k$ matrix consisting of the eigenvectors (multiplied by $\sqrt{n}$) corresponding to the k largest eigenvalues of $\mathbf{Z}\mathbf{Z}^\intercal$. The factor loading matrix is estimated by

$\widehat{\mathbf{\Lambda}} = \left(\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_d\right)^\intercal = \mathbf{Z}^\intercal \widehat{\mathbf{F}}/n$. Consequently, the common component $\mathbf{\Lambda} F_t$is estimated by $\widehat{\mathbf{\Lambda}}\widehat{F}_t$ and the idiosyncratic error component X_t is estimated by

$$\widehat{X}_t = Z_t - \widehat{\mathbf{\Lambda}}\widehat{F}_t, \quad t = 1, \ldots, n. \tag{5.3}$$

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:

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

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

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

where we've kept the same notation $\widehat{X}_t$ 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 $\widehat{X}_t$, we can apply the three-stage estimation procedure proposed in Section 3. Denote the preliminary time-varying LASSO estimate as $\widetilde{\alpha}_{ij}^\dagger(\cdot)$, the second-stage weighted group LASSO estimate as $\widehat{\alpha}_{ij}^\dagger(\cdot)$, and the factor-adjusted time-varying precision matrix estimate as $\widehat{\mathbf{\Omega}}^\dagger(\cdot) = \left[\widehat{\omega}_{ij}^\dagger(\cdot)\right]_{d\times d}$. Subsequently, we may construct the uniform network estimates $\widehat{\mathbb{E}}_n^{G,\dagger}$ and $\widehat{\mathbb{E}}_n^{P,\dagger}$, defined similarly to $\widehat{\mathbb{E}}_n^G$ and $\widehat{\mathbb{E}}_n^P$ in (3.13) and (3.14), but with $\widehat{\alpha}_{ij}(\cdot)$ and $\widehat{\omega}_{ij}(\cdot)$ replaced by $\widehat{\alpha}_{ij}^\dagger(\cdot)$ and $\widehat{\omega}_{ij}^\dagger(\cdot)$, 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* $\delta_X = \max_{1\leqslant t\leqslant n}\left|\widehat{X}_t - X_t\right|_{\max}$. *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* $\zeta_{n,d}^\dagger = \zeta_{n,d} + [\log(n \vee d)]^{1/2} s\delta_X$.

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

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

Assumption 6(i) imposes a high-level condition on the approximation of the 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, Liao and Mincheva (2013), a typical rate for the approximation error from PCA estimation of the conventional factor model (5.1) is

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

where $\upsilon > 2$ 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 $d \gg n^{1+2/\upsilon}$, then the rate in (5.5) can be simplified to $\delta_X = O_P\left((\log d) n^{-1/2}\right) = o_P(h^2)$ and thus $\zeta_{n,d} \propto \zeta^\dagger_{n,d}$. 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

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

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

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left[\widehat{\alpha}^\dagger_{ij}(\tau_t) - \alpha_{ij}(\tau_t)\right]^2 = O_P\left(s\left(\zeta^\dagger_{n,d}\right)^2\right). \tag{5.7}$$

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

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\mathbf{\Omega}}^\dagger(\tau) - \mathbf{\Omega}(\tau)\right\|_{\max} = O_P\left(\nu^\diamond_{n,d} + \nu^\dagger_{n,d}\right), \tag{5.8}$$

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\mathbf{\Omega}}^\dagger(\tau) - \mathbf{\Omega}(\tau)\right\| = O_P\left(\xi_d(\nu^\diamond_{n,d} + \nu^\dagger_{n,d})^{1-q}\right), \tag{5.9}$$

$$\sup_{0\leqslant\tau\leqslant1}\frac{1}{d}\left\|\widehat{\mathbf{\Omega}}^\dagger(\tau) - \mathbf{\Omega}(\tau)\right\|_F^2 = O_P\left(\xi_d(\nu^\diamond_{n,d} + \nu^\dagger_{n,d})^{2-q}\right). \tag{5.10}$$

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

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

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

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

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 (infeasible) time-varying oracle estimation, denoted as tv-Oracle, which estimates only the true significant coefficient functions (assuming they were known), and the unpenalised full time-varying estimation, 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 estimation, 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 VAR errors (rather than residuals) in the estimation of the 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, Ke and Zhang (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). We set the sample size n as 200 and 400, and the dimension d as 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 derivative, $2d = 200$. 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 E in the supplement gives definitions of these information criteria.

To evaluate whether the time-varying model structure is accurately estimated, 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 available in Appendix E of the supplement. To evaluate the performance of the coefficient estimators, we report the average R square (average R^2) over all the dimensions, the average scaled Frobenius norm of estimation errors of coefficient functions (EE_A), and the root-mean-squared error of the errors ($RMSE_e$). Taking our proposed tv-wgLASSO estimator for time-varying VAR(1) as an example,

$$EE_A = \frac{1}{n\sqrt{d}} \sum_{t=1}^{n} \left\| \widehat{\mathbf{A}}_1(\tau_t) - \mathbf{A}_1(\tau_t) \right\|_F \quad \text{and} \quad RMSE_e = \sqrt{\frac{1}{nd} \sum_{i=1}^{d} \sum_{t=1}^{n} (\widehat{e}_{t,i} - e_{t,i})^2}.$$

To evaluate the performance of the precision matrix estimators, we report the average scaled Frobenius norm of estimation error (EE_Ω) defined as

$$EE_\Omega = \frac{1}{n\sqrt{d}} \sum_{t=1}^{n} \left\| \widehat{\mathbf{\Omega}}(\tau_t) - \mathbf{\Omega}(\tau_t) \right\|_F.$$

All the above measures are calculated for each Monte Carlo replication and then averaged over 100 replications.

Example 1. The data is generated from a time-varying VAR(1) model with $\mathbf{A}_1(\tau)$ being a diagonal matrix for all $\tau \in [0,1]$. Each diagonal entry of $\mathbf{A}_1(\tau)$ independently takes a 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 $\mathbf{\Omega}(\tau)$ to be a block diagonal matrix: $\mathbf{\Omega}(\tau) = \mathbf{I}_{d/2} \otimes \mathbf{\Omega}_*(\tau)$, where $\mathbf{\Omega}_*(\tau) = [\omega_{ij,*}(\tau)]_{2\times 2}$ 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 $\mathbf{A}_1(\tau)$ implies that no Granger causality exists between variables, whereas the block diagonal structure of $\mathbf{\Omega}(\tau)$ 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 (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 recover 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 of 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 recover 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 $\mathbf{A}_1(\tau)$ being an upper triangular matrix for all $\tau \in [0, 1]$. Each diagonal entry of $\mathbf{A}_1(\tau)$ 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 $\mathbf{\Omega}(\tau) = [\omega_{ij}(\tau)]_{d\times d}$ 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 of 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 of 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 directly 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 recovering 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 $\mathbf{A}_1(\tau) = [a_{ij}(\tau)]_{d\times d}$ being a Toeplitz

Table 1: Transition matrix and Granger network estimation in Example 1.

		tv-wgLASSO		tv-Oracle		tv-Full	
measure	dimension	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
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

In all the tables, except for exact values of 0's and 1's, the FP and FN measures are rounded to 2 decimal places, while the others are rounded to 3 decimal places.

Table 2: Precision matrix and partial correlation network estimation in Example 1.

		tv-CLIME		infeasible tv-CLIME		tv-GLASSO	
measure	dimension	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
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

matrix and $a_{ij}(\tau) = (0.4 - 0.1\tau)^{|i-j|+1}$. We also set $\mathbf{\Omega}(\tau) = [\omega_{ij}(\tau)]_{d\times d}$ to be 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, and we aim to examine how our proposed methods perform when the (exact) sparsity assumption fails.

Table 5 reports the estimation errors of the various methods considered. In this example, the tv-Oracle is equivalent to tv-Full and both suffer from the curse of dimensionality in the conventional local linear estimation procedure for the time-varying transition matrices (in particular 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 Oracle tv-CLIME are generally close to those of tv-GLASSO. The simulation results show that the proposed tv-wgLASSO and tv-CLIME perform reasonably well when the sparsity assumption on transition and precision matrices is not satisfied.

Example 4. The data is generated from a factor-adjusted time-varying VAR model in the form of (5.2). The idiosyncratic errors of the time-varying factor model are generated from a VAR(1) model in Example 2. The two factors in $F_t = (F_{t,1}, F_{t,2})^\intercal$ are generated from two univariate

Table 3: Transition matrix and Granger network estimation in Example 2.

		tv-wgLASSO		tv-Oracle		tv-Full	
measure	dimension	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
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.

		tv-CLIME		infeasible tv-CLIME		tv-GLASSO	
measure	dimension	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
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

Table 5: Estimation accuracy of dual networks in Example 3.

		tv-wgLASSO		tv-Oracle		tv-Full	
measure	dimension	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
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

AR(1) processes: $F_{t,1} = 0.6F_{t-1,1} + \sqrt{1-0.6^2}u^F_{t,1}$ and $F_{t,2} = 0.3F_{t-1,2} + \sqrt{1-0.3^2}u^F_{t,2}$, where $u^F_{t,1}$ and $u^F_{t,2}$ are independently drawn from a standard normal distribution. The factor-loading matrix is defined as $\mathbf{\Lambda}_t = (\Lambda_{t,1}, \Lambda_{t,2})$ where $\Lambda_{t,1} \equiv \Lambda_1$ is a time-invariant vector drawn from a d-dimensional standard multivariate normal distribution and $\Lambda_{t,2} = (\Lambda_{1t,2}, \ldots, \Lambda_{dt,2})^\intercal$ with $\Lambda_{it,2} = 2/\left(1 + \exp\{-2[10(t/n) - 5(i/d) - 2]\}\right)$ for $i = 1, \ldots, d$.

Table 6 reports the estimation results of the time-varying transition matrices and Granger networks for the idiosyncratic errors, and Table 7 reports the estimation results of the time-varying precision matrices and partial correlation networks. Comparing with the results in Tables 3 and 4, 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 onto the three-stage estimation procedure, their impact on the estimation of the networks is not significant when the sample size is moderately large ($n = 400$).

7 An empirical application

In this section, we apply the proposed methods to estimate the Granger causality and partial correlation networks using the FRED-MD macroeconomic dataset. The dataset, available on the Fred-MD website[1], consists of 127 U.S. macroeconomic variables observed monthly over the period from January 1959 to July 2022. These macroeconomic variables can be 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).

We follow McCracken and Ng (2016) and McCracken and Ng (2020) to remove outliers and fill missing values. Each variable is standardised to 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 E in the supplement for 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, denoted as $\widehat{x}_{t,i}$,

[1] https://research.stlouisfed.org/econ/mccracken/fred-databases/

Table 6: Factor-adjusted transition matrix and Granger network estimation in Example 4.

		tv-wgLASSO	
measure	dimension	$n = 200$	$n = 400$
FP	$d = 50$	11.35	10.60
	$d = 100$	20.40	10.41
FN	$d = 50$	35.97	14.77
	$d = 100$	65.45	20.68
TPR	$d = 50$	0.637	0.851
	$d = 100$	0.671	0.896
TNR	$d = 50$	0.995	0.996
	$d = 100$	0.998	0.999
PPV	$d = 50$	0.852	0.890
	$d = 100$	0.869	0.945
NPV	$d = 50$	0.985	0.994
	$d = 100$	0.993	0.998
F1	$d = 50$	0.725	0.869
	$d = 100$	0.756	0.920
MCC	$d = 50$	0.725	0.865
	$d = 100$	0.759	0.919
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

Table 7: Factor-adjusted precision matrix and partial correlation network estimation in Example 4.

		tv-CLIME	
measure	dimension	$n = 200$	$n = 400$
FP	$d = 50$	0.01	0.01
	$d = 100$	0	0.02
FN	$d = 50$	38.22	5.36
	$d = 100$	65.99	2.21
TPR	$d = 50$	0.220	0.891
	$d = 100$	0.333	0.978
TNR	$d = 50$	1.000	1.000
	$d = 100$	1	1.000
PPV	$d = 50$	0.999	1.000
	$d = 100$	1	1.000
NPV	$d = 50$	0.969	0.995
	$d = 100$	0.987	1.000
F1	$d = 50$	0.349	0.941
	$d = 100$	0.496	0.989
MCC	$d = 50$	0.448	0.941
	$d = 100$	0.570	0.988
EE_Ω	$d = 50$	0.670	0.585
	$d = 100$	0.628	0.534

$i = 1, \ldots, 127$, $t = 1, \ldots, 763$, are then used for our empirical analysis. Miao, Phillips and Su (2022) suggest determining the optimal order of a high-dimensional VAR model via a ratio criterion, comparing the Frobenius norms of the estimated transition matrices over different lags. We extend their criterion to the time-varying VAR model context (see Appendix E in the supplement for detail) and subsequently select the time-varying VAR(1) model for $\widehat{X}_t = (\widehat{x}_{t,1}, \ldots, \widehat{x}_{t,127})^\tau$.

Figure 1 plots the estimated Granger networks from the static VAR(1) and the time-varying VAR(1) models. From the estimated time-varying transition matrix, 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 matrix, 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 ones 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 captures more linkages in the network estimation. Figure 2 plots the Granger networks estimated without factor adjustment. Compared with the factor-adjusted version, the Granger network via time-varying VAR(1) is more dense with 1118 directed linkages, among which 104 are self-linkages and 432 are within categories. As pointed out by McCracken and Ng (2016), common factors, which may be interpreted as business cycles, are the main sources of the Granger causalities between macroeconomic variables, leading to a rather dense network structure. On the other hand, the estimated Granger network via static VAR(1) without factor adjustment has only 450 linkages.

We further explore the dynamic smooth structural changes of Gaussian causality linkages. 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: 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 logarithmic growth rate of the S&P PE ratio which is a self-linkage. We re-estimate the corresponding time-varying coefficients using the nonparametric autoregression model with only the four selected predictors, and draw the 90% confidence bands using the R package "tvReg". Figure 3 plots the estimated curves of the four coefficient functions. We find that the logarithmic growth rate of S&P PE ratio is generally persistent and positively correlated to BOGMBASE in the most recent two decades. The estimated time-varying coefficient of the S&P 500 industrials index return is significant but close to zero. It is thus unsurprising that the static VAR(1) model with classic LASSO penalty does not detect the Granger causality linkage from this

[2]We show in the parentheses the variable names used in the FRED-MD dataset. The variable transformation is conducted following the guideline in the dataset.

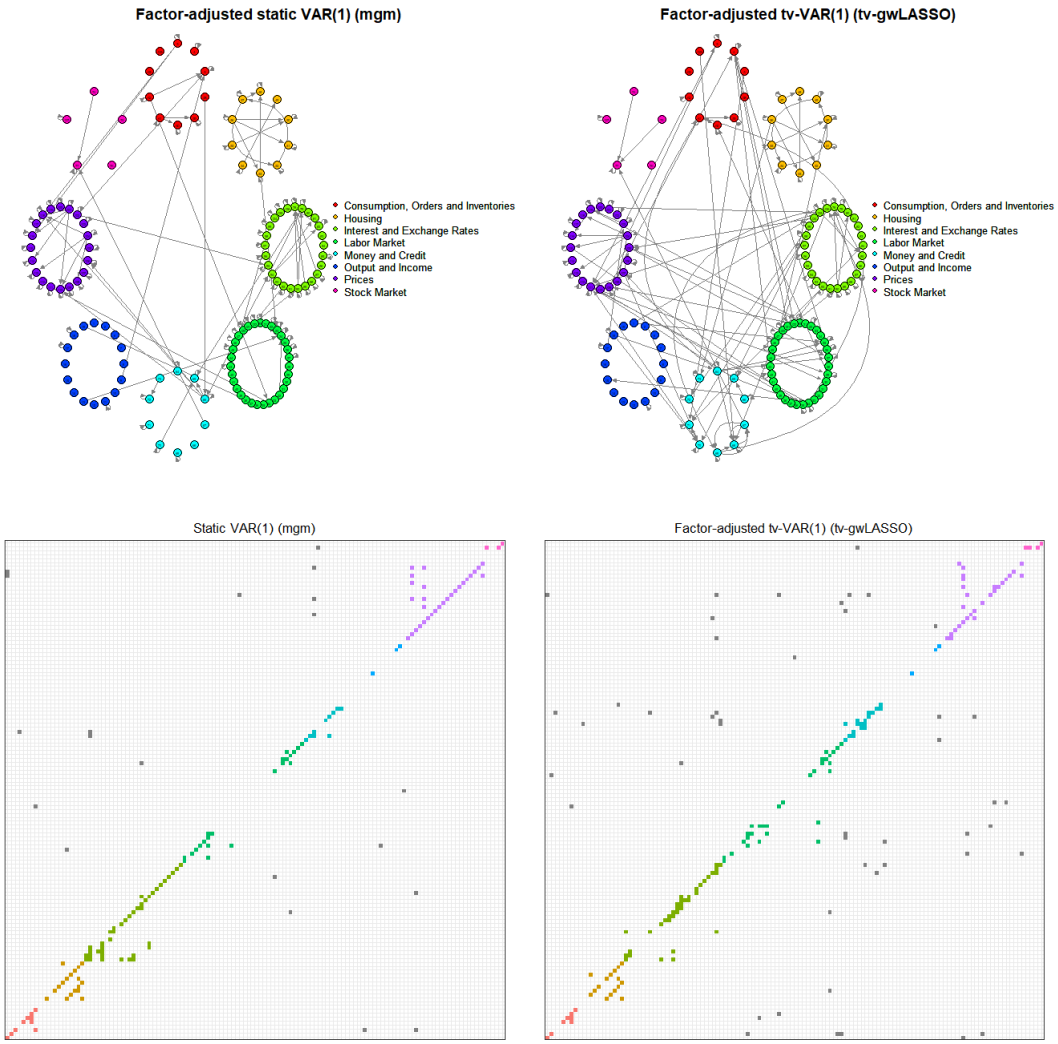

Figure 1: The estimated Granger causality networks using the factor-adjusted static VAR(1) model (left) and time-varying VAR(1) model (right).

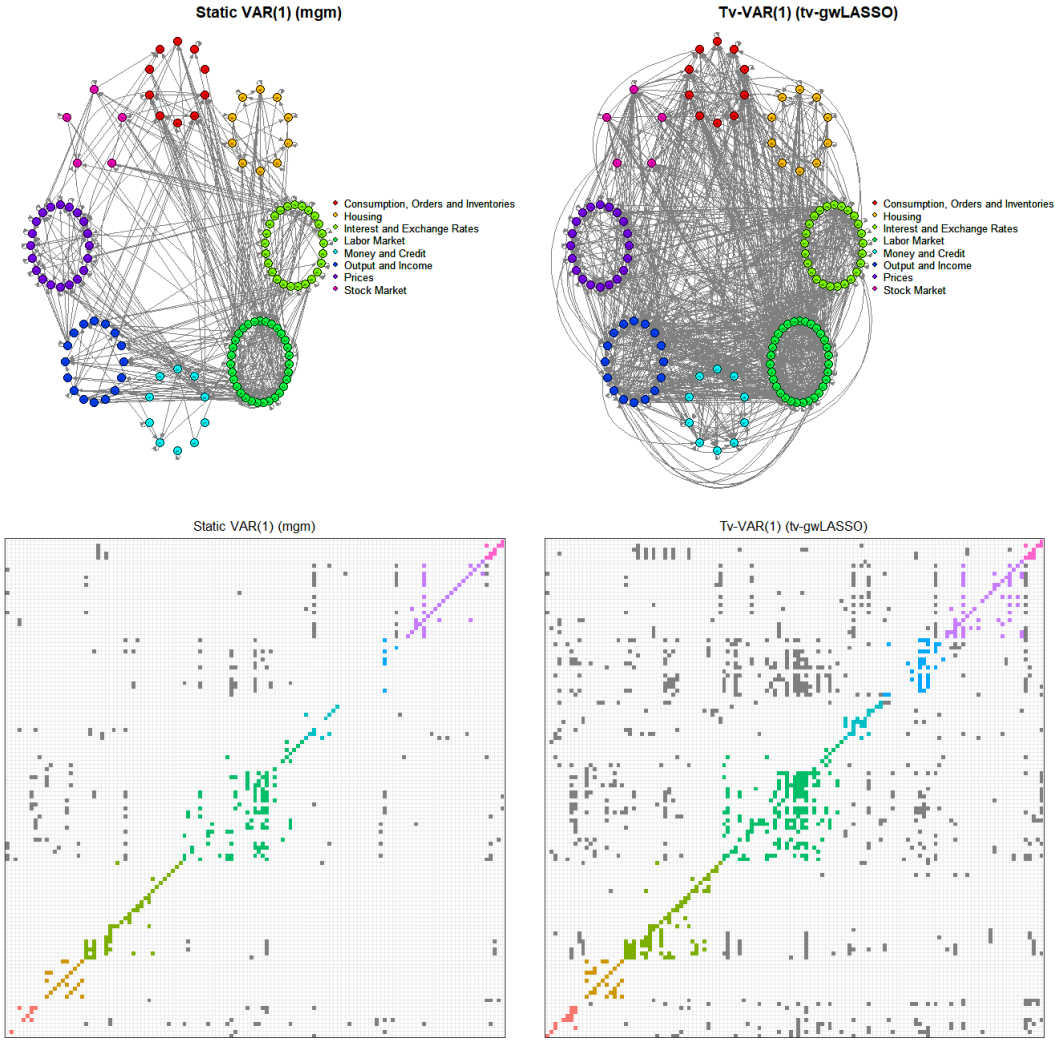

Figure 2: The estimated Granger causality networks using the static VAR(1) model (left) and time-varying VAR(1) model (right) without factor-adjustment.

variable. In fact, LASSO tends to select only one variable in a group of highly-correlated predictors. Due to high correlation between the two index returns, only the S&P 500 Index return is selected in the static VAR(1) model. In contrast, the proposed time-varying LASSO selects both of the two index returns at different time periods, and the second-stage weighted group LASSO aggregates the information over time and selects both index returns.

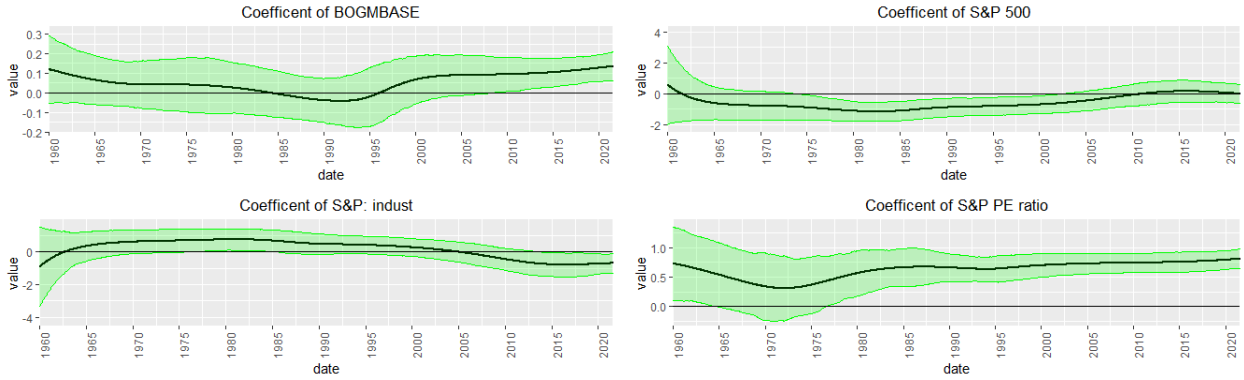

Figure 3: The estimated time-varying coefficients linked to S&P PE ratio with 90% confident bands.

We plot the estimated partial correlation networks in Figure 4, which are generally sparse. Using the factor-adjusted time-varying CLIME, 234 undirected linkages are detected in the estimated network, among which 205 linkages are within the same category. In contrast, the estimated network without factor adjustment contains 236 linkages with 211 in the same category. Unlike the Granger network estimation, it seems that whether to make factor adjustment or not has little impact on the partial correlation network estimation.

We next examine the time-varying pattern of partial correlation linkages between S&P PE ratio and four other variables: S&P 500, S&P: indust, S&P div yield (the increment of S&P composite common stock: dividend yield), and BAAFFM (the spread between Moody's seasoned baa corporate bond and effective federal funds rate). We re-estimate the relevant time-varying functions with a 200-month moving window (Jankova and van de Geer, 2015), and draw the 90% confidence bands using R package "SILGGM" in Figure 5. Note that the partial correlation has a sign opposite to the corresponding entry in the precision matrix. We find 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. The confidence bands in Figure 5 suggest that time-invariant partial correlation linkages are inappropriate to describe the network structure of the FRED-MD data.

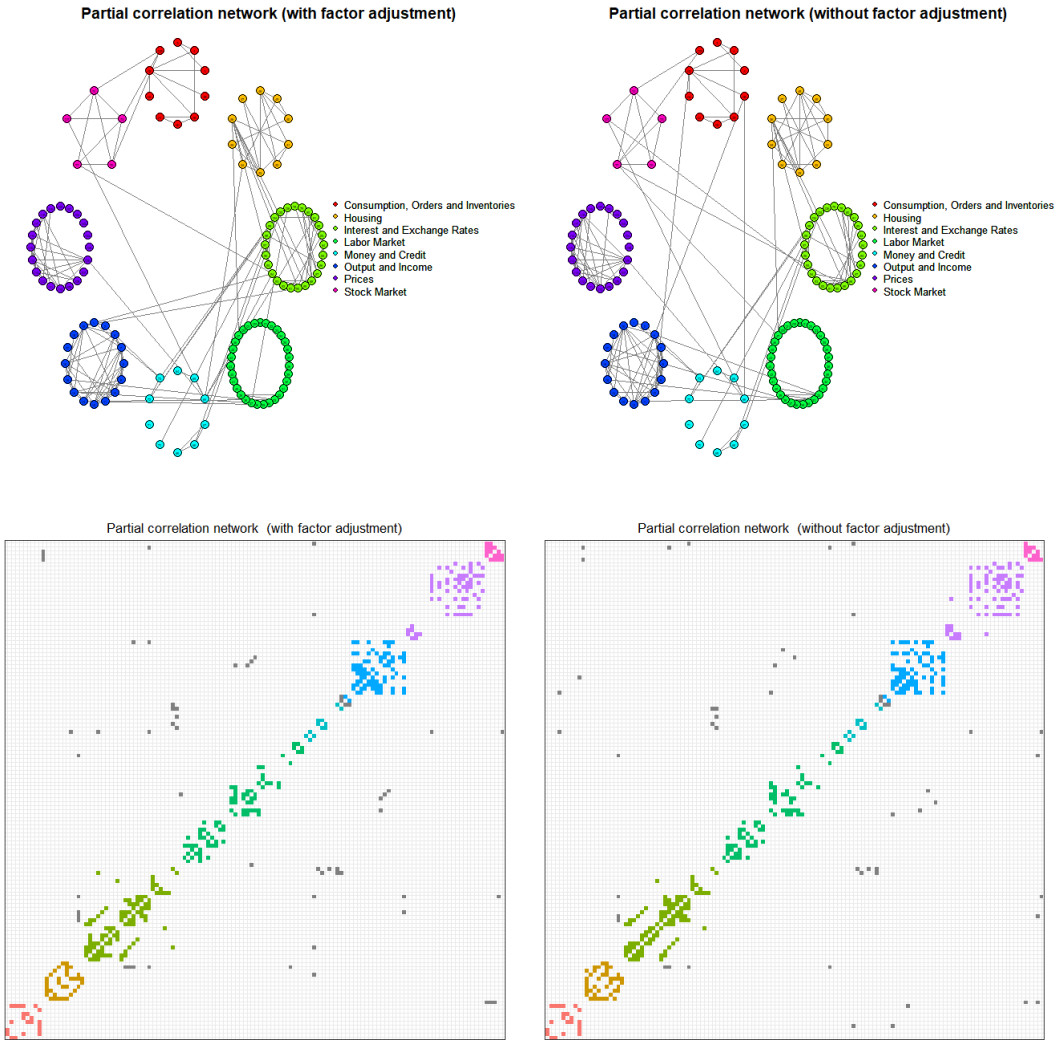

Figure 4: The estimated partial correlation networks with (left) and without (right) factor adjustment.

8 Conclusion

In this paper we estimate a general time-varying VAR model for high-dimensional locally stationary time series. A three-stage estimation procedure combining time-varying LASSO, weighted group LASSO and time-varying CLIME is developed to estimate both transition and error precision matrices, allowing smooth structural changes over time. The estimated transition and precision matrices are further used to construct dual network structures with directed Granger causality linkages and undirected partial correlation linkages, respectively. Under the sparse structural assumption and other technical conditions, we derive the uniform consistency and oracle properties for the developed estimates. In order to accommodate high correlation among large-scale time

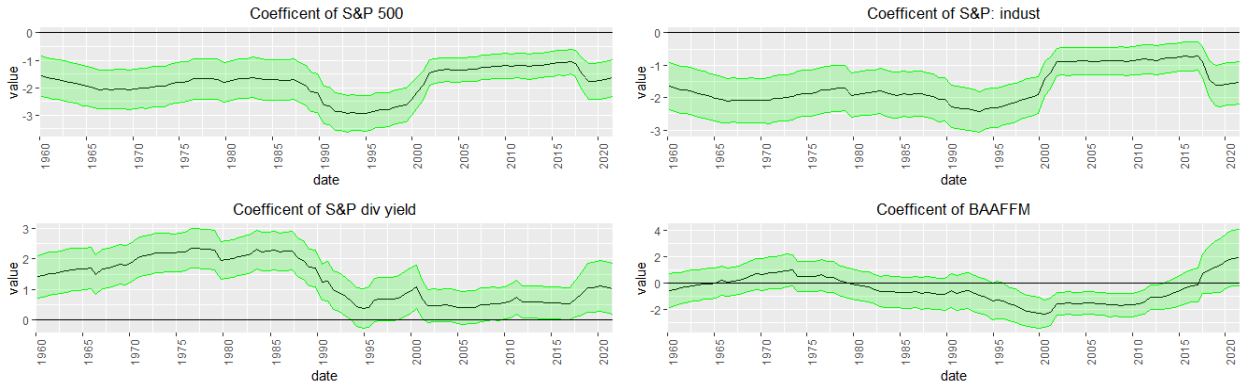

Figure 5: The estimated time-varying elements in the precision matrix linked to S&P PE ratio with 90% confident bands.

series and avoid directly imposing the sparsity assumption, we also extend the methodology and theory to a more general factor-adjusted time-varying VAR and network structures. Both the simulation and empirical studies show that the developed network model and methodology have reliable numerical performance in finite samples.

Supplementary materials

The supplement contains proofs of the main asymptotic theorems, some technical lemmas with proofs, verification of Assumption 3(ii) and discussions on tuning parameter selection.

References

BAI, J. AND NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 90, 191–221.

BARIGOZZI, M. AND BROWNLEES, C. (2019). NETS: Network estimation for time series. *Journal of Applied Econometrics* 34, 347–364.

BARIGOZZI, M., CHO, H. AND OWENS, D. (2022). FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series. Working paper available at https://arxiv.org/pdf/2201.06110.pdf.

BASU, S. AND MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* 43, 1535–1567.

BASU, S. SHOJAIE, A. AND MICHAILIDIS, G. (2015). Network Granger causality with inherent grouping structure. *Journal of Machine Learning Research* 16, 417–453.

BICKEL, P., RITOV, Y. AND TSYBAKOV, A. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37, 1705–1732.

BURT, R. S., KILDUFF, M. AND TASSELLI, S. (2013). Social network analysis: foundations and frontiers on advantage. *Annual Review of Psychology* 64, 527–547.

CAI, T. T., LIU, W. AND LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106, 594–607.

CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.

CHAMBERLAIN, G. AND ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51, 1305–1324.

CHEN, E., FAN, J. AND ZHU, X. (2020). Community network autoregression for high-dimensional time series. Working paper available at https://arxiv.org/abs/2007.05521.

CHEN, J., LI, D., WEI, L. AND ZHANG, W. (2021). Nonparametric homogeneity pursuit in functional-coefficient models. *Journal of Nonparametric Statistics* 33, 387–416.

CHENG, M., HONDA, T., LI, J. AND PENG, H. (2014). Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *The Annals of Statistics*, 42, 1819–1849.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* 25, 1–37.

DAHLHAUS, R. AND SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics* 34, 1075–1114.

DAVIS, R., ZANG, P. AND ZHENG, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* 25, 1077–1096.

DEMPSTER, A.P. (1972). Covariance selection. *Biometrics* 28, 157–175.

DIEBOLD, F. AND YILMAZ, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182, 119–134.

DIEBOLD, F. AND YILMAZ, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.

DING, X., QIU, Z. AND CHEN, X. (2017). Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electronic Journal of Statistics* 11, 3871–3902.

FAN, J., FENG, Y. AND WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *The Annals of Applied Statistics* 3, 521–541.

FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall.

FAN, J., MASINI, R. AND MEDEIROS, M. (2021). Bridging factor and sparse models. Working paper available at https://arxiv.org/abs/2102.11341.

FAN, J. AND LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1360.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* 75, 603–680.

FAN, J., MA, Y. AND DAI, W. (2014). Nonparametric independence screening in sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association* 109, 1270–1284.

FAN, J., XUE, L. AND ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics* 42, 819–849.

GRANGER, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

HAFNER, C. AND LINTON, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* 159, 55–73.

HAN, F., LU, H. AND LIU H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research* 16, 3115–3150.

HAUTSCH, N., SCHAUMBURG, J. AND SCHIENLE, M. (2014). Forecasting systemic impact in financial networks. *International Journal of Forecasting* 30, 781–794.

JANKOVA, J. AND VAN DE GEER S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* 9, 1205–1229.

KOCK, A.B. AND CALLOT, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* 186, 325–344.

KOLAR, M., SONG, L. AHMED, A. AND XING, E. (2010). Estimating time-varying networks. *The Annals of Applied Statistics* 4, 94–123.

KOO, B. AND LINTON O. (2012). Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics* 170, 210–233.

Krampe, J. and Margaritella, L. (2022). Factor models with sparse VAR idiosyncratic components. Working paper available at https://arxiv.org/pdf/2112.07149.pdf.

LAM, C. AND FAN, J. (2009). Sparsity and rates of convergence in large covariance matrix estimation. *The Annals of Statistics* 37, 4254–4278.

LI, D., CHEN, J. AND GAO, J. (2011). Nonparametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal* 14, 387–408.

LI, D., KE, Y. AND ZHANG, W. (2015). Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. *The Annals of Statistics* 43, 2676–2705.

LIAN, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statistica Sinica*, 22, 1563–1588.

LIU, J., LI, R. AND WU, R. (2014). Feature selection for varying coefficient models with ultrahigh dimensional covariates. *Journal of the American Statistical Association*, 109, 266–274.

LOH, P. AND WAINWRIGHT, M. (2013). Structural estimation for discrete graphical models: Generalized covariance matrices and their inverse. *The Annals of Statistics* 41, 3022–3049.

LIU, L. AND ZHANG, D. (2021). Robust estimation of high-dimensional vector autoregressive models. Working paper available at https://arxiv.org/abs/2109.10354.

LÜTKEPOHL, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.

MCCRACKEN, M.W. AND NG, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34, 574–589.

MCCRACKEN, M.W. AND 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. AND SU, L. (2022). High-dimensional VARs with common factors. Forthcoming in *Journal of Econometrics*.

Motta, G., Hafner, C. and 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. *Physics Review, Series E* 66, 016128.

SAFIKHANI, A. AND SHOJAIE, A. (2022). Joint structural break detection and parameter estimation in high-dimensional non-stationary VAR models. *Journal of the American Statistical Association* 117, 251–264.

SCOTT, J. (2017). *Social Network Analysis* (4th Edition). Sage, London.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

SU, L. AND WANG, X. (2017). On time-varying factor models: estimation and testing. *Journal of 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. *The Annals of Statistics* 40, 2601–2633.

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics.

WANG, L., LI, H. AND HUANG, J. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association* 103, 1556–1569.

WANG, H. AND XIA, Y. (2009). Shrinkage estimation of the varying-coefficient model. *Journal of the American Statistical Association* 104, 747–757.

WANG, D., YU, Y. AND RINALDO, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics* 49, 203–232.

XU, M., CHEN, X. AND WU, W. (2020). Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* 22, 55.

YAN, Y., GAO, J. AND 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. *Journal of Machine Learning Research* 11, 2261–2286.

YUAN, M. AND LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* 94, 19–35.

ZHANG, T. AND WU, W. B. (2012). Inference of time varying regression models. *The Annals of Statistics* 40, 1376–1402.

ZHANG, D. AND WU, W. (2021). Convergence of covariance and spectral density estimators for high-dimensional locally stationary processes. *The Annals of Statistics* 49, 233–254.

ZHAO, J., LIU, X., WANG, H. AND LENG, C. (2022). Dimension reduction for covariates in network data. *Biometrika* 109, 85–102.

ZHOU, S., LAFFERTY, J. AND WASSERMAN, L. (2010). Time varying undirected graphs. *Machine Learning* 80, 295–319.

ZHU, X., CHANG, X., LI, R. AND WANG, H. (2019). Portal nodes screening for large scale social networks. *Journal of Econometrics* 209, 145–157.

ZHU, X., PAN, R., LI, G., LIU, Y. AND WANG, H. (2017). Network vector autoregression. *The Annals of Statistics* 45, 1096–1123.

ZOU, H. AND LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *The Annals of Statistics*, 36, 1509–1566.

Supplement to "Estimating Time-Varying Networks for High-Dimensional Time Series"

Appendix A: Proofs of Theorems 4.1–4.3

Proof of Theorem 4.1. The main idea to be used in this proof is similar to that in Bickel, Ritov and Tsybakov (2009), Lian (2012) and Li, Ke and Zhang (2015) which study high-dimensional data under the classic independence assumption. In the following proof, we need to use the uniform convergence properties of the kernel-weighted quantities for time-varying VAR (say, Lemma B.3 in Appendix B). In fact, we next prove a strengthened version of (4.4) which also includes a uniform consistency of the derivative function estimates:

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left(\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\|+h\left\|\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)-\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\right\|\right)=O_P\left(\sqrt{s}\lambda_1\right). \tag{A.1}$$

As we only consider the time-varying VAR (1) model,

$$\boldsymbol{\alpha}_{i\bullet}(\tau_t)=[\alpha_{i,1}(\tau_t),\alpha_{i,2}(\tau_t),\cdots,\alpha_{i,d}(\tau_t)]^\tau \quad\text{and}\quad \boldsymbol{\alpha}'_{i\bullet}(\tau_t)=\left[\alpha'_{i,1}(\tau_t),\alpha'_{i,2}(\tau_t),\cdots,\alpha'_{i,d}(\tau_t)\right]^\tau.$$

Recall that $\mathcal{J}_i(\tau_t)=\{j:\ \alpha_{i,j}(\tau_t)\neq 0\}$ and define $\mathcal{J}'_i(\tau_t)=\left\{j:\ \alpha'_{i,j}(\tau_t)\neq 0\right\}$. We first prove that for any $i=1,\cdots,d$ and $t=1,\cdots,n$,

$$\sum_{j\notin\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\notin\mathcal{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\leqslant 2\left(\sum_{j\in\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\in\mathcal{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\right), \tag{A.2}$$

where $\delta_{i,j}(\tau_t)=\widetilde{\alpha}_{i,j}(\tau_t)-\alpha_{i,j}(\tau_t)$ and $\delta'_{i,j}(\tau_t)=h\left[\widetilde{\alpha}'_{i,j}(\tau_t)-\alpha'_{i,j}(\tau_t)\right]$.

By the definition of the preliminary time-varying LASSO, we have

$$\mathcal{L}_i^*\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t),\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)\mid\tau_t\right)\leqslant\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)$$

for any $i=1,\cdots,d$ and $t=1,\cdots,n$, where $\mathcal{L}_i^*(\boldsymbol{\alpha},\boldsymbol{\beta}\mid\tau_t)$ is defined in (3.4). Then, we readily have that

$$\begin{aligned}&\mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)-\mathcal{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t),\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)\mid\tau_t\right)\\&\geqslant\lambda_1\left[\sum_{j=1}^d|\widetilde{\alpha}_{i,j}(\tau_t)|+h\sum_{j=1}^d|\widetilde{\alpha}'_{i,j}(\tau_t)|-\sum_{j=1}^d|\alpha_{i,j}(\tau_t)|-h\sum_{j=1}^d|\alpha'_{i,j}(\tau_t)|\right].\end{aligned} \tag{A.3}$$

Let

$$\delta_i(\tau_t) = [\delta_{i,1}(\tau_t), \cdots, \delta_{i,d}(\tau_t)]^\tau \text{ and } \delta_i'(\tau_t) = \left[\delta_{i,1}'(\tau_t), \cdots, \delta_{i,d}'(\tau_t)\right]^\tau.$$

Note that

$$\begin{aligned}
& \mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}_{i\bullet}'(\tau_t) \mid \tau_t\right) - \mathcal{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t), \widetilde{\boldsymbol{\alpha}}_{i\bullet}'(\tau_t) \mid \tau_t\right) \\
= \; & 2\left[L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right] - \frac{1}{n}\sum_{s=1}^{n}\left\{\left[\delta_i(\tau_t) + \delta_i'(\tau_t)\left(\frac{\tau_s - \tau_t}{h}\right)\right]^\tau X_{s-1}\right\}^2 K_h(\tau_s - \tau_t) \\
\leqslant \; & 2\left[L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right], \qquad \text{(A.4)}
\end{aligned}$$

where $L_{i,0}(\tau_t)$ and $L_{i,1}(\tau_t)$ are defined in Appendix B. By Lemma B.3, we may show that

$$\left|L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right| \leqslant O_P\left(\zeta_{n,d}\right) \cdot \left(\sum_{j=1}^{d}|\delta_{i,j}(\tau_t)| + \sum_{j=1}^{d}|\delta_{i,j}'(\tau_t)|\right) \qquad \text{(A.5)}$$

uniformly over $i = 1, \cdots, d$ and $t = 1, \cdots, n$.

On the other hand, by the triangle inequality, we may prove that

$$\begin{aligned}
& \lambda_1\left[\sum_{j=1}^{d}|\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j=1}^{d}|\widetilde{\alpha}_{i,j}'(\tau_t)| - \sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)| - h\sum_{j=1}^{d}|\alpha_{i,j}'(\tau_t)|\right] \\
= \; & \lambda_1\left[\sum_{j\in\mathcal{J}_i(\tau_t)}\left(|\widetilde{\alpha}_{i,j}(\tau_t)| - |\alpha_{i,j}(\tau_t)|\right) + h\sum_{j\in\mathcal{J}_i'(\tau_t)}\left(|\widetilde{\alpha}_{i,j}'(\tau_t)| - |\alpha_{i,j}'(\tau_t)|\right)\right] + \\
& \lambda_1\left[\sum_{j\notin\mathcal{J}_i(\tau_t)}|\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j\notin\mathcal{J}_i'(\tau_t)}|\widetilde{\alpha}_{i,j}'(\tau_t)|\right] \\
\geqslant \; & -\lambda_1\left(\sum_{j\in\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)| + \sum_{j\in\mathcal{J}_i'(\tau_t)}|\delta_{i,j}'(\tau_t)|\right) + \lambda_1\left(\sum_{j\notin\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)| + \sum_{j\notin\mathcal{J}_i'(\tau_t)}|\delta_{i,j}'(\tau_t)|\right). \qquad \text{(A.6)}
\end{aligned}$$

By (A.3)–(A.6) and the condition $\zeta_{n,d} = o(\lambda_1)$ in Assumption 3(i), we complete the proof of (A.2).

Let $u_1 = (u_{1,1}, \cdots, u_{1,d})^\tau$ and $u_2 = (u_{2,1}, \cdots, u_{2,d})^\tau$ be two d-dimensional vectors and

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

where M is a positive constant which may be sufficiently large. Note that for any $i = 1, \cdots, d$,

$t = 1, \cdots, n$, and $u \in \mathcal{B}_i(\tau_t; M)$,

$$\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, \boldsymbol{\alpha}'_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_2/h \mid \tau_t\right) - \mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}'_{i\bullet}(\tau_t) \mid \tau_t\right) = \sum_{k=1}^{3} \Xi_{i,k}(\tau_t), \quad \text{(A.7)}$$

where

$$\begin{aligned}
\Xi_{i,1}(\tau_t) &= \mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, \boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_2/h \mid \tau_t\right) - \mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}'_{i\bullet}(\tau_t) \mid \tau_t\right), \\
\Xi_{i,2}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d} |\alpha_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{1,j}| - \sum_{j=1}^{d} |\alpha_{i,j}(\tau_t)|\right), \\
\Xi_{i,3}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d} |h\alpha'_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{2,j}| - \sum_{j=1}^{d} |h\alpha'_{i,j}(\tau_t)|\right).
\end{aligned}$$

For $\Xi_{i,1}(\tau_t)$, it can be written as

$$\Xi_{i,1}(\tau_t) = -2\sqrt{s}\lambda_1 u^\intercal L_i(\tau_t) + s\lambda_1^2 u^\intercal \boldsymbol{\Psi}(\tau_t) u, \quad \text{(A.8)}$$

where $L_i(\tau) = \left[L_{i,0}^\intercal(\tau), L_{i,1}^\intercal(\tau)\right]^\intercal$, and $\boldsymbol{\Psi}(\tau)$ is defined in (4.2). By the definition of $\mathcal{B}_i(\tau_t; M)$, Lemma B.3 and the Cauchy-Schwarz inequality, we have

$$\max_{1\leqslant i\leqslant d} \left|\sqrt{s}\lambda_1 u^\intercal L_i(\tau_t)\right| = o_P\left(s\lambda_1^2\right) \cdot \|u\|. \quad \text{(A.9)}$$

By (A.8), (A.9) and the uniform restricted eigenvalue condition (4.3), when n is sufficiently large and M is chosen to be large enough, we have

$$\min_{1\leqslant i\leqslant d} \min_{1\leqslant t\leqslant n} \inf_{u\in\mathcal{B}_i(\tau_t;M)} u^\intercal \Xi_{i,1}(\tau_t) = s\lambda_1^2 u^\intercal \boldsymbol{\Psi}(\tau_t) u(1 + o_P(1)) > \frac{1}{2}\kappa_0 s\lambda_1^2 \|u\|^2, \ \text{w.p.a.1.} \quad \text{(A.10)}$$

We next consider $\Xi_{i,2}(\tau_t)$ and $\Xi_{i,3}(\tau_t)$. It is easy to show that

$$\begin{aligned}
\Xi_{i,2}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d} |\alpha_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{1,j}| - \sum_{j=1}^{d} |\alpha_{i,j}(\tau_t)|\right) \\
&= \lambda_1 \sum_{j\in\mathcal{J}_i(\tau_t)} \left[|\alpha_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{1,j}| - |\alpha_{i,j}(\tau_t)|\right] + \lambda_1 \sum_{j\notin\mathcal{J}_i(\tau_t)} |\sqrt{s}\lambda_1 u_{1,j}| \\
&= O\left(s\lambda_1^2\right) \cdot \|u_1\| + \lambda_1 \sum_{j\notin\mathcal{J}_i(\tau_t)} |\sqrt{s}\lambda_1 u_{1,j}| = O\left(s\lambda_1^2\right) \cdot \|u_1\|, \quad \text{(A.11)}
\end{aligned}$$

and similarly,

$$\Xi_{i,3}(\tau_t) = O\left(s\lambda_1^2\right) \cdot \|u_2\| + \lambda_1 \sum_{j \notin \mathcal{J}_i'(\tau_t)} |\sqrt{s}\lambda_1 u_{2,j}| = O\left(s\lambda_1^2\right) \cdot \|u_2\|, \tag{A.12}$$

uniformly over $i = 1, \cdots, d$ and $t = 1, \cdots, n$.

With (A.7) and (A.10)–(A.12), letting M be large enough, we can prove that the leading term of

$$\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, \boldsymbol{\alpha}_{i\bullet}'(\tau_t) + \sqrt{s}\lambda_1 u_2/h \mid \tau_t\right) - \mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}_{i\bullet}'(\tau_t) \mid \tau_t\right)$$

is positive uniformly over $i = 1, \cdots, d$ and $t = 1, \cdots, n$. Hence, we may find a local minimiser to $\mathcal{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau_t)$, denoted by $\left[\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t), h\widetilde{\boldsymbol{\alpha}}_{i\bullet}'(\tau_t)\right]$, in the interior of

$$\left\{\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, h\boldsymbol{\alpha}_{i\bullet}'(\tau_t) + \sqrt{s}\lambda_1 u_2\right) : \ u \in \mathcal{B}_i(\tau_t; M)\right\},$$

which, together with (A.2), completes the proof of (A.1). ■

Proof of Theorem 4.2. Define

$$\begin{aligned}
\mathbf{L}_{i,j}^{\alpha} &= \left[l_{i,j}^{\alpha}(\boldsymbol{\alpha}_{\bullet 1}, \boldsymbol{\beta}_{\bullet 1} \mid \tau_1), \cdots, l_{i,j}^{\alpha}(\boldsymbol{\alpha}_{\bullet n}, \boldsymbol{\beta}_{\bullet n} \mid \tau_n)\right]^{\tau}, \\
\mathbf{L}_{i,j}^{\beta} &= \left[l_{i,j}^{\beta}(\boldsymbol{\alpha}_{\bullet 1}, \boldsymbol{\beta}_{\bullet 1} \mid \tau_1), \cdots, l_{i,j}^{\beta}(\boldsymbol{\alpha}_{\bullet n}, \boldsymbol{\beta}_{\bullet n} \mid \tau_n)\right]^{\tau}, \\
\mathbf{P}_{i,j}^{\alpha} &= \left[p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \frac{\alpha_{j|1}}{\|\boldsymbol{\alpha}_j\|}, \cdots, p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \frac{\alpha_{j|n}}{\|\boldsymbol{\alpha}_j\|}\right]^{\tau}, \\
\mathbf{P}_{i,j}^{\beta} &= \left[p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right) \frac{\beta_{j|1}}{\|\boldsymbol{\beta}_j\|}, \cdots, p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right) \frac{\beta_{j|n}}{\|\boldsymbol{\beta}_j\|}\right]^{\tau},
\end{aligned}$$

where

$$\begin{aligned}
l_{i,j}^{\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) &= \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\tau} X_{t-1}\right\} x_{t-1,j} K_h(\tau_t - \tau), \\
l_{i,j}^{\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) &= \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\tau} X_{t-1}\right\} x_{t-1,j} \left(\frac{\tau_t - \tau}{h}\right) K_h(\tau_t - \tau).
\end{aligned}$$

From the KKT condition (e.g., Fan and Lv, 2011; Fan, Xue and Zou, 2014; Li, Ke and Zhang, 2015), the oracle estimate $\left(\widehat{\mathbf{A}}_i^o, \widehat{\mathbf{B}}_i^o\right)$ is the unique minimiser to the objective function $\mathcal{Q}_i(\mathbf{A}, \mathbf{B})$ if

$$\mathbf{L}_{i,j}^{\alpha} - \mathbf{P}_{i,j}^{\alpha} = \mathbf{0}_n \ \text{ for } \ j \in \mathcal{J}_i, \ \ \mathbf{L}_{i,j}^{\beta} - \mathbf{P}_{i,j}^{\beta} = \mathbf{0}_n \ \text{ for } \ j \in \mathcal{J}_i', \tag{A.13}$$

$$\max_{j \in \overline{\mathcal{J}}_i} \left\|\mathbf{L}_{i,j}^{\alpha}\right\| < \min_{j \in \overline{\mathcal{J}}_i} p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right), \ \ \max_{j \in \overline{\mathcal{J}}_i} \left\|\mathbf{L}_{i,j}^{\beta}\right\| < \min_{j \in \overline{\mathcal{J}}_i} p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right), \tag{A.14}$$

hold at $\mathbf{A} = \widehat{\mathbf{A}}_i^o$ and $\mathbf{B} = \widehat{\mathbf{B}}_i^o$, where $\mathbf{0}_n$ is an n-dimensional vector of zeros.

Note that the equalities in (A.13) automatically hold by the definition of the oracle estimates $\widehat{\mathbf{A}}_i^o$ and $\widehat{\mathbf{B}}_i^o$. It remains to prove (A.14). We next only show the proof of the first assertion in (A.14) as the proof of the second one is analogous. By Theorem 4.1 and the condition of $(ns)^{1/2}\lambda_1 = o(\lambda_2)$ in Assumption 4(i), we may show that $\min_{j\in\overline{\mathcal{J}}_i} p'_{\lambda_2}(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|) = \lambda_2$ *w.p.a.1*. Meanwhile, by Lemmas B.3 and B.4 as well as Assumption 4(i), we may prove that

$$\max_{j\in\overline{\mathcal{J}}_i} \left\|\mathbf{L}_{i,j}^{\alpha}\right\| = O_P\left(\sqrt{n}s\log(n\vee d)\zeta_{n,d}\right) = o_P(\lambda_2)$$

when $\mathbf{A} = \widehat{\mathbf{A}}_i^o$ and $\mathbf{B} = \widehat{\mathbf{B}}_i^o$, leading to the first assertion in (A.14). Then, the mean squared convergence result (4.8) follows from Lemma B.4. ■

Proof of Corollary 4.1. By Theorem 4.2 and Assumption 4(ii), we may show that

$$P\left(\min_{(i,j)\in\mathbb{E}_n^G}\sum_{t=1}^{n}\widehat{a}_{ij}^2(\tau_t) \geqslant a_0\lambda_2 > 0\right) \to 1$$

and

$$P\left(\sum_{t=1}^{n}\widehat{a}_{ij}^2(\tau_t) = 0,\ \forall\,(i,j)\notin\mathbb{E}_n^G\right) \to 1,$$

leading to (4.9). ■

Proof of Theorem 4.3. By Lemma B.5 in Appendix B, we have

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} = O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right). \tag{A.15}$$

By (A.15), the sparsity assumption (3.7) and the inequality: $\|\mathbf{W}_1\mathbf{W}_2\|_{max} \leqslant \|\mathbf{W}_1\|_1\|\mathbf{W}_2\|_{max}$ for any two square matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ with the same size,

$$\begin{aligned}
\sup_{0\leqslant\tau\leqslant1}\left\|\mathbf{I}_d - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{max} &= \sup_{0\leqslant\tau\leqslant1}\left\|\boldsymbol{\Sigma}(\tau)\boldsymbol{\Omega}(\tau) - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{max} \\
&\leqslant \sup_{0\leqslant\tau\leqslant1}\|\boldsymbol{\Omega}(\tau)\|_1\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} \\
&\leqslant C_2\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} \\
&= O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right),
\end{aligned} \tag{A.16}$$

where C_2 is defined in (3.7). By (A.16), the triangle inequality, Assumption 5(ii) and the definition

of the time-varying CLIME estimate, we readily have that

$$\begin{aligned}
&\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\widetilde{\boldsymbol{\Omega}}(\tau)-\mathbf{I}_d\right\|_{\max}+\sup_{0\leqslant\tau\leqslant1}\left\|\mathbf{I}_d-\widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{\max}\\
\leqslant\ &\lambda_3+O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)=O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right). \qquad (A.17)
\end{aligned}$$

By Lemma 1 in Cai, Liu and Luo (2011), $\left\|\widetilde{\boldsymbol{\Omega}}(\tau)\right\|_1\leqslant\|\boldsymbol{\Omega}(\tau)\|_1\leqslant C_2$ uniformly over $0\leqslant\tau\leqslant1$. Then, by (A.16) and (A.17), we readily have that

$$\begin{aligned}
&\sup_{0\leqslant\tau\leqslant1}\left\|\boldsymbol{\Sigma}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}+\sup_{0\leqslant\tau\leqslant1}\left\|\left[\widehat{\boldsymbol{\Sigma}}(\tau)-\boldsymbol{\Sigma}(\tau)\right]\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)+2C_2\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)-\boldsymbol{\Sigma}(\tau)\right\|_{\max}=O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right). \qquad (A.18)
\end{aligned}$$

Using the assumption $\|\boldsymbol{\Omega}(\tau)\|_1\leqslant C_2$ again and (A.18), we have

$$\begin{aligned}
\sup_{0\leqslant\tau\leqslant1}\left\|\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right\|_{\max}&\leqslant\sup_{0\leqslant\tau\leqslant1}\|\boldsymbol{\Omega}(\tau)\|_1\left\|\boldsymbol{\Sigma}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
&=O_P\left(\nu^{*}_{n,d}+\nu^{\diamond}_{n,d}\right). \qquad (A.19)
\end{aligned}$$

By (A.19) and the definition of $\widehat{\boldsymbol{\Omega}}(\tau)$ in (3.10), we prove (4.10).

We next give the proof of (4.11). By Lemma 1 in Cai, Liu and Luo (2011), we have

$$\sum_{i=1}^{d}|\widehat{\omega}_{ij}(\tau)|\leqslant\sum_{i=1}^{d}|\widetilde{\omega}_{ij}(\tau)|\leqslant\sum_{i=1}^{d}|\omega_{ij}(\tau)|.$$

Noting that

$$\begin{aligned}
\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|\leqslant\lambda_3\right)&=\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|-\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)\\
&\leqslant\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|-\sum_{j=1}^{d}|\omega_{ij}(\tau)|+\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)-\omega_{ij}(\tau)|\\
&\leqslant\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)-\omega_{ij}(\tau)|,
\end{aligned}$$

we have

$$\begin{aligned}
\sup_{0\leqslant\tau\leqslant 1}\left\|\widehat{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\| &\leqslant \sup_{0\leqslant\tau\leqslant 1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)| \\
&\leqslant 2\sup_{0\leqslant\tau\leqslant 1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)+ \\
&\quad 2\sup_{0\leqslant\tau\leqslant 1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|\leqslant\lambda_3\right) \\
&=: \Delta_1+\Delta_2. \qquad (A.20)
\end{aligned}$$

Define an event

$$\mathcal{E}_\epsilon=\left\{\sup_{0\leqslant\tau\leqslant 1}\left\|\widehat{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\|_{\max}\leqslant c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)\right\},$$

where c_ϵ is a positive constant such that $\mathsf{P}\left(\mathcal{E}_\epsilon\right)\geqslant 1-\epsilon$ with any $\epsilon>0$. Conditional on $\mathcal{E}_\epsilon$,

$$\Delta_1\leqslant c_\epsilon(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*})\sup_{0\leqslant\tau\leqslant 1}\left[\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)\right]. \qquad (A.21)$$

Note that on $\mathcal{E}$,

$$|\widehat{\omega}_{ij}(\tau)|\leqslant|\omega_{ij}(\tau)|+|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\leqslant|\omega_{ij}(\tau)|+c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right).$$

Choosing $C_3=2c_\epsilon$ in Assumption 5(ii), the event $\{|\widehat{\omega}_{ij}(\tau)|>\lambda_3\}$ implies that $\left\{|\omega_{ij}(\tau)|>c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)\right\}$ holds. Then, by (3.7) and (A.21), we may show that on $\mathcal{E}_\epsilon$,

$$\begin{aligned}
\Delta_1 &\leqslant c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)\left[\sup_{0\leqslant\tau\leqslant 1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}I\left(|\omega_{ij}(\tau)|>c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)\right)\right] \\
&\leqslant c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)\left[\sup_{0\leqslant\tau\leqslant 1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}\frac{|\omega_{ij}(\tau)|^q}{c_\epsilon^q\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)^q}\right] \\
&= O_P\left(\xi_d\cdot\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)^{1-q}\right). \qquad (A.22)
\end{aligned}$$

On the other hand, by the triangle inequality,

$$|\widehat{\omega}_{ij}(\tau)|\geqslant|\omega_{ij}(\tau)|-|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\geqslant|\omega_{ij}(\tau)|-c_\epsilon\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)$$

on $\mathcal{E}_\epsilon$. Hence, we readily show that $\{|\widehat{\omega}_{ij}(\tau)| \leqslant \lambda_3\}$ indicates $\left\{|\omega_{ij}(\tau)| \leqslant 3c_\epsilon \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)\right\}$. Then, by (3.7) again, we have

$$\begin{aligned}
\Delta_2 &\leqslant \sup_{0\leqslant\tau\leqslant 1} \max_{1\leqslant i\leqslant d} \sum_{j=1}^{d} |\omega_{ij}(\tau)| \, I\left(|\omega_{ij}(\tau)| \leqslant 3c_\epsilon \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)\right) \\
&\leqslant (3c_\epsilon)^{1-q} \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q} \sup_{0\leqslant\tau\leqslant 1} \max_{1\leqslant i\leqslant d} \sum_{j=1}^{d} |\omega_{ij}(\tau)|^q \\
&= O_P\left(\xi_d \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q}\right). \qquad \text{(A.23)}
\end{aligned}$$

The proof of (4.11) can be completed by (A.20), (A.22) and (A.23).

Following the proof of (4.11), we also have

$$\sup_{0\leqslant\tau\leqslant 1} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_1 = O_P\left(\xi_d \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q}\right),$$

which, together with the following inequalities:

$$\frac{1}{d}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_F^2 \leqslant \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{max} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_1,$$

leads to (4.12). The proof of Theorem 4.3 is completed. ■

Proof of Corollary 4.2. By (4.10) in Theorem 4.3 and the condition of $\min_{(i,j)\in\mathbb{E}^P} \min_{1\leqslant t\leqslant n} |\omega_{ij}(\tau_t)| \gg \lambda_3$, we have

$$P\left(\min_{(i,j)\in\mathbb{E}^P_n} \min_{1\leqslant t\leqslant n} |\widehat{\omega}_{ij}(\tau_t)| \geqslant \lambda_3 > 0\right) \to 1. \qquad \text{(A.24)}$$

Letting $\mathcal{E}_\epsilon$ and c_ϵ be defined as in the proof of Theorem 4.3 and choosing $C_3 = 2c_\epsilon$ in Assumption 5(ii), we may prove that

$$\max_{(i,j)\notin\mathbb{E}^P_n} \max_{1\leqslant t\leqslant n} |\widehat{\omega}_{ij}(\tau_t)| \leqslant c_\epsilon(\nu^{*}_{n,d} + \nu^{\diamond}_{n,d}) < \lambda_3 \qquad \text{(A.25)}$$

conditional on $\mathcal{E}_\epsilon$. By virtue of (A.24) and (A.25), letting $\epsilon \to 0$, we prove (4.13). ■

Appendix B: Technical lemmas

In this appendix, we give some technical lemmas which are crucial to proofs of the main theoretical results in Appendix A. Without loss of generality, we focus on the time-varying VAR (1) model

framework. Throughout the proofs, we let M denote a generic positive constant whose value may change from line to line.

Lemma B.1. *Suppose that Assumption 1 is satisfied. Let*

$$\iota_2 = \iota_1/C_*, \;\; \iota_3 = \iota_1(1-\rho)/(C_1^2 C_*), \;\; C_* = \max_{1\leqslant t\leqslant n} \|\mathbf{\Sigma}_t\| < \infty$$

where ι_1 and ρ are defined in Assumption 1, and C_1 is defined in (2.4). For any d-dimensional vector u satisfying $\|u\| = 1$,

$$\max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau e_t\right)^2\right\}\right] \leqslant C_0 < \infty, \tag{B.1}$$

and

$$\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_3 x_{t,i}^2\right\}\right] \leqslant C_0^{1/(1-\rho)} < \infty, \tag{B.2}$$

where C_0 is a positive constant defined in Assumption 1(iii).

Proof of Lemma B.1. Writing $u_t^\tau = u^\tau \mathbf{\Sigma}_t^{1/2}$ and using Assumption 1(ii)(iii), we may show that

$$\begin{aligned}
\max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau e_t\right)^2\right\}\right] &= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau \mathbf{\Sigma}_t^{1/2}\varepsilon_t\right)^2\right\}\right] \\
&= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \|u_t\|^2 \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \\
&\leqslant \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 C_* \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \\
&= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_1 \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \leqslant C_0,
\end{aligned}$$

completing the proof of (B.1).

By the time-varying linear process representation (2.3), we have

$$x_{t,i}^2 = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \left(\Phi_{t,k_1,i}^\tau e_{t-k_1}\right)\left(\Phi_{t,k_2,i}^\tau e_{t-k_2}\right)$$

where $\Phi_{t,k,i}^\tau$ is the i-th row vector of $\mathbf{\Phi}_{t,k}$. Without loss of generality, assume (2.4) for all $k \geqslant 0$. Letting $u_{t,k,i} = \Phi_{t,k,i}/\|\Phi_{t,k,i}\|$ and noting that

$$\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \|\Phi_{t,k,i}\| \leqslant \max_{1\leqslant t\leqslant n} \|\mathbf{\Phi}_{t,k}\| \leqslant C_1 \rho^k,$$

we may show that

$$\begin{aligned} x_{t,i}^2 &\leqslant C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1} \sum_{k_2=0}^{\infty} \rho^{k_2} \left| \left(u_{t,k_1,i}^\intercal e_{t-k_1}\right) \left(u_{t,k_2,i}^\intercal e_{t-k_2}\right) \right| \\ &\leqslant C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1} \sum_{k_2=0}^{\infty} \rho^{k_2} \left(u_{t,k_2,i}^\intercal e_{t-k_2}\right)^2 \\ &= \frac{C_1^2}{1-\rho} \sum_{k=0}^{\infty} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2, \end{aligned}$$

which, together with the independence assumption over e_t and (A.1), indicates that

$$\begin{aligned} \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_3 x_{t,i}^2\right\}\right] &\leqslant \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\frac{\iota_3 C_1^2}{1-\rho} \sum_{k=0}^{\infty} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &= \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \prod_{k=0}^{\infty} \mathsf{E}\left[\exp\left\{\frac{\iota_3 C_1^2}{1-\rho} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &= \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \prod_{k=0}^{\infty} \mathsf{E}\left[\exp\left\{\iota_2 \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &\leqslant \prod_{k=0}^{\infty} \left(\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_2 \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right]\right)^{\rho^k} \\ &\leqslant \prod_{k=0}^{\infty} C_0^{\rho^k} = C_0^{1/(1-\rho)}, \end{aligned}$$

completing the proof of (B.2). ■

The following lemma is a well-known Bernstein-type inequality for martingale differences (e.g., Freedman, 1975; de la Peña, 1999).

Lemma B.2. *Let $(z_t, \mathcal{F}_t)_{t\geqslant 1}$ be a sequence of martingale differences and $\sigma_n^2 = \sum_{t=1}^n \mathsf{E}(z_t^2|\mathcal{F}_{t-1})$. Suppose that there exists a constant $a > 0$ such that $\mathsf{P}(|z_t| \leqslant a|\mathcal{F}_{t-1}) = 1$ for all $t \geqslant 2$. Then, for all $x, y > 0$,*

$$\mathsf{P}\left(\sum_{t=1}^n z_t > x,\ \sigma_n^2 \leqslant y\right) \leqslant \exp\left\{-\frac{x^2}{2(y+ax)}\right\}.$$

Define

$$L_{i,0}(\tau) = \frac{1}{n}\sum_{t=1}^n e_{t,i}(\tau) X_{t-1} K_h(\tau_t - \tau) \text{ and } L_{i,1}(\tau) = \frac{1}{n}\sum_{t=1}^n e_{t,i}(\tau) X_{t-1} \left(\frac{\tau_t - \tau}{h}\right) K_h(\tau_t - \tau),$$

where $e_{t,i}(\tau) = x_{t,i} - [\boldsymbol{\alpha}_{i\bullet}(\tau) + \boldsymbol{\alpha}'_{i\bullet}(\tau)(\tau_t - \tau)]^\intercal X_{t-1}$. Lemma B.3 below gives the uniform asymptotic orders for the kernel-weighted quantities $L_{i,k}(\cdot)$, $k = 0, 1$.

Lemma B.3. *Suppose that Assumptions 1 and 2 are satisfied. Then we have*

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}|L_{i,k}(\tau_t)|_{\max} = O_P\left(\zeta_{n,d}\right), \quad k = 0, 1, \tag{B.3}$$

where $\zeta_{n,d} = \log(n\vee d)\left[(nh)^{-1/2} + sh^2\right]$ *as in Assumption3(i).*

Proof of Lemma B.3. We only prove (B.3) for $k = 0$ as the proof is analogous for $k = 1$. Noting that

$$e_{l,i}(\tau_t) = e_{l,i} + [\boldsymbol{\alpha}_{i\bullet}(\tau_l) - \boldsymbol{\alpha}_{i\bullet}(\tau_t) - \boldsymbol{\alpha}'_{i\bullet}(\tau_t)(\tau_l - \tau_t)]^\intercal X_{l-1} =: e_{l,i} + b_{l,i}^\intercal(\tau_t)X_{l-1},$$

we write

$$L_{i,0}(\tau_t) = \frac{1}{n}\sum_{l=1}^{n} e_{l,i}X_{l-1}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n} b_{l,i}^\intercal(\tau_t)X_{l-1}X_{l-1}K_h(\tau_l - \tau_t).$$

In order to prove (B.3) with $k = 0$, it is sufficient to show that

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n} e_{l,i}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left((nh)^{-1/2}\log(n\vee d)\right) \tag{B.4}$$

and

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n} b_{l,i}^\intercal(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left(sh^2\log(n\vee d)\right). \tag{B.5}$$

Define

$$\overline{e}_{l,i} = e_{l,i}I\left(|e_{l,i}| \leqslant 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right), \quad \widetilde{e}_{l,i} = e_{l,i} - \overline{e}_{l,i},$$

and

$$\overline{x}_{l,i} = x_{l,i}I\left(|x_{l,i}| \leqslant 2\sqrt{\iota_3^{-1}\log(n\vee d)}\right), \quad \widetilde{x}_{l,i} = x_{l,i} - \overline{x}_{l,i},$$

where ι_2 and ι_3 are defined in Lemma B.1. Then, we have the following decomposition:

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n} e_{l,i}x_{l-1,j}K_h(\tau_l - \tau_t) \;=\; & \frac{1}{n}\sum_{l=1}^{n}\overline{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n}\overline{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l - \tau_t) + \\ & \frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l - \tau_t).\end{aligned}$$

By the Bonferroni and Markov inequalities as well as (B.1), for any $\epsilon > 0$, we have

$$\begin{aligned}
& \mathsf{P}\left(\max_{1\leqslant i\leqslant d} \max_{1\leqslant j\leqslant d} \max_{1\leqslant t\leqslant n} \left| \frac{1}{n} \sum_{l=1}^{n} \widetilde{e}_{l,i} \overline{x}_{l-1,j} K_h(\tau_l - \tau_t) \right| > \epsilon (nh)^{-1/2} \log(n \vee d) \right) \\
\leqslant\ & \mathsf{P}\left(\max_{1\leqslant i\leqslant d} \max_{1\leqslant t\leqslant n} |e_{t,i}| > 2\sqrt{\iota_2^{-1} \log(n \vee d)} \right) \\
\leqslant\ & \sum_{i=1}^{d} \sum_{t=1}^{n} \mathsf{P}\left(|e_{t,i}| > 2\sqrt{\iota_2^{-1} \log(n \vee d)} \right) \\
\leqslant\ & \sum_{i=1}^{d} \sum_{t=1}^{n} (n \vee d)^{-4} \mathsf{E}\left(\exp\left\{ \iota_2 e_{t,i}^2 \right\} \right) \\
\leqslant\ & M(n \vee d)^{-2} = o(1). \qquad \text{(B.6)}
\end{aligned}$$

Hence, we have

$$\max_{1\leqslant i\leqslant d} \max_{1\leqslant j\leqslant d} \max_{1\leqslant t\leqslant n} \left| \frac{1}{n} \sum_{l=1}^{n} \widetilde{e}_{l,i} \overline{x}_{l-1,j} K_h(\tau_t - \tau_t) \right| = o_P\left((nh)^{-1/2} \log(n \vee d) \right). \qquad \text{(B.7)}$$

Following the proof of (B.7), we also have

$$\max_{1\leqslant i\leqslant d} \max_{1\leqslant j\leqslant d} \max_{1\leqslant t\leqslant n} \left| \frac{1}{n} \sum_{l=1}^{n} \overline{e}_{l,i} \widetilde{x}_{l-1,j} K_h(\tau_l - \tau_t) \right| = o_P\left((nh)^{-1/2} \log(n \vee d) \right) \qquad \text{(B.8)}$$

and

$$\max_{1\leqslant i\leqslant d} \max_{1\leqslant j\leqslant d} \max_{1\leqslant t\leqslant n} \left| \frac{1}{n} \sum_{l=1}^{n} \widetilde{e}_{l,i} \widetilde{x}_{l-1,j} K_h(\tau_l - \tau_t) \right| = o_P\left((nh)^{-1/2} \log(n \vee d) \right). \qquad \text{(B.9)}$$

By the Cauchy-Schwarz and Markov inequalities and (B.1), we may show that

$$\begin{aligned}
\mathsf{E}\left(|\widetilde{e}_{l,i}| \right) &\leqslant \left[\mathsf{E}\left(|e_{l,i}|^2 \right) \right]^{1/2} \left[\mathsf{P}\left(|e_{l,i}| > 2\sqrt{\iota_2^{-1} \log(n \vee d)} \right) \right]^{1/2} \\
&= \left[\mathsf{E}\left(|e_{l,i}|^2 \right) \right]^{1/2} \left[\mathsf{P}\left(\exp\left\{ \iota_2 e_{l,i}^2 \right\} > (n \vee d)^4 \right) \right]^{1/2} \\
&\leqslant \left[\mathsf{E}\left(|e_{l,i}|^2 \right) \right]^{1/2} \left[\mathsf{E}\left(\exp\left\{ \iota_2 e_{l,i}^2 \right\} \right) \right]^{1/2} (n \vee d)^{-2} \\
&\leqslant M(n \vee d)^{-2},
\end{aligned}$$

which, together with the definition of $\overline{x}_{l-1,j}$ and the condition on the kernel function, indicates that

$$\begin{aligned}\left|\frac{1}{n}\sum_{l=1}^{n}\mathsf{E}\left[\overline{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathcal{F}_{l-1}(X)\right]\right| &= \left|\frac{1}{n}\sum_{l=1}^{n}\mathsf{E}\left[\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathcal{F}_{l-1}(X)\right]\right| \\ &= O_P\left((n\vee d)^{-2}\sqrt{\log(n\vee d)}\right) \\ &= o_P\left((nh)^{-1/2}\log(n\vee d)\right), \end{aligned} \tag{B.10}$$

where $\mathcal{F}_l(X)=\sigma(X_t:\ t\leqslant l)$. With (B.7)–(B.10), we readily have that

$$\frac{1}{n}\sum_{l=1}^{n}e_{l,i}x_{l-1,j}K_h(\tau_l-\tau_t)=\frac{1}{n}\sum_{l=1}^{n}\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathcal{F}_{l-1}(X)\right]\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+o_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.11}$$

By the Bonferroni inequality and the Bernstein inequality in Lemma B.2, we prove that

$$\begin{aligned}&\mathsf{P}\left(\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathcal{F}_{l-1}(X)\right]\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|>M_0(nh)^{-1/2}\log(n\vee d)\right) \\ \leqslant\ &\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{t=1}^{n}\mathsf{P}\left(\left|\frac{1}{n}\sum_{l=1}^{n}\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathcal{F}_{l-1}(X)\right]\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|>M_0(nh)^{-1/2}\log(n\vee d)\right) \\ \leqslant\ &\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{t=1}^{n}\exp\{-g_0(M_0)\log(n\vee d)\}=O\left(nd^2(n\vee d)^{-g_0(M_0)}\right)=o(1),\end{aligned}$$

letting $M_0>0$ be sufficiently large, where $g_0(\cdot)$ is a positive function satisfying $g_0(z)\to\infty$ as $z\to+\infty$. Consequently, we have

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathcal{F}_{l-1}(X)\right]\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|=O_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.12}$$

By virtue of (B.11) and (B.12), we complete the proof of (B.4).

Letting $\overline{X}_l=(\overline{x}_{l,1},\cdots,\overline{x}_{l,d})^\tau$ and $\widetilde{X}_l=(\widetilde{x}_{l,1},\cdots,\widetilde{x}_{l,d})^\tau$, we have the following decomposition:

$$\begin{aligned}&\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^\tau(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l-\tau_t) \\ =\ &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^\tau(\tau_t)\overline{X}_{l-1}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^\tau(\tau_t)\overline{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t) \\ &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^\tau(\tau_t)\widetilde{X}_{l-1}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^\tau(\tau_t)\widetilde{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t).\end{aligned} \tag{B.13}$$

Similarly to the proof of (B.11), we may show that the last three terms on the right side of (B.13) are of order $o_P\left(h^2\log(n\vee d)\right)$. By Assumption 1(i) and the Taylor expansion of $\boldsymbol{\alpha}_{i\bullet}(\cdot)$, we can prove that the first term on the right side of (B.13) is of order $O_P\left(sh^2\log(n\vee d)\right)$ uniformly over i,j,t. The proof of (B.5) is completed. ■

Lemma B.4 below gives the mean squared convergence rates of the infeasible oracle estimates $\widehat{\mathbf{A}}_i^o$ and $\widehat{\mathbf{B}}_i^o$ defined in (4.6) and (4.7) of Section 4.2.

Lemma B.4. *Suppose Assumptions 1–4 are satisfied. Then we have*

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}^o(\tau_t)-\alpha_{i,j}(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2\right), \tag{B.14}$$

and

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}'^o(\tau_t)-\alpha_{i,j}'(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2h^{-2}\right). \tag{B.15}$$

Proof of Lemma B.4. For any $1\leqslant i\leqslant d$, let

$$\mathbf{U}^o=\left[(v_1^o)^\intercal,(w_1^o)^\intercal,(v_2^o)^\intercal,(w_2^o)^\intercal,\cdots,(v_n^o)^\intercal,(w_n^o)^\intercal\right]^\intercal,$$

where $v_t^o=\left(v_{1|t}^o,\cdots,v_{d|t}^o\right)^\intercal$ with $v_{j|t}^o=0$ for $j\in\overline{\mathcal{J}}_i$, and $w_t^o=\left(w_{1|t}^o,\cdots,w_{d|t}^o\right)^\intercal$ with $w_{j|t}^o=0$ for $j\in\overline{\mathcal{J}}_i'$. Define

$$\mathcal{B}_i^*(M_*)=\left\{\mathbf{U}^o:\ \sum_{t=1}^{n}\left(\|v_t^o\|^2+\|w_t^o\|^2\right)=\|\mathbf{V}^o\|^2+\|\mathbf{W}^o\|^2=nM_*\right\},$$

where M_* is a positive constant which can be sufficiently large,

$$\mathbf{V}^o=\left[(v_1^o)^\intercal,(v_2^o)^\intercal,\cdots,(v_n^o)^\intercal\right]^\intercal\ \text{ and }\ \mathbf{W}^o=\left[(w_1^o)^\intercal,(w_2^o)^\intercal,\cdots,(w_n^o)^\intercal\right]^\intercal.$$

Write

$$\begin{aligned}\mathbf{A}_i&=(\boldsymbol{\alpha}_{i,1},\cdots,\boldsymbol{\alpha}_{i,d})\ \text{ with }\ \boldsymbol{\alpha}_{i,j}=[\alpha_{i,j}(\tau_1),\cdots,\alpha_{i,j}(\tau_n)]^\intercal,\\ \mathbf{B}_i&=(\boldsymbol{\alpha}_{i,1}',\cdots,\boldsymbol{\alpha}_{i,d}')\ \text{ with }\ \boldsymbol{\alpha}_{i,j}'=\left[\alpha_{i,j}'(\tau_1),\cdots,\alpha_{i,j}'(\tau_n)\right]^\intercal,\end{aligned}$$

as the matrices of true time-varying parameters. Observe that

$$\mathcal{Q}_i\left(\mathbf{A}_i+\sqrt{\zeta_{n,d}^*}\mathbf{V}^o,\mathbf{B}_i+\sqrt{\zeta_{n,d}^*}\mathbf{W}^o/h\right)-\mathcal{Q}_i\left(\mathbf{A}_i,\mathbf{B}_i\right)=\Pi_{i,1}^o+\Pi_{i,2}^o+\Pi_{n,3}^o, \tag{B.16}$$

where $\zeta_{n,d}^* = s\zeta_{n,d}^2$,

$$\begin{aligned}
\Pi_{i,1}^o &= \sum_{t=1}^{n} \left[\mathcal{L}_i \left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{\zeta_{n,d}^*} v_t^o, \boldsymbol{\alpha}'_{i\bullet}(\tau_t) + \sqrt{\zeta_{n,d}^*} w_t^o / h \mid \tau_t \right) - \mathcal{L}_i \left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}'_{i\bullet}(\tau_t) \mid \tau_t \right) \right], \\
\Pi_{i,2}^o &= \sum_{j=1}^{d} p'_{\lambda_2} \left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\| \right) \left(\left\| \boldsymbol{\alpha}_{i,j} + \sqrt{\zeta_{n,d}^*} \mathbf{v}_j^o \right\| - \|\boldsymbol{\alpha}_{i,j}\| \right), \\
\Pi_{i,3}^o &= \sum_{j=1}^{d} p'_{\lambda_2} \left(\widetilde{D}_{i,j} \right) \left(\left\| h\boldsymbol{\alpha}'_{i,j} + \sqrt{\zeta_{n,d}^*} \mathbf{w}_j^o \right\| - \left\| h\boldsymbol{\alpha}'_{i,j} \right\| \right),
\end{aligned}$$

in which $\mathbf{v}_j^o = \left(v_{j|1}^o, \cdots, v_{j|n}^o \right)^\tau$ and $\mathbf{w}_j^o = \left(w_{j|1}^o, \cdots, w_{j|n}^o \right)^\tau$.

By the definition of the local linear objective function, we readily have

$$\Pi_{i,1}^o = -2\sqrt{\zeta_{n,d}^*} \sum_{t=1}^{n} \left[(v_t^o)^\tau, (w_t^o)^\tau \right] L_i(\tau_t) + \zeta_{n,d}^* \sum_{t=1}^{n} \left[(v_t^o)^\tau, (w_t^o)^\tau \right] \boldsymbol{\Psi}(\tau_t) \left[(v_t^o)^\tau, (w_t^o)^\tau \right]^\tau. \tag{B.17}$$

By the definition of $\mathcal{B}_i^*(M_*)$, Lemma B.3 and the Cauchy-Schwarz inequality, we prove

$$\left| \sum_{t=1}^{n} \left[(v_t^o)^\tau, (w_t^o)^\tau \right] L_i(\tau_t) \right| = O_P \left(\sqrt{\zeta_{n,d}^*} n^{1/2} \right) \cdot \|\mathbf{U}^o\| \tag{B.18}$$

uniformly over i. By the uniform restricted eigenvalue condition in Assumption 3(ii), we have

$$\sum_{t=1}^{n} \left[(v_t^o)^\tau, (w_t^o)^\tau \right] \boldsymbol{\Psi}(\tau_t) \left[(v_t^o)^\tau, (w_t^o)^\tau \right]^\tau \geqslant \kappa_0 \sum_{t=1}^{n} \left(\|v_t^o\|^2 + \|w_t^o\|^2 \right) = n\kappa_0 M_* \tag{B.19}$$

for $\mathbf{U}^o \in \mathcal{B}_i^*(M_*)$. Combining (B.17)–(B.19) and letting $M_* > 0$ be sufficiently large, we have

$$\min_{1 \leqslant i \leqslant d} \Pi_{i,1}^o \geqslant \kappa_0 \zeta_{n,d}^* \|\mathbf{U}^o\|^2 + O_P \left(\zeta_{n,d}^* n^{1/2} \right) \cdot \|\mathbf{U}^o\| > \frac{\kappa_0}{2} \zeta_{n,d}^* \|\mathbf{U}^o\|^2 \ \ w.p.a.1. \tag{B.20}$$

On the other hand, by Theorem 4.1 and Assumption 4(ii), we have

$$P \left(\min_{1 \leqslant i \leqslant d} \min_{j \in \mathcal{J}_i} \|\widetilde{\boldsymbol{\alpha}}_{i,j}\| > a_0 \lambda_2 \right) \to 1,$$

and

$$P \left(\min_{1 \leqslant i \leqslant d} \min_{j \in \mathcal{J}_i'} \widetilde{D}_{i,j} > a_0 \lambda_2 \right) \to 1.$$

As $\alpha_{i,j}(\tau_t) = 0$ and $u^o_{1,j} = 0$ for $j \in \overline{\mathcal{J}}_i(\tau_t)$, we thus have

$$\Pi_{i,2}(\tau_t) = \sum_{j \in \mathcal{J}_i(\tau_t)} p'_{\lambda_2}\left(|\widetilde{\alpha}_{i,j}(\tau_t)|\right)\left(\left|\alpha_{i,j}(\tau_t) + \sqrt{\zeta^*_{n,d}}(\tau_t) u^o_{1,j}\right| - |\alpha_{i,j}(\tau_t)|\right) = 0 \ \ w.p.a.1, \tag{B.21}$$

and similarly

$$\Pi_{i,3}(\tau_t) = \sum_{j \in \mathcal{J}_i(\tau_t)} p'_{\lambda_2}\left(\left|\widetilde{\alpha}'_{i,j}(\tau_t)\right|\right)\left(\left|h\alpha'_{i,j}(\tau_t) + \sqrt{\zeta^*_{n,d}}(\tau_t) u^o_{2,j}\right| - \left|h\alpha'_{i,j}(\tau_t)\right|\right) = 0 \ \ w.p.a.1. \tag{B.22}$$

Hence, by (B.20)–(B.22), letting $M_* > 0$ be large enough, we can prove that

$$\min_{1 \leqslant i \leqslant d}\left[\sup_{U^o \in \mathcal{B}^*_i(M_*)} Q_i\left(\mathbf{A}_i + \sqrt{\zeta^*_{n,d}}\mathbf{V}^o, \mathbf{B}_i + \sqrt{\zeta^*_{n,d}}\mathbf{W}^o/h\right) - Q_i\left(\mathbf{A}_i, \mathbf{B}_i\right)\right] > 0 \ \ w.p.a.1,$$

indicating that there exists a local minimiser $\left(\widehat{\mathbf{A}}^o_i, \widehat{\mathbf{B}}^o_i\right)$ in the interior of

$$\left\{\left(\mathbf{A}_i + \sqrt{\zeta^*_{n,d}}\mathbf{V}^o, \mathbf{B}_i + \sqrt{\zeta^*_{n,d}}\mathbf{W}^o/h\right) : \ \mathbf{U}^o \in \mathcal{B}^*_i(c_1)\right\}$$

for any $1 \leqslant i \leqslant d$. The proof of Lemma B.4 is completed. ■

Lemma B.5 below gives the uniform convergence rates for the time-varying volatility function estimates, a crucial result to prove uniform consistency of the time-varying CLIME estimates.

Lemma B.5. *Suppose that Assumptions 1–4 are satisfied. Then we have*

$$\max_{1 \leqslant i,j \leqslant d} \sup_{0 \leqslant \tau \leqslant 1} |\widehat{\sigma}_{ij}(\tau) - \sigma_{ij}(\tau)| = O_P\left(\nu^\diamond_{n,d} + \nu^*_{n,d}\right), \tag{B.23}$$

where $\sigma_{ij}(\tau)$ is the (i,j)-entry of $\mathbf{\Sigma}(\tau)$, $\nu^\diamond_{n,d}$ and $\nu^*_{n,d}$ are defined in Assumption 5(ii).

Proof of Lemma B.5. By the definition of $\widehat{\sigma}_{ij}(\tau)$ in (3.8), we have

$$\begin{aligned}\widehat{\sigma}_{ij}(\tau) - \sigma_{ij}(\tau) &= \left\{\frac{\sum_{t=1}^n \varpi_{n,t}(\tau) e_{t,i} e_{t,j}}{\sum_{t=1}^n \varpi_{n,t}(\tau)} - \sigma_{ij}(\tau)\right\} + \left\{\frac{\sum_{t=1}^n \varpi_{n,t}(\tau)\left(\widehat{e}_{t,i} - e_{t,i}\right) e_{t,j}}{\sum_{t=1}^n \varpi_{n,t}(\tau)} + \right. \\ &\quad \left. \frac{\sum_{t=1}^n \varpi_{n,t}(\tau) e_{t,i}\left(\widehat{e}_{t,j} - e_{t,j}\right)}{\sum_{t=1}^n \varpi_{n,t}(\tau)} + \frac{\sum_{t=1}^n \varpi_{n,t}(\tau)\left(\widehat{e}_{t,i} - e_{t,i}\right)\left(\widehat{e}_{t,j} - e_{t,j}\right)}{\sum_{t=1}^n \varpi_{n,t}(\tau)}\right\} \\ &=: \chi^\diamond_{ij}(\tau) + \chi^*_{ij}(\tau).\end{aligned} \tag{B.24}$$

We first prove that

$$\max_{1 \leqslant i,j \leqslant d} \sup_{0 \leqslant \tau \leqslant 1} \left|\chi^\diamond_{ij}(\tau)\right| = O_P\left(\nu^\diamond_{n,d}\right). \tag{B.25}$$

Note that

$$\chi_{ij}^{\diamond}(\tau) = \frac{\sum_{t=1}^{n} \varpi_{n,t}(\tau)\left[e_{t,i}e_{t,j} - \sigma_{ij}(\tau_t)\right]}{\sum_{t=1}^{n} \varpi_{n,t}(\tau)} + \frac{\sum_{t=1}^{n} \varpi_{n,t}(\tau)\sigma_{ij}(\tau_t)}{\sum_{t=1}^{n} \varpi_{n,t}(\tau)} - \sigma_{ij}(\tau).$$

By the Taylor expansion of $\sigma_{ij}(\cdot)$ and the definition of the local linear weights $\varpi_{n,t}(\tau)$, we have

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d} \sup_{0\leqslant\tau\leqslant 1} \left| \frac{\sum_{t=1}^{n} \varpi_{n,t}(\tau)\sigma_{ij}(\tau_t)}{\sum_{t=1}^{n} \varpi_{n,t}(\tau)} - \sigma_{ij}(\tau) \right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d} \sup_{0\leqslant\tau\leqslant 1} \left|\sigma_{ij}''(\tau)\right| \cdot \left| \frac{\sum_{t=1}^{n} (\tau_t - \tau)^2 \varpi_{n,t}(\tau)}{\sum_{t=1}^{n} \varpi_{n,t}(\tau)} \right| \\
\leqslant\ &M \sup_{0\leqslant\tau\leqslant 1} \left| \frac{\sum_{t=1}^{n} (\tau_t - \tau)^2 \varpi_{n,t}(\tau)}{\sum_{t=1}^{n} \varpi_{n,t}(\tau)} \right| = O\left(b^2\right). \qquad \text{(B.26)}
\end{aligned}$$

Let $\overline{e}_{t,i}$ and $\widetilde{e}_{t,j}$ be defined as in the proof of Lemma B.3. Then, we have

$$\begin{aligned}
\sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) e_{t,i}e_{t,j} \ =\ & \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) \overline{e}_{t,i}\overline{e}_{t,j} + \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) \overline{e}_{t,i}\widetilde{e}_{t,j} + \\
& \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) \widetilde{e}_{t,i}\overline{e}_{t,j} + \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) \widetilde{e}_{t,i}\widetilde{e}_{t,j}. \qquad \text{(B.27)}
\end{aligned}$$

Following the proof of (B.11), the first term on the right side of (B.27) is the asymptotic leading term. Consider covering the closed interval $[0,1]$ by some disjoint intervals $\mathcal{I}_k$, $k = 1, \cdots, N$, with the center τ_k^* and length $b^2[nb\log(n \vee d)]^{-1/2}$. By the Lipschitz continuity of $K(\cdot)$ in Assumption 2(i), we have

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d} \sup_{0\leqslant\tau\leqslant 1} \left| \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau}{b}\right) \left[\overline{e}_{t,i}\overline{e}_{t,j} - E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right] \right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d} \max_{1\leqslant k\leqslant N} \left| \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau_k^*}{b}\right) \left[\overline{e}_{t,i}\overline{e}_{t,j} - E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right] \right| + \\
&\max_{1\leqslant i,j\leqslant d} \max_{1\leqslant k\leqslant N} \sup_{\tau\in\mathcal{I}_k} \left| \frac{1}{nb} \sum_{k=1}^{n} \left[K\left(\frac{\tau_t - \tau}{b}\right) - K\left(\frac{\tau_t - \tau_k^*}{b}\right) \right] \left[\overline{e}_{t,i}\overline{e}_{t,j} - E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right] \right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d} \max_{1\leqslant k\leqslant N} \left| \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{\tau_t - \tau_k^*}{b}\right) \left[\overline{e}_{t,i}\overline{e}_{t,j} - E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right] \right| + O_P\left(\left[\frac{\log(n \vee d)}{nb}\right]^{1/2} \right) \qquad \text{(B.28)}
\end{aligned}$$

By the Bonferroni inequality and Lemma B.2 as well as the condition $nb/[\log(n \vee d)]^3 \to \infty$ in

Assumption 5(i), we may show that

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|>M_1\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right)\\
\leqslant\ &\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{N}\mathsf{P}\left(\left|\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|>M_1\left[nb\log(n\vee d)\right]^{1/2}\right)\\
=\ &O\left(d^2N\exp\{-g_1(M_1)\log(n\vee d)\}\right)=O\left(d^2N(n\vee d)^{g_1(M_1)}\right)=o(1),
\end{aligned}$$

where $M_1>0$ is sufficiently large and $g_1(\cdot)$ is a positive function satisfying that $g_1(z)\to\infty$ as $z\to+\infty$. Therefore, we have

$$\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.29}$$

Combining (B.28) and (B.29), we can prove that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.30}$$

By the definitions of $\overline{e}_{t,i}$ and $\widetilde{e}_{t,i}$, we have

$$\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)-\sigma_{ij}(\tau_t)=\mathsf{E}\left(\widetilde{e}_{t,i}\widetilde{e}_{t,j}\right)-\mathsf{E}\left(\overline{e}_{t,i}\widetilde{e}_{t,j}\right)-\mathsf{E}\left(\widetilde{e}_{t,i}\overline{e}_{t,j}\right).$$

Meanwhile, by the Cauchy-Schwarz and Markov inequalities and (B.1) in Lemma B.1,

$$\begin{aligned}
\mathsf{E}\left(|\overline{e}_{t,i}\widetilde{e}_{t,j}|\right)&\leqslant M\left[\mathsf{E}\left(\widetilde{e}_{t,j}^2\right)\right]^{1/2}\\
&\leqslant M\left[\mathsf{E}\left(|e_{t,i}|^4\right)\right]^{1/4}\left[\mathsf{P}\left(|e_{t,i}|>2\sqrt{\iota_2^{-1}\log(n\vee d)}\right)\right]^{1/4}\\
&\leqslant M\left[\mathsf{P}\left(\exp\left\{\iota_2 e_{t,i}^2\right\}>(n\vee d)^4\right)\right]^{1/4}\\
&\leqslant M\left[\mathsf{E}\left(\exp\left\{\iota_2 e_{t,i}^2\right\}\right)\right]^{1/4}(n\vee d)^{-1}\\
&\leqslant O\left((n\vee d)^{-1}\right)=o\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right),
\end{aligned}$$

and similarly,

$$\mathsf{E}\left(|\widetilde{e}_{t,i}\overline{e}_{t,j}|\right)+\mathsf{E}\left(|\widetilde{e}_{t,i}\widetilde{e}_{t,j}|\right)=o\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right).$$

Hence, we can prove that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)-\sigma_{ij}(\tau_t)\right]\right|=o_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.31}$$

With (B.27), (B.30) and (B.31), we can prove that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.32}$$

Analogously, we also have

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K_1\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.33}$$

Using (B.32), (B.33) and the definition of $\varpi_{n,t}(\tau)$, we may show that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right), \tag{B.34}$$

which, together with (B.26), leads to (B.25).

Using the arguments in the proof of Lemma B.4, we may prove that

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left\|\widehat{\alpha}_{i\bullet}^{o}(\tau_t)-\alpha_{i\bullet}(\tau_t)\right\|=O_P\left(\sqrt{s}\zeta_{n,d}\right), \tag{B.35}$$

which, together with (B.2) in Lemma B.1, indicates that

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}|\widehat{e}_{t,i}-e_{t,i}|=O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right). \tag{B.36}$$

By (B.25), (B.36) and the Cauchy-Schwarz inequality, letting $\varpi_{n,t}^{*}(\tau)=\varpi_{n,t}(\tau)/\sum_{t=1}^{n}\varpi_{n,t}(\tau)$, we can prove that

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi_{n,t}^{*}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)e_{t,j}\right| \\
\leqslant\ &\max_{1\leqslant j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left(\sum_{t=1}^{n}\left|\varpi_{n,t}^{*}(\tau)\right|e_{t,j}^{2}\right)^{1/2}\max_{1\leqslant i\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left(\sum_{t=1}^{n}\left|\varpi_{n,t}^{*}(\tau)\right|\left(\widehat{e}_{t,i}-e_{t,i}\right)^{2}\right)^{1/2} \\
=\ &O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right)=O_P\left(\nu_{n,d}^{*}\right).
\end{aligned} \tag{B.37}$$

Similarly, we can also show that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi^*_{n,t}(\tau)e_{t,i}\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_P\left(\nu^*_{n,d}\right) \tag{B.38}$$

and

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi^*_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_P\left(\left[\nu^*_{n,d}\right]^2\right)=o_P\left(\nu^*_{n,d}\right). \tag{B.39}$$

From (B.37)–(B.39), we readily have that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\chi^*_{ij}(\tau)\right|=O_P\left(\nu^*_{n,d}\right),$$

which, together with (B.24) and (B.25), completes the proof of Lemma B.5. ■

Appendix C: Proofs of Propositions 5.1 and 5.2

In this appendix, we provide proofs of the convergence properties for the factor-adjusted estimators stated in Propositions 5.1 and 5.2. Define

$$\widehat{L}_{i,0}(\tau)=\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_{t,i}(\tau)\widehat{X}_{t-1}K_h(\tau_t-\tau)\ \text{ and }\ \widehat{L}_{i,1}(\tau)=\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_{t,i}(\tau)\widehat{X}_{t-1}\left(\frac{\tau_t-\tau}{h}\right)K_h(\tau_t-\tau),$$

where $\widehat{X}_t=(\widehat{x}_{t,1},\cdots,\widehat{x}_{t,d})^\intercal$ is defined in (5.3) or (5.4), and $\widehat{e}_{t,i}(\tau)=\widehat{x}_{t,i}-[\boldsymbol{\alpha}_{i\bullet}(\tau)+\boldsymbol{\alpha}'_{i\bullet}(\tau)(\tau_t-\tau)]^\intercal\widehat{X}_{t-1}$. The following lemma extends Lemma B.3 to the factor-adjusted kernel-weighted quantities.

Lemma C.1. *Suppose that Assumptions 1, 2 and 6(i) are satisfied. Then we have*

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left|\widehat{L}_{i,k}(\tau_t)\right|_{\max}=O_P\left(\zeta^\dagger_{n,d}\right),\ \ k=0,1, \tag{C.1}$$

where $\zeta^\dagger_{n,d}=\zeta_{n,d}+[\log(n\vee d)]^{1/2}s\delta_X$ *as in Assumption 6(ii).*

Proof of Lemma C.1. As in the proof of Lemma B.3, we only consider $k=0$. As

$$\widehat{e}_{t,i}(\tau)=e_{t,i}(\tau)+(\widehat{x}_{t,i}-x_{t,i})+[\boldsymbol{\alpha}_{i\bullet}(\tau)+\boldsymbol{\alpha}'_{i\bullet}(\tau)(\tau_t-\tau)]^\intercal\left(X_{t-1}-\widehat{X}_{t-1}\right),$$

by Assumption 6(i), we may show that

$$\begin{aligned}\widehat{L}_{i,0}(\tau) &= L_{i,0}(\tau) + \frac{1}{n}\sum_{t=1}^{n}(\widehat{x}_{t,i} - x_{t,i})\,\widehat{X}_{t-1}K_h(\tau_t - \tau) + \\ &\quad \frac{1}{n}\sum_{t=1}^{n}[\alpha_{i\bullet}(\tau) + \alpha'_{i\bullet}(\tau)(\tau_t - \tau)]^{\intercal}\left(X_{t-1} - \widehat{X}_{t-1}\right)\widehat{X}_{t-1}K_h(\tau_t - \tau) - \\ &\quad \frac{1}{n}\sum_{t=1}^{n} e_{t,i}(\tau)\left(X_{t-1} - \widehat{X}_{t-1}\right)K_h(\tau_t - \tau) \\ &= L_{i,0}(\tau) + O_P\left([\log(n \vee d)]^{1/2} s\delta_X\right).\end{aligned}$$

Then, by Lemma B.3, we complete the proof of (C.1) for $k = 0$. ■

Write

$$\widehat{e}_t^{\dagger} = \left(\widehat{e}_{t,1}^{\dagger}, \cdots, \widehat{e}_{t,d}^{\dagger}\right)^{\intercal} = \widehat{X}_t - \widehat{A}_1^{\dagger}(\tau_t)\widehat{X}_{t-1}, \;\; \widehat{A}_1^{\dagger}(\tau_t) = \left[\widehat{\alpha}_{ij}^{\dagger}(\tau_t)\right]_{d\times d}.$$

Let $\widehat{\sigma}_{ij}^{\dagger}(\tau)$ be the factor-adjusted local linear estimate $\sigma_{ij}(\tau)$, i.e., replace $\widehat{e}_{t,i}$ by $\widehat{e}_{t,i}^{\dagger}$ in (3.8). The following lemma extends Lemma B.5 to the factor-adjusted volatility function estimate.

Lemma C.2. *Suppose that the assumptions of Proposition 5.1(iii) are satisfied. Then we have*

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\widehat{\sigma}_{ij}^{\dagger}(\tau) - \sigma_{ij}(\tau)\right| = O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{\dagger}\right), \tag{C.2}$$

where $\nu_{n,d}^{\diamond}$ is defined in Assumption 5(ii) and $\nu_{n,d}^{\dagger}$ is defined in Assumption 6(iv).

Proof of Lemma C.2. As in (B.24), we have

$$\begin{aligned}\widehat{\sigma}_{ij}^{\dagger}(\tau) - \sigma_{ij}(\tau) &= \left\{\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)e_{t,i}e_{t,j}}{\sum_{t=1}^{n}\omega_{n,t}(\tau)} - \sigma_{ij}(\tau)\right\} + \left\{\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)\left(\widehat{e}_{t,i}^{\dagger} - e_{t,i}\right)e_{t,j}}{\sum_{t=1}^{n}\omega_{n,t}(\tau)} + \right. \\ &\quad \left. \frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)e_{t,i}\left(\widehat{e}_{t,j}^{\dagger} - e_{t,j}\right)}{\sum_{t=1}^{n}\omega_{n,t}(\tau)} + \frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)\left(\widehat{e}_{t,i}^{\dagger} - e_{t,i}\right)\left(\widehat{e}_{t,j}^{\dagger} - e_{t,j}\right)}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}\right\} \\ &=: \chi_{ij}^{\diamond}(\tau) + \chi_{ij}^{\dagger}(\tau).\end{aligned} \tag{C.3}$$

By (B.25), we only need to show

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\chi_{ij}^{\dagger}(\tau)\right| = O_P\left(\nu_{n,d}^{\dagger}\right). \tag{C.4}$$

Following the proof of (B.36), we have

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left|\widehat{e}^{\dagger}_{t,i}-e_{t,i}\right|=O_P\left(s\zeta^{\dagger}_{n,d}\sqrt{\log(n\vee d)}\right). \tag{C.5}$$

By (B.25), (C.5) and the Cauchy-Schwarz inequality, we can prove that

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i}-e_{t,i}\right)e_{t,j}\right|=O_P\left(s\zeta^{\dagger}_{n,d}\sqrt{\log(n\vee d)}\right)=O_P\left(\nu^{\dagger}_{n,d}\right), \tag{C.6}$$

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)e_{t,i}\left(\widehat{e}^{\dagger}_{t,j}-e_{t,j}\right)\right|=O_P\left(\nu^{\dagger}_{n,d}\right), \tag{C.7}$$

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i}-e_{t,i}\right)\left(\widehat{e}^{\dagger}_{t,j}-e_{t,j}\right)\right|=o_P\left(\nu^{\dagger}_{n,d}\right). \tag{C.8}$$

With (C.6)–(C.8), we complete the proof of (C.4). ■

Define

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

Proof of Proposition 5.1. We start with the proof of

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\widehat{\Psi}(\tau_t)u\geqslant\kappa_0/2,\ \ w.p.a.1, \tag{C.9}$$

where $\mathcal{B}_i(\tau)$ is defined as in (4.2). In fact, combining Assumption 6(i) with the arguments in the proofs of Lemmas B.3 and C.1, we may show that

$$\max_{1\leqslant t\leqslant n}\left\|\widehat{\Psi}(\tau_t)-\Psi(\tau_t)\right\|_{\max}=O_P\left([\log(n\vee d)]^{1/2}\delta_X\right). \tag{C.10}$$

Then, using (C.10) and the arguments in the proof of Lemma D.1, we have

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\widehat{\Psi}(\tau_t)u\geqslant\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\Psi(\tau_t)u+O_P\left([\log(n\vee d)]^{1/2}s\delta_X\right),$$

which, together with Assumptions 3(ii) and 6(i), completes the proof of (C.9).

The proofs of (5.6) and (5.7) are similar to the proofs of Theorems 4.1 and 4.2 but with Lemma B.3 and (4.3) replaced by Lemma C.1 and (C.9), respectively. The proof of (5.8) is similar to the proof of Theorem 4.3 but with Lemma B.5 replaced by Lemma C.2. Details are omitted here to save

the space. ■

Proof of Proposition 5.2. With Proposition 5.1(ii), the proof of (5.9) is similar to the proof of Corollary 4.1. With Proposition 5.1(iii), the proof of (5.10) is similar to the proof of Corollary 4.2.■

Appendix D: Verification of Assumption 3(ii)

In this appendix, we verify the uniform restricted eigenvalue condition (4.3) for the time-varying VAR under the Gaussian assumption, i.e., $e_t \sim \mathsf{N}(\mathbf{0}_d, \mathbf{\Sigma}_t)$. Recall that

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

We first give some technical lemmas together with their proofs.

Lemma D.1. *Conditional on the event that*

$$\mathcal{E}_{\mathbf{\Psi}}(\delta) = \left\{ \max_{1\leqslant t\leqslant n} \|\mathbf{\Psi}(\tau_t) - \mathsf{E}[\mathbf{\Psi}(\tau_t)]\|_{\max} \leqslant \delta \right\},$$

we have

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathbf{\Psi}(\tau_t)u \geqslant \min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - 18\delta s,$$

where $\mathcal{B}_i(\tau)$ is defined in Section 4.1 and s is defined in Assumption 2(ii).

Proof of Lemma D.1. The proof is similar to Lemma 6 in Kock and Callot (2015). Write $\mathcal{J}_{i,t} = \mathcal{J}_i(\tau_t)$ and $\mathcal{J}'_{i,t} = \mathcal{J}'_i(\tau_t)$. For $u = \left(u_1^\intercal, u_2^\intercal\right)^\intercal \in \mathcal{B}_i(\tau_t)$ and given $\mathcal{E}_{\mathbf{\Psi}}(\delta)$, we have

$$\begin{aligned} u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - u^\intercal \mathbf{\Psi}(\tau_t)u &\leqslant \left|u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - u^\intercal \mathbf{\Psi}(\tau_t)u\right| = \left|u^\intercal(\mathbf{\Psi}(\tau_t) - \mathsf{E}[\mathbf{\Psi}(\tau_t)])u\right| \\ &\leqslant \delta|u|_1^2 \leqslant 9\delta\left(|u_1(\mathcal{J}_{i,t})|_1 + \left|u_2(\mathcal{J}'_{i,t})\right|_1\right)^2 \\ &\leqslant 18\delta s\left(\|u_1(\mathcal{J}_{i,t})\|^2 + \left\|u_2(\mathcal{J}'_{i,t})\right\|^2\right) \leqslant 18\delta s, \end{aligned}$$

where $u(\mathcal{J})$ denotes the vector consisting only the elements of u index by $\mathcal{J}$. This indicates that

$$u^\intercal \mathbf{\Psi}(\tau_t)u \geqslant u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - 18\delta s.$$

Taking $\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}$ on both sides of the above inequality, we complete the proof of Lemma D.1. ■

Letting

$$\overline{X}_t(\tau) = \begin{bmatrix} X_t \\ X_t\left(\frac{\tau_t-\tau}{h}\right) \end{bmatrix} \text{ and } \overline{X}_{K,t}(\tau) = K^{1/2}\left(\frac{\tau_t-\tau}{h}\right)\overline{X}_t(\tau),$$

we may re-write

$$(nh)u^\intercal \mathbf{\Psi}(\tau)u = \sum_{t=1}^{n} u^\intercal \left[\overline{X}_{K,t}(\tau)\overline{X}_{K,t}^\intercal(\tau)\right] u = \|\overline{\mathbf{X}}_u(\tau)\|^2,$$

with $\overline{\mathbf{X}}_u(\tau) = [u^\intercal \overline{X}_{K,1}(\tau), \cdots, u^\intercal \overline{X}_{K,n}(\tau)]^\intercal$. Since $\overline{X}_{K,t}(\tau)$ is a Gaussian random vector, we can adopt the following lemma (e.g., Lemma 7 of Kock and Callot, 2015).

Lemma D.2. *Let $\mathbf{Z}$ be an $n \times 1$ vector with $\mathbf{Z} \sim N(\mathbf{0}_n, \mathbf{Q})$. Then, for any $\delta, m > 0$,*

$$P\left(\|\mathbf{Z}\|^2 - E\|\mathbf{Z}\|^2 > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2}{8n\|\mathbf{Q}\|_\infty^2 m^2}\right) + n\exp\left(-m^2/2\right).$$

The inequality in Lemma D.2 is crucial to derive the probability of the event $\mathcal{E}_{\mathbf{\Psi}}(\delta)$ defined in Lemma D.1, as shown in the following lemma.

Lemma D.3. *Suppose that Assumptions 1 and 2(i) are satisfied. Then, for any $\delta, m > 0$, we have*

$$P\left(\mathcal{E}_{\mathbf{\Psi}}(\delta)\right) \leqslant 4nd^2\left[6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right)\right], \tag{D.1}$$

where $C_\diamond = \frac{2C_* C_K C_1^2}{(1-\rho)(1-\rho^2)}$, C_* is defined in Lemma B.1, C_K is the upper bound of the kernel function $K(\cdot)$, and C_1 and ρ are defined in (2.4).

Proof of Lemma D.3. Let the (i,j)-entry of $\mathbf{\Psi}(\tau_t)$ be $\Psi_{i,j}(\tau_t)$. For any $\delta > 0$, we note that

$$P\left(\max_{1\leqslant t\leqslant n}\max_{1\leqslant i,j\leqslant 2d}|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right) \leqslant \sum_{t=1}^{n}\sum_{i=1}^{2d}\sum_{j=1}^{2d} P\left(|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right).$$

Hence, it suffices to show

$$P\left(|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right) \leqslant 6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right). \tag{D.2}$$

By removing the zero elements of $\overline{\mathbf{X}}_u(\tau)$, we define a sub-vector $\widetilde{\mathbf{X}}_u(\tau)$ which only contains the non-zero elements. We apply Lemma D.2 with $\mathbf{Z} = \widetilde{\mathbf{X}}_u(\tau_t)$ and $\mathbf{Q} = \mathbf{Q}(\tau_t) = \mathrm{Cov}(\widetilde{\mathbf{X}}_u(\tau_t))$. Consider a typical entry in $\mathbf{Q}(\tau_t)$: $\mathrm{Cov}\left(u^\intercal \overline{X}_{K,l_1}(\tau_t), u^\intercal \overline{X}_{K,l_2}(\tau_t)\right)$ when $|\tau_{l_1} - \tau_t| \leqslant h$ and $|\tau_{l_2} - \tau_t| \leqslant h$, where $u = \left(u_1^\intercal, u_2^\intercal\right)^\intercal$ is an appropriately selected vector with dimension $2d$ and $\|u\| = 1$. Letting

$u_{\tau,l} = (u_1 + \frac{\tau_l - \tau}{h} u_2)/\|u_1 + \frac{\tau_l - \tau}{h} u_2\|$, we have

$$
\begin{aligned}
& \mathsf{Cov}\left(u^\intercal \overline{X}_{K,l_1}(\tau_t), u^\intercal \overline{X}_{K,l_2}(\tau_t)\right) \\
= \; & \mathsf{E}\left[\left(u_1 + \frac{\tau_{l_1} - \tau_t}{h} u_2\right)^\intercal X_{l_1} X_{l_2}^\intercal \left(u_1 + \frac{\tau_{l_2} - \tau_t}{h} u_2\right)\right] K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right) \\
\leqslant \; & \left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| \left\|(u_1 + \frac{\tau_l - \tau}{h} u_2\right\| \left\|(u_1 + \frac{\tau_t - \tau}{h} u_2\right\| K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right) \\
\leqslant \; & \left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right).
\end{aligned}
$$

For $1 \leqslant l_1, l_2 \leqslant n$ with $|\tau_{l_1} - \tau_t| \leqslant h$ and $|\tau_{l_2} - \tau_t| \leqslant h$, by (2.3) and (2.4),

$$
\begin{aligned}
\left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| &= \left|\mathsf{E}\left[\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\left(u_{\tau_t,l_1}^\intercal \mathbf{\Phi}_{l_1,k_1} e_{l_1-k_1}\right)\left(u_{\tau_t,l_2}^\intercal \mathbf{\Phi}_{l_2,k_2} e_{l_2-k_2}\right)^\intercal\right]\right| \\
&\leqslant C_* C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1}\rho^{|l_2-l_1|+k_1} = \frac{C_* C_1^2 \rho^{|l_2-l_1|}}{1-\rho^2}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\max_{1\leqslant t\leqslant n} \|\mathbf{Q}(\tau_t)\|_\infty &\leqslant \frac{C_* C_1^2}{1-\rho^2} \max_{1\leqslant l_1\leqslant n} \sum_{l_2=1}^{n} \rho^{|l_2-l_1|}\left[\max_{1\leqslant t\leqslant n} K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right)\right] \\
&\leqslant \frac{2C^* C_1^2 C_K}{1-\rho^2} \sum_{k=0}^{\infty} \rho^k \leqslant \frac{2C^* C_1^2 C_K}{(1-\rho)(1-\rho^2)} = C_\diamond.
\end{aligned}
$$

Using Lemma 7 in Kock and Callot (2015) and noting that the dimension of $\widetilde{\mathbf{X}}_u(\tau_t)$ is $(2nh)$, we obtain that for any $\delta, m > 0$,

$$
\mathsf{P}\left(\left\|\widetilde{\mathbf{X}}_u(\tau_t)\right\|^2 - \mathsf{E}\left\|\widetilde{\mathbf{X}}_u(\tau_t)\right\|^2 > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2}{16C_\diamond^2 m^2 (nh)}\right) + 2nh\exp\left(-m^2/2\right),
$$

indicating that

$$
\mathsf{P}\left(u^\intercal \mathbf{\Psi}(\tau_t) u - \mathsf{E}\left[u^\intercal \mathbf{\Psi}(\tau_t) u\right] > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2 (nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right). \tag{D.3}
$$

Choosing u as a vector with the i-th element being one and the others being zeros, by (D.3), we have

$$
\mathsf{P}\left(|\Psi_{i,i}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t)]| > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2 (nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right) \tag{D.4}
$$

for $i = 1, \cdots, 2d$. Analogously, we may further show that, for $1 \leqslant i \neq j \leqslant 2d$,

$$\begin{aligned}
& \mathsf{P}\left(|\Psi_{i,j}(\tau_t) - \mathsf{E}[\Psi_{i,j}(\tau_t)]| > \delta\right) \\
\leqslant\ & \mathsf{P}\left(|\Psi_{i,i}(\tau_t) - 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) - 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t)]|/2 > \delta/2\right) + \\
& \mathsf{P}\left(|\Psi_{i,i}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) + \Psi_{j,j}(\tau_t)]|/2 > \delta/2\right) \\
\leqslant\ & \mathsf{P}\left(|\Psi_{i,i}(\tau_t) + 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) + 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t)]| > \delta\right) + \\
& \mathsf{P}\left(|\Psi_{i,i}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t)]| > \delta/2\right) + \mathsf{P}\left(|\Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{j,j}(\tau_t)]| > \delta/2\right) \\
\leqslant\ & 6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right). \qquad \text{(D.5)}
\end{aligned}$$

By virtue of (D.4) and (D.5), we complete the proof of (D.2). ■

The following proposition verifies the uniform restricted eigenvalue condition.

Proposition D.1. *Suppose that Assumptions 1 and 2(i) are satisfied. If*

$$\min_{1\leqslant t\leqslant n} \inf_{u\in\mathcal{B}} u^\intercal \mathsf{E}[X_t X_t^\intercal] u \geqslant 2\kappa_0, \qquad \text{(D.6)}$$

where $\mathcal{B} = \{u : \|u\| = 1, |u|_1 \leqslant 3|u_{\mathcal{J}}|_1\}$, $\mathcal{J}$ *is any index set satisfying* $\mathcal{J} \subset \{1, \cdots, d\}$ *with cardinality*

$$s = o\left((nh)^{1/2}/\log(ndh^{1/2})\right),$$

we have (4.3) w.p.a.1.

Proof of Proposition D.1. Taking $\delta = c_\circ/s$ and $m^2 = \left(\frac{c_\circ^2 nh}{32C_\diamond^2 s^2}\right)^{1/2}$ in Lemma D.3 with $c_\circ$ being a proper constant to be determined later, we have

$$\begin{aligned}
\mathsf{P}\left(\max_{1\leqslant t\leqslant n} \|\boldsymbol{\Psi}(\tau_t) - \mathsf{E}[\boldsymbol{\Psi}(\tau_t)]\|_{\max} > \frac{c_\circ}{s}\right) &\leqslant 4nd^2\left[6\exp\left(\frac{-c_\circ^2 nh}{64C_\diamond^2 s^2 m^2}\right) + 6nh\exp\left(-m^2/2\right)\right] \\
&\leqslant 48\exp\left(\log(n^2 d^2 h) - \frac{c_\circ (nh)^{1/2}}{16C_\diamond s}\right),
\end{aligned}$$

which converges to 0 if $s = o\left((nh)^{1/2}/\log(ndh^{1/2})\right)$. By Lemma D.1, we then have

$$\min_{1\leqslant i\leqslant d} \min_{1\leqslant t\leqslant n} \inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \boldsymbol{\Psi}(\tau_t) u \geqslant \min_{1\leqslant i\leqslant d} \min_{1\leqslant t\leqslant n} \inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathsf{E}[\boldsymbol{\Psi}(\tau_t)] u - 18c_\circ \ \ w.p.a.1. \qquad \text{(D.7)}$$

It remains to prove that the first term on the right side of (D.7) has a lower bound and find a proper value for $c_\circ$. In fact, by (D.6), we have

$$\min_{1\leqslant i\leqslant d} \min_{1\leqslant t\leqslant n} \inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathsf{E}[\boldsymbol{\Psi}(\tau_t)] u$$

$$= \min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}\frac{1}{nh}\sum_{l=1}^{n}\mathsf{E}\left[\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)^{\intercal}X_lX_l^{\intercal}\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)\right]K\left(\frac{\tau_l-\tau_t}{h}\right)$$

$$\geqslant 2\kappa_0\min_{1\leqslant t\leqslant n}\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{\tau_l-\tau_t}{h}\right)=2\kappa_0-\epsilon,$$

where ϵ is an arbitrary small number. Choosing $c<(\kappa_0-\epsilon)/18$ in (D.7), we can complete the proof of (4.3). ■

Appendix E: Tuning parameter selection

The numerical performance of the proposed three-state shrinkage estimation procedure depends on a careful selection of the three tuning parmaeters: λ_1 in the preliminary time-varying LASSO estimation, λ_2 in the time-varying weighted group LASSO, and λ_3 in the 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. We next briefly introduce these three criteria.

The local linear regression smoothing in (3.3) is essentially the weighted least squares with kernel weights $K_h(\tau_t-\tau)$. The BIC objective function is thus defined as

$$\mathsf{BIC}_i(\lambda_1;\tau)=\log\left[\frac{\mathcal{L}_i\left(\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1),\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)\right)}{\sum_{t=1}^{n}K_h(\tau_t-\tau)}\right]+\frac{\log(n_e)}{n_e}\cdot\left[|\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1)|_0+\left|\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)\right|_0\right], \quad \text{(E.1)}$$

where $\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1)$ and $\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)$ are the local linear estimates using the tuning parameter λ_1 at the point τ, and the effective sample size n_e is defined as $\sum_{t=1}^{n}K_h(\tau_t-\tau)/\max_t\{K_h(\tau_t-\tau)\}$. We select the tuning parameter in the preliminary time-varying LASSO by minimising $\mathsf{BIC}_i(\lambda_1;\tau)$ defined in (E.1) with respect to λ_1. The selected tuning parameter depends on both the index i and the (scaled) time point τ.

The GIC is introduced by Fan and Tang (2013) in the context of high-dimensional penalised likelihood estimation. As our model involves unknown time-varying coefficients and the estimation procedure involves local linear smoothing, we need to modify the GIC as in Li, Ke and Zhang (2015). For example, Cheng, Zhang and Chen (2009) suggest that each unknown functional parameter would amount to $36/(35h)$ unknown constant parameters when the Epanechnikov kernel is used. Hence, we define the GIC objective function as

$$\mathsf{GIC}_i(\lambda_2)=\log\left[\frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i}-\widehat{\alpha}_{i\bullet}^{\intercal}(\tau_t\mid\lambda_2)X_{t-1}\right\}^2\right]+\frac{\gamma_{n,d}}{n}\cdot\frac{36s_i(\lambda_2)}{35h}, \quad \text{(E.2)}$$

where $\gamma_{n,d}$ is a function of n and d, $\widehat{\boldsymbol{\alpha}}_{i\bullet}(\tau \mid \lambda_2)$ is the time-varying weighted group LASSO estimate using the tuning parameter λ_2 and $s_i(\lambda_2)$ is the number of selected time-varying coefficients using λ_2. We choose $\gamma_{n,d} = \gamma \log(\log(n)) \log(36d/(35h))$ with $\gamma \in (0, 1]$. We determine the tuning parameter by minimising $\mathsf{GIC}_i(\lambda_2)$ defined in (E.2) with respect to λ_2. The selected tuning parameter depends on the index i. A smaller γ leads to denser network estimation. The intuition to select a γ less than 1 is that when a functional parameter is zero in most of the sampling period and non-zero otherwise, the marginal contribution to the sum of squared error by including the corresponding variable is small, and a smaller γ adjusts the the information criterion to be more adaptive and sensitive. For example, when we want to select variables whose functional parameter is not zero in at least 10% of the sampling period, we can choose $\gamma = 0.1$. We choose $\gamma = 1$ in the simulation and $\gamma = 0.1$ in the empirical study.

The EBIC is proposed by Chen and Chen (2008) and has been applied to Gaussian graphical model estimation by Foygel and Drton (2010). The EBIC objective function is defined as

$$\mathsf{EBIC}(\lambda_3; \tau) = -\log\left(\det(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3))\right) + \mathsf{Tr}(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3)\widehat{\boldsymbol{\Sigma}}(\tau)) + \frac{\log(n_e)}{n_e} \cdot \sum_{i<j} \mathrm{I}(|\widehat{\omega}_{ij}(\tau \mid \lambda_3)| > 0), \quad \text{(E.3)}$$

where $\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3) = [\widehat{\omega}_{ij}(\tau \mid \lambda_3)]_{d \times d}$ denotes the time-varying CLIME estimate obtained using the tuning parameter λ_3. We determine the tuning parameter by minimising $\mathsf{EBIC}(\lambda_3; \tau)$ defined in (E.3) with respect to λ_3. Note that the selected tuning parameter changes with τ.

The numerical performance of the factor-adjusted VAR model and methodology depends on a careful selection of the factor number. Let $\widehat{X}_t(q)$ be the estimated idiosyncratic component in (5.3) or (5.4), when the number of factors is set to be q, and define the sum of squared residuals as $V_n(q) = \sum_{t=1}^{n} |\widehat{X}_t(q)|_2^2$. When we consider the approximate factor model (5.1), we select the factor number by the information criterion developed by Bai and Ng (2002), i.e., maximise the following objective function with respect to q

$$\mathsf{IC}(q) = \log[V_n(q)] + q \cdot \left(\frac{n+d}{nd}\right) \log(n \wedge d),$$

and obtain $\widehat{q}$ as the estimated number of factors. When we consider the time-varying factor model (5.2), we adopt Su and Wang (2017)'s information criterion, i.e., maximise the following objective function with respect to q

$$\mathsf{IC}(q) = \log[V_n(q)] + q \cdot \left(\frac{nh_* + d}{nh_* d}\right) \log(nh_* \wedge d),$$

and obtain $\widehat{q}$ as the estimated number of factors, where h_* is the bandwidth used in the local PCA.

The above two criteria are used in the empirical data analysis to determine the factor numbers.

In practice, we need to select an appropriate order for the time-varying VAR model. For the high-dimensional VAR model with constant transition matrices, Miao, Phillips and Su (2022) introduces a ratio criterion which compares Frobenius norms of the estimated transition matrices over different lags. We next extend their criterion to the time-varying VAR model context. Define

$$R(k) = \frac{\sum_{l=k}^{2k_{max}} \sum_{t=1}^{n} (\|\widehat{\mathbf{A}}_{t,l}\|_F \vee \xi_A)}{\sum_{l=k+1}^{2k_{max}} \sum_{t=1}^{n} (\|\widehat{\mathbf{A}}_{t,k}\|_F \vee \xi_A)},$$

where k_{max} and ξ_A are user-specified. In Section 7 of the main document, we set $k_{max} = 10$ and $\xi_A = 0.1$ and use the estimated transition matrices of time-varying VAR(20) in computing $R(k)$. The order of the time-varying VAR is selected by the integer which maximises $R(k)$, $1 \leqslant k \leqslant k_{max}$. In the empirical analysis, we use the above criterion to select the time-varying VAR(1).

In Tables 1–7 of the main document, in order to evaluate the accuracy of the estimated time-varying VAR and network structures, 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). The FP is defined as the number of insignificant predictor variables falsely identified as the significant ones; FN is defined as the number of significant predictor variables falsely identified as the insignificant ones; TPR and TNR are defined by

$$TPR = \frac{TP}{TP + FN} \text{ and } TNR = \frac{TN}{TN + FP}$$

with TP denoting true positive whereas TN denoting true negative; PPV and NPV are defined by

$$PPV = \frac{TP}{TP + FP} \text{ and } NPV = \frac{TN}{TN + FN};$$

the F1 score is the harmonic mean of precision and sensitivity defined by

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR};$$

and MCC is defined as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

References

BAI, J. AND NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 90, 191–221.

BICKEL, P., RITOV, Y. AND TSYBAKOV, A. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37, 1705–1732.

CAI, T. T., LIU, W. AND LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106, 594–607.

FAN, J. AND LV, J. (2011). Non-concave penalized likelihood with NP-Dimensionality. *IEEE: Information Theory*, 57, 5467–5484.

FAN, J., XUE, L. AND ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics* 42, 819–849.

CHEN, J. AND CHEN, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95, 759–771.

CHENG, M. -Y., ZHANG, W. AND CHEN, L. -H. (2009). Statistical estimation in generalized multiparameter likelihood models. *Journal of the American Statistical Association* 104, 1179–1191.

FAN, Y. AND TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75, 531–552.

Foygel, R. and Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems* 23.

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *The Annals of Probability*, 3, 100–118.

KOCK, A.B. AND CALLOT, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* 186, 325–344.

LI, D., KE, Y. AND ZHANG, W. (2015). Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. *The Annals of Statistics* 43, 2676–2705.

LIAN, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statistica Sinica*, 22, 1563–1588.

MIAO, K., PHILLIPS, P.C.B. AND SU, L. (2022). High-dimensional VARs with common factors. Forthcoming in *Journal of Econometrics*.

DE LA PEÑA, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *The Annals of Probability*, 27, 537–564.

SU, L. AND WANG, X. (2017). On time-varying factor models: estimation and testing. *Journal of Econometrics* 198, 84–101.