Three Essays on the Econometric Methods for High-Dimensional Economic and Financial Data Using Factor Structures

Yuning Li

PhD

University of York

Economics and Related Studies

July, 2023

Abstract

This dissertation is a collection of three independent essays that explore novel econometric approaches for analyzing high-dimensional economic and financial data. A key commonality among these essays is the use of a factor structure, which aims to capture the underlying latent factors that drive the dynamics of the data. Moreover, each essay focuses on a unique aspect of the high-dimensional factor model, delving into diverse areas such as high-frequency data analysis, network analysis, and portfolio management in financial markets.

Chapter 1 studies high-frequency cross-sectional intraday stock returns which are contaminated with microstructure noise and exhibit co-movements. A dual factor model is introduced to capture the underlying dynamics of efficient prices and microstructure noise. Then a Double Principal Component Analysis (DPCA) method is proposed for the estimation of common factors for both efficient prices and microstructure noise.

Chapter 2 shifts the focus to network analysis for high-dimensional time series. Using a high-dimensional time-varying factor-adjusted vector autoregressive (VAR) model framework, two types of networks are investigated: a directed Granger causality network and an undirected partial correlation network. To estimate the transition and precision matrices, a penalized local linear method with a time-varying weighted group LASSO and a time-varying CLIME method is proposed.

Chapter 3 addresses the estimation of large dynamic precision matrices with multiple conditioning variables. To overcome the challenges of high dimensionality and cross-dependence, an approximate factor structure is introduced. A semiparametric method based on model averaging marginal regression is employed to approximate the underlying dynamic covariance matrices of the factors and the idiosyncratic components. The estimate of the dynamic precision matrices for the original time series is then obtained by utilising the Sherman-Morrison-Woodbury formula, and is applied in the construction of the minimum variance portfolio.

Throughout each chapter, asymptotic properties of the proposed estimates are established and validated through extensive Monte Carlo simulations. These methods are further applied to stock return datasets or a macroeconomic dataset to demonstrate their strong performance.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. Where individual chapters were co-authored with other researchers, this is indicated with the necessary specifications in this declaration.

Chapter 1 is a coauthored paper with Prof. Jia Chen and Prof. Oliver Linton and has been published in the *Journal of Econometrics*. I am the first author and contributed to every part of the research and my co-authors contributed to various points with revisions and comments.

Chapter 2 is a coauthored paper with Prof. Jia Chen, Prof. Degui Li, and Prof. Oliver Linton. I contributed to coding and executing the Monte Carlo simulations, designing and conducting all the empirical analyses and writing the paper. My co-authors contributed at various points with revisions and comments on design and methodology.

Chapter 3 is a coauthored paper with Prof. Jia Chen. I contributed to coding and executing the Monte Carlo simulations, designing and conducting all the empirical analyses and writing the paper. My co-author contributed at various points with revisions and comments on design and methodology.

An earlier version of Chapter 1 was presented at the 15^{th} International Conference on Computational and Financial Econometrics in December 2021, the 13^{th} York Econometrics Symposium in May 2022, York-Maastricht Partnership Workshop in September 2022, the AFR Summer Institute of Economics and Finance in 2022, York Financial Econometrics Workshop in March 2023. An earlier version of Chapter 2 was presented at the 16^{th} International Conference on Computational and Financial Econometrics in December 2022 and Virtual Workshop for Junior Researchers in Time Series in April 2023.

Yuning Li
July 2023, York

Acknowledgements

I would like to express my deepest gratitude to everyone who has supported me throughout my journey in pursuing this second PhD, especially during the challenging times of the COVID-19 pandemic.

First and foremost, I am extremely grateful to my supervisor, Prof. Jia Chen, for her continuous guidance, patience, and invaluable mentorship. Her expertise and dedication have been instrumental in shaping the direction of my research and providing valuable insights at every stage of this dissertation.

Second, I would also like to thank Prof. Peter Smith and Prof. Oliver Linton for serving as my thesis advisory panel members, and for offering me insightful advice throughout my research journey. I am also grateful to Prof. Degui Li for his collaboration and valuable contributions.

Third, I would like to acknowledge the support and guidance I have received from Prof. Laura Coroneo and Dr Jorgen Kratz, with whom I have worked as a teaching assistant. Their expertise has greatly contributed to my academic development. Additionally, I would like to thank Prof. Subir Chattopadhyay, Prof. Andrew Jones, Prof. Kostas Koufopoulos, Prof. Michael Thornton, Dr Francesco Bravo, Dr Adam Golinski, Dr Michael Shallcross, Mr Vin McDermott, and Maigen Savory for their teaching or assistance in various aspects of my academic journey.

Fourth, I want to extend my gratitude to my friends at UoY, particularly Rui Lin, Jiaxin Peng, Songqiao Tang, Wenting Wang, Ting Xie, Wenkun Zhang, Chaowen Zheng, and many others. Your friendship and shared experiences have brought joy and memorable moments to my life during my time at the university.

Fifth, I would like to express my heartfelt appreciation to my wife, Huilei Kang, for her unwavering support, understanding, and encouragement throughout this challenging period. Her presence in my life has been a constant source of strength and inspiration.

Finally, I would like to express my deepest thanks to my parents. Although we have not been able to meet for almost four years due to the pandemic, their unwavering support, love, and sacrifices have been the foundation of my personal and academic growth. I am forever grateful for everything

they have done for me, and I wish them all the best.

Although it is impossible to name everyone individually, I extend my sincere thanks to all those who have played a part, big or small, in supporting me throughout this academic journey. Your support and encouragement have been immeasurable, and I am truly grateful for the opportunity to undertake this second PhD.

Contents

List of Figures

List of Tables

Introduction

The exponential growth in data availability has sparked the development of data science, a multi-disciplinary field that encompasses statistics, computer science, and econometrics. This data-driven era presents unique challenges in handling complex datasets with intricate structures, a large number of variables, and diverse sources of noise. To effectively extract meaningful insights and capture the underlying dynamics of the data, it is essential to employ robust methods that can handle high dimensionality while accounting for cross-sectional dependence in the data.

The factor model has emerged as a powerful framework for analyzing high-dimensional data. By assuming that the observed variables can be expressed as linear combinations of a few latent factors plus an approximation error, the factor model provides a means of dimension reduction and capturing the essential dynamics of the data. From a machine learning perspective, the factor model can be viewed as a special case of the encoder-decoder framework in unsupervised learning. It aims to learn a low-dimensional representation of the observed data by mapping it to a smaller set of latent factors, which can then be used to reconstruct the original variables. From an econometric standpoint, the factor model offers more than just dimension reduction. It also enhances the interpretability of the underlying (economic and financial) phenomena, because the latent factors in the factor model can be viewed as fundamental drivers that explain the common variation or trends among the observed variables. By uncovering and examining these latent factors, economists and researchers gain valuable insights into the economic mechanisms that govern the data dynamics.

Owing to the advantages mentioned above, the latent factor model has found widespread applications in various areas, especially in economics and finance (e.g., Chamberlain and Rothschild, 1983; Stock and Watson, 2002; Giglio *et al.*, 2022). Recognizing the increasing importance of factor models in high-dimensional data analysis, this PhD dissertation aims to contribute to the existing literature by introducing novel econometric methods that use factor structures to address specific challenges encountered in high-dimensional economic and financial data analysis.

- High-frequency high-dimensional time series analysis

The increasing availability of high-frequency transaction data motivates applying factor models to intraday stock prices. However, this setting raises certain theoretical and computational challenges. Compared to the discrete-time factor model, new mathematical tools are required to deal with a continuous-time setting, where long-span asymptotics (also called increasing domain asymptotics) gives way to infill asymptotics (also called fixed domain asymptotics). Market microstructure is an additional challenge that must be faced. The specifics of market organisation and market participants' behaviour induce certain short-run patterns in security prices. These patterns, such as bid-ask bounce and price-discreteness, lead to a deviation from the fundamental values (also known as efficient prices) of the securities. The security prices are thereby contaminated with market microstructure noise, which affects the estimation of parameters of interest.

High-dimensional models with microstructure noise have been developed more recently. Wang and Zou (2010) propose the first noise-robust estimators of the integrated volatility matrix and establish an asymptotic theory that allows both the sample size and the number of assets to approach infinity, see also Tao *et al.* (2011, 2013a,b), and Kim *et al.* (2016) for related results. However, these papers assume that the integrated volatility matrix is sparse, which often contradicts our intuition from low-frequency data analysis. To solve this problem, Pelger (2019) and Dai *et al.* (2019) develop a continuous-time factor model with microstructure noise.

Bollerslev *et al.* (2019) investigate a continuous-time factor model and assume that microstructure noise can have a factor structure itself. To eliminate the influence of microstructure noise on estimation, they employ the modulated realised volatility estimator proposed by Christensen *et al.* (2010). Notably, their approach avoids explicitly estimating the factors for microstructure noise and separating them from those corresponding to efficient prices. However, given that the factor structure for microstructure noise can be of independent interest, Chapter 1 aims to separately estimate these factors from the observed prices. To the best of our knowledge, this is the first work to specifically look at common factors for microstructure noise.

Chapter 1 studies high-frequency cross-sectional intraday stock returns which are contaminated with microstructure noise and exhibit co-movements. A dual factor model is introduced to capture the underlying dynamics of efficient prices and microstructure noise. Then a Double Principal Component Analysis (DPCA) method is proposed for the estimation of common factors for both efficient prices and microstructure noise. The uniform consistency of these estimators is established as the number of assets and sampling frequency increase. Moreover, a Monte Carlo exercise shows that our DPCA method outperforms the PCA-VECM method. Lastly, an empirical analysis of intraday returns of S&P 500 Index constituents provides evidence of co-movement of the microstructure noise, highlighting its distinguishing features from latent systematic risk factors.

- Time-varying network analysis

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. These models extend the standard factor model by taking into account the weak dependence of the idiosyncratic errors. Chapter 2 aims to further study these dependencies and unveil underlying relationships between variables via network analysis. Motivated by the stable network time series analysis in Barigozzi and Brownlees (2019), time-varying Vector autoregression (VAR) is used to construct two dynamic networks of interest: the Granger causality network and the partial correlation network.

In recent years, there has been increasing interest in extending the finite-dimensional VAR to the high-dimensional setting. Under appropriate sparsity restrictions on the transition (or autoregressive coefficient) matrices, various regularised methods have been proposed to estimate high-dimensional VAR models and identify non-zero entries in the transition matrices (e.g., Basu and Michailidis, 2015; Han *et al.*, 2015; Kock and Callot, 2015; Davis *et al.*, 2016). Moreover, to capture smooth structural changes in the underlying data generating process, time-varying VAR models are developed (e.g., Ding *et al.*, 2017; Xu *et al.*, 2020; Safikhani and Shojaie, 2022). Chapter 2 reconsider the estimation problem from a network perspective, combining the kernel smoothing with LASSO regularisation in a preliminary estimation step and combining the kernel smoothing with weighted-group-LASSO regularisation in a second step to construct the estimator of the Granger causality network.

Regarding the partial correlation network, the so-called graphical model is commonly used 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 (e.g., Yuan and Lin, 2007; Fan *et al.*, 2009; Loh and Wainwright, 2013; Basu *et al.*, 2015; Zhao *et al.*, 2022), or dynamic network models with independent data (e.g., Kolar *et al.*, 2010; Zhou *et al.*, 2010; Wang *et al.*, 2021a). Chapter 2, on the other hand, investigates the dynamic partial correlation network based on the time-varying error precision matrix, considering the estimation error of both factor analysis and VAR estimation.

In Chapter 2, theoretical properties, including consistency and oracle properties, are derived under the assumption of sparsity. In the case where the time series are highly cross-sectionally correlated and the sparsity assumption is likely to be violated, we introduce a factor structure to account for the cross-sectional dependencies so that the residual component follows a sparse network structure, for which our methods are valid again. Simulation studies and an empirical application to a large U.S. macroeconomic dataset demonstrate the good performance of our methods.

- Large precision matrix estimation

The estimation of large covariance matrices or precision matrices is a prominent subject in high-dimensional statistics. It not only plays a vital role in network analysis but also holds significance in portfolio management within the field of finance. Existing literature often assumes that the precision matrix of a high-dimensional random vector satisfies an approximate sparsity condition similar to that often imposed on large covariance matrices (e.g., Bickel and Levina, 2008), and then uses various techniques, such as penalised likelihood (Lam and Fan, 2009), graphical Danzig selector (Yuan, 2010) and constrained ℓ_1-minimisation for inverse matrix estimation (CLIME) (Cai *et al.*, 2011), to estimate it. A comprehensive review of recent developments in large precision matrix estimation can be found in Cai *et al.* (2016) and Fan *et al.* (2016c).

Chapter 3 aims to estimate a large dynamic precision matrix with a latent factor structure, avoiding both the static and the sparsity assumptions. As in Tang *et al.* (2020) and Wu *et al.* (2017), a sparsity assumption is imposed on the error precision matrix, which leads to a "low-rank plus sparse" structure for the precision matrix of the time series. To capture the dynamics, conditioning variables are utilised in the estimation. To address the curse of dimensionality, an easy-to-implement semiparametric method, known as Model Averaging MArginal Regression (MAMAR), is used to estimate each entry of the conditional factor/error covariance matrices. Subsequently, the CLIME method is employed to obtain the estimate of the dynamic error precision matrices and the Sherman-Morrison-Woodbury formula is utilised to obtain the dynamic precision matrix for the time series. Under mild assumptions, such as the approximate sparsity assumption of the precision matrix of the error component, the uniform consistency of the proposed precision matrix estimator is established. Extensive simulations show the advantage of the low-rank plus sparse structure for covariance and precision matrices estimation. Furthermore, the developed methodology is applied to the returns of S&P 500 constituents, to demonstrate its effectiveness in the portfolio selection problem.

The dissertation is organised as follows. Chapter 1 focuses on high-frequency data analysis, presenting novel econometric methods to tackle the challenges associated with continuous-time factor models and market microstructure noise. Chapter 2 delves into network analysis, introducing methodologies to model time-varying networks to study the dependencies among a large panel of time series. Finally, Chapter 3 explores precision matrix estimation in high-dimensional settings, proposing innovative approaches to estimate time-varying precision matrices with a "low-rank plus sparse" structure. The dissertation concludes with a summary of the findings, contributions, and potential avenues for future research.

Chapter 1

Estimation of Common Factors for Microstructure Noise and Efficient Price in a High-frequency Dual Factor Model

Abstract We develop the Double Principal Component Analysis (DPCA) based on a dual factor structure for high-frequency intraday returns contaminated with microstructure noise. The dual factor structure allows a factor structure for microstructure noise in addition to the factor structure for efficient log-prices. We construct estimators of factors for both efficient log-prices and microstructure noise as well as their common components, and provide uniform consistency of these estimators when the number of assets and the sampling frequency go to infinity. In a Monte Carlo exercise, we compare our DPCA method to a PCA-VECM method. Finally, an empirical analysis of intraday returns of S&P 500 Index constituents provides evidence of the existence of co-movement in microstructure noise, and this co-movement is distinct from latent systematic risk factors.

Key Words: Cointegration, Factor model, High-frequency data, Microstructure noise, Non-stationarity.

1.1 Introduction

Factor models are widely used in many scientific fields, and in particular in the study of financial data. Their popularity is partly due to the easiness of their implementation and their effectiveness in dimension reduction. More and more observable factors have been investigated and reported (see, e.g., Ross (1976), Sharpe (1994), Fama and French (1993, 2015) and Carhart (1997)) as driving stock returns. Researchers have also found common components in other attributes of financial assets such as volatility and liquidity. For example, Chordia *et al.* (2000) document the commonality in liquidity, which remains significant after controlling for volatility, volume, and price. The factor structure is not found in isolation. Indeed, price and other attributes of stocks have been found to have correlated common factors. Hasbrouck and Seppi (2001) use principal component analysis to show that common factors exist in order flows and equity returns. In addition, using canonical correlation analysis, they find that the common factor in returns is highly correlated with the common factor in order flows. Hallin and Liška (2011) propose a two-step general dynamic factor method to account for a joint factor structure of sub-panels, which is further developed by Barigozzi and Hallin (2016) and Barigozzi and Hallin (2017) for extracting the market volatility shocks. They find that returns and volatilities can be decomposed into four mutually orthogonal components: a strongly idiosyncratic component, a strongly common component, a weakly common component, and a weakly idiosyncratic component.

The increasing availability of high-frequency transaction data motivates applying this methodology to intraday stock prices. However, this setting raises certain theoretical and computational challenges. Compared to the discrete-time factor model, new mathematical tools are required to deal with a continuous-time setting, where long-span asymptotics (also called increasing domain asymptotics) gives way to infill asymptotics (also called fixed domain asymptotics). For example, Fan *et al.* (2016b) and Aït-Sahalia and Xiu (2017) extend Fan *et al.* (2013)'s Principal Orthogonal complEment Thresholding (POET) method to high-frequency factor models. Market microstructure is an additional challenge that must be faced. The specifics of market organisation and market participants' behaviour induce certain short-run patterns in security prices. These patterns, such as bid-ask bounce and price-discreteness, lead to a deviation from the fundamental values (also known as efficient prices) of the securities. The security prices are thereby contaminated with market microstructure noise, which affects the estimation of parameters of interest such as volatility. Market microstructure models have been used to capture a variety of frictions inherent in the trading process. Roll (1984) is among the first to propose a dichotomous structure in which the observed market price is the sum of the efficient price and an exogenous i.i.d. bid-ask spread. After that, Hasbrouck and Ho (1987), Choi *et al.* (1988) and Hasbrouck (1993) consider extended models with

positive dependence in the bid and ask transactions. More complicated price patterns arising from microstructure noise, such as asynchronous trading, have been investigated by researchers under the fundamental dichotomous structure.

High-dimensional models with microstructure noise have been developed more recently. Wang and Zou (2010) propose the first noise-robust estimators of the integrated volatility matrix and establish an asymptotic theory that allows both the sample size and the number of assets to approach infinity, see also Tao *et al.* (2011, 2013a,b), and Kim *et al.* (2016) for related results. However, these papers assume that the integrated volatility matrix is sparse, which often contradicts our intuition from low-frequency data analysis. To solve this problem, Pelger (2019) and Dai *et al.* (2019) develop a continuous-time factor model with microstructure noise. Likewise, Bollerslev *et al.* (2019) investigate a continuous-time factor model and assume that microstructure noise can have a factor structure itself. They use the modulated realised volatility estimator (henceforth MRC) of Christensen *et al.* (2010) to eliminate the effect of the microstructure noise on the estimation without explicitly estimating the factors for the microstructure noise and separating them from those of the efficient prices. They establish the consistency and bound the rate of convergence of the estimated integrated covolatility matrix of the efficient price process in the large dimensional case. Related to this, Pelger (2019) classifies factors in a high-frequency factor model into jump factors and continuous factors.

We consider the dual factor model of Bollerslev *et al.* (2019) but we take a different approach to estimation. Our goal is to identify and separate the factors and common components from both sources: the efficient price process and the microstructure noise process. Factors for the efficient prices arise from information about future security cash flows and thereby are long-lasting, whereas factors for the microstructure noise are transient and due to the nature of trading behaviour; both are of interest. We develop a methodology that is inspired by Bai and Ng (2004), who propose a test procedure called Panel Analysis of Non-stationarity in Idiosyncratic and Common Components (PANIC), which can be used to identify non-stationary factors in discrete time series. We extend the PANIC approach to our high-frequency dual factor model. Our methodology is in two parts. First, we estimate the common factors and loadings of both signal and noise components simultaneously from the observed returns via PCA. The PCA uses the eigen-decomposition of a sample variance-covariance matrix to identify the common component and the idiosyncratic error. The second step separates the return factors into efficient price factors and microstructure noise factors. This involves a second PCA on the cumulative form of the factors found in the first step, following the approach of Bai and Ng (2004). Intuitively, after cumulation, the efficient price common components are nonstationary and thus have a larger magnitude than the stationary microstructure noise common

components, which enables us to separate the two different types of factors. Another way to obtain separate estimates of the factors is to: (i) first obtain a consistent estimate of the integrated covolatility matrix of the efficient price using a noise-robust method (such as pre-averaging) and an estimate of the covariance matrix of the microstructure noise; and (ii) apply PCA separately to these estimated matrices to obtain estimates of their respective common factors. However, such an approach may result in a loss of efficiency. By contrast, our approach does not suffer from this problem. We establish the consistency of our procedures as the number of assets increases and the number of infill observations for each asset increases. Our asymptotic framework allows for a rich diversity in the relative size of the efficient price process and the microstructure noise process and in the relative size of the common component of the microstructure noise and the idiosyncratic components of the noise. This is important because a number of authors have documented that in frequently traded assets, the microstructure noise component can be quite small. Also, the Epps effect, whereby observed cross-asset correlations shrink with sampling frequency, can be captured in our framework when the idiosyncratic component of the noise is larger element by element than the common component. Our model allows the so-called "weak factors", c.f., Briggs and MacCallum (2003); Onatski (2010) and Freyaldenhoven (2022). We provide a full analysis of the convergence rates of all our estimators, which are affected by the magnitudes of the microstructure noise. We apply our method to the intraday returns of S&P 500 Index constituents. The empirical analysis provides evidence of co-movement of the microstructure noise.

The rest of this chapter is organised as follows. Section 1.2 specifies the model and its assumptions. Section 1.3 proposes the high-frequency PANIC estimation procedure and presents the asymptotic properties for the estimators. Section 1.4 provides finite-sample simulation results and Section 1.5 demonstrates the applicability of our proposed method through an empirical study. Section 1.6 concludes. The proofs of our main results are relegated to Appendix A.

Throughout this chapter, we use $||\cdot||_2$ to denote the Euclidean norm of a vector. For a real symmetric matrix $\mathbf{S}$, we denote its k-th largest eigenvalue and trace by $\mu_k(\mathbf{S})$ and $\mathsf{tr}(\mathbf{S})$, respectively. For any $m\times n$ matrix $\mathbf{M}=(m_{ij})$, let $||\mathbf{M}||_{\mathsf{O}}$, $||\mathbf{M}||_1$, $||\mathbf{M}||_\infty$, $||\mathbf{M}||_F$ and $||\mathbf{M}||_{\max}$ denote the spectral norm, the l_1 norm, the l_∞ norm, the Frobenius norm, and the max norm of $\mathbf{M}$, respectively. Specifically, $\|\mathbf{M}\|_{\mathsf{O}}=\sqrt{\mu_1(\mathbf{M}^\intercal\mathbf{M})}$, $||\mathbf{M}||_1=\max_j\sum_i|m_{ij}|$, $||\mathbf{M}||_\infty=\max_i\sum_j|m_{ij}|$, $\|\mathbf{M}\|_F=\sqrt{\mathsf{tr}(\mathbf{M}^\intercal\mathbf{M})}=\sqrt{\sum_{i,j}m_{ij}^2}$ and $||\mathbf{M}||_{\max}=\max_{i,j}|m_{ij}|$. Let $\mathbf{1}_n$ denote an n-dimensional vector of 1's. Also let $a\vee b$ and $a\wedge b$ denote $\max\{a,b\}$ and $\min\{a,b\}$, and x_+ and x_- denote $\max\{0,x\}$ and $\min\{0,x\}$, respectively.

1.2 Model setup and assumptions

1.2.1 Dual factor structure

Let X_{it} denote the observed log transaction price of stock i at time t, for $i = 1, ..., d$. We allow d to diverge with n, although we have suppressed the subscript n for d. For the sake of simplicity, we assume that price observations of all stocks are synchronously collected, and that price observations for each stock are equidistantly collected in the fixed time interval $[0, T]$. Thus we do not consider non-synchronous trading explicitly. Without loss of generality, we let $T = 1$. Let n be the number of observations and $\Delta = 1/n$. Then, the prices are observed at the time points $t = 0, \Delta, 2\Delta, \ldots, n\Delta$.

We assume that the observed log transaction price, X_{it}, can be decomposed into the unobserved efficient log-price X^*_{it} plus a noise component Z_{it}, i.e.,

$$X_{it} = X^*_{it} + Z_{it} \quad \text{or} \quad \boldsymbol{X}_t = \boldsymbol{X}^*_t + \boldsymbol{Z}_t, \tag{1.2.1}$$

where $\boldsymbol{X}^*_t = (X^*_{1t}, \ldots, X^*_{dt})^\intercal$ and $\boldsymbol{Z}_t = (Z_{1t}, \ldots, Z_{dt})^\intercal$. For each component of X_{it}, we introduce a factor structure (see Assumptions 1.A and 1.B below) and therefore, name the model as a *dual factor model*.

Assumption 1.A. *(Factor Structure for Efficient Log-price)*

(i) The efficient log-price $\boldsymbol{X}^*_t$ follows a factor model of the form,

$$\begin{cases} \mathsf{d}\boldsymbol{X}^*_t &= \boldsymbol{\Lambda}_F \mathsf{d}\boldsymbol{F}_t + \mathsf{d}\boldsymbol{U}_t, \\ \mathsf{d}\boldsymbol{F}_t &= \boldsymbol{\sigma}_{Ft} \mathsf{d}\boldsymbol{B}^F_t, \\ \mathsf{d}\boldsymbol{U}_t &= \boldsymbol{\sigma}_{Ut} \mathsf{d}\boldsymbol{B}^U_t, \end{cases}$$

where $\boldsymbol{\Lambda}_F = (\lambda_{F,ik})_{1\le i\le d, 1\le k\le K_F}$ denotes the $d \times K_F$ matrix of factor loadings, K_F is the number of factors, $\boldsymbol{F}_t = (F_{1t}, ..., F_{K_F t})^\intercal$ denotes latent factors, $\boldsymbol{U}_t = (U_{1t}, ..., U_{dt})^\intercal$ is the idiosyncratic component, $\boldsymbol{\sigma}_{Ft}$ is a $K_F \times K_F$ càd-làg spot volatility matrix for factors, $\boldsymbol{\sigma}_{Ut}$ is a $d \times d$ càd-làg spot volatility matrix for idiosyncratic errors, and $\boldsymbol{B}^F_t = (B^F_{1t}, \ldots, B^F_{K_F t})^\intercal$ and $\boldsymbol{B}^U_t = (B^U_{1t}, \ldots, B^U_{dt})^\intercal$ are independent Brownian motions.

(ii) There exists a locally bounded process Q_t such that $\|\boldsymbol{\sigma}_{Ft}\boldsymbol{\sigma}^\intercal_{Ft}\|_{\max}$ and $\|\boldsymbol{\sigma}_{Ut}\boldsymbol{\sigma}^\intercal_{Ut}\|_{\max}$ are bounded

by Q_t for all paths. [1]

We adopt the same high-frequency factor structure in Bollerslev *et al.* (2019). The existence of uniform bounds on all volatility processes is natural in a continuous-time factor model (see, for example, Aït-Sahalia and Xiu (2017)), and is necessary to the development of the large dimensional asymptotic results. Locally bounded processes are more general than uniformly bounded processes. For example, $Q_t \equiv Q_0$, where Q_0 follows a normal distribution, is not a uniformly bounded process but is a locally bounded process.

Remark: Our model inherits several limitations from Bollerslev *et al.* (2019)'s factor structure, which does not allow for drift terms or jump terms in the diffusion model. We refer to Dai *et al.* (2019) and Aït-Sahalia and Xiu (2019) for high-frequency factor models that both common factors and idiosyncratic errors follow (continuous or general) Itô semimartingale processes. Moreover, our model does not allow for time-varying loadings or an infinite number of factors. For these extensions, we refer to Su and Wang (2017); Fan *et al.* (2011, 2016b); Kong (2017, 2018); Aït-Sahalia *et al.* (2020) and Kong *et al.* (2023), respectively.

Assumption 1.B. *(Factor Structure for Market Microstructure Noise)*

The microstructure noise $\boldsymbol{Z}_t$ follows a factor model whose magnitude may depend on the sampling frequency, that is

$$\boldsymbol{Z}_t = \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{G}_t + \mathbf{D}_V \boldsymbol{V}_t, \tag{1.2.2}$$

where $\boldsymbol{\Lambda}_G = (\lambda_{G,ik})_{1\le i\le d, 1\le k\le K_G}$ denotes the $d \times K_G$ matrix of factor loadings with K_G being the number of factors for the microstructure noise, $\boldsymbol{G}_t = (G_{1t}, ..., G_{K_G t})^\intercal$ denotes the latent factors, $\boldsymbol{V}_t = (V_{1t}, ..., V_{dt})^\intercal$ is the vector of idiosyncratic components, and $\mathbf{D}_G$ and $\mathbf{D}_V$ are two diagonal matrices satisfying $\mu_1(\mathbf{D}_G) = O(n^{\bar{\tau}_G^\diamond})$, $\mu_1(\mathbf{D}_G^{-1}) = O(n^{-\underline{\tau}_G^\diamond})$ and $\mu_1(\mathbf{D}_V) = O(n^{\bar{\tau}_V^\diamond})$, where $\bar{\tau}_G^\diamond$, $\underline{\tau}_G^\diamond$, and $\bar{\tau}_V^\diamond$ are constants, whose values may be positive, negative or zero.

In a stock market, microstructure noise can be much larger in magnitude than efficient prices when sampling frequency is high. However, the observed prices are close to the efficient prices in a long horizon due to the efficiency of the market. In other words, efficient returns accumulate over time, but microstructure noise does not. Therefore, when the sampling frequency is low, the

[1] A process $\{Q_t\}_{t\in[0,1]}$ is locally bounded if there exists a sequence of stopping times $\{\tau_s\}$, with $0 \le \tau_s \le \tau_{s+1}$ for $s = 1, 2, ...,$ and $\tau_s \to \infty$, a.s., as $s \to \infty$, such that the stopped process $1_{\{\tau_s>0\}} Q_{t\wedge\tau_s}$ is a uniformly bounded process for each s, where $1_{\{\cdot\}}$ is an indicator function.

efficient prices will dominate microstructure noise in magnitude, while the reverse may happen when sampling frequency is high. The introduction of $\mathbf{D}_G$ and $\mathbf{D}_V$ in (1.2.2) allows our model to capture this phenomenon (see also Kalnina and Linton (2008) for a model where the microstructure noise can be large or small, depending on the value of a magnitude parameter). With such a setup, the magnitude of $\mathbf{\Lambda}_G$, $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ is independent of n. A similar treatment can be found in Kim *et al.* (2016). Our model is an extension of the model of Bollerslev *et al.* (2019), who only consider $\mathbf{D}_G = \mathbf{I}_{K_G}$ and $\mathbf{D}_V = \mathbf{I}_d$, where $\mathbf{I}_K$ denotes the $K \times K$ identity matrix. However, they use a similar setting when generating simulation data (also see Section 1.4) without discussing the asymptotic impacts of $\mathbf{D}_G$ and $\mathbf{D}_V$.

To introduce the first-differenced form of the dual factor model, we use little letters to denote the first-order differences of random variables. Specifically, define the return as $\boldsymbol{x}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-\Delta}$, and the efficient return (or frictionless return) as $\boldsymbol{x}_t^* = \int_{t-\Delta}^t \mathsf{d}\boldsymbol{X}_s^* = \boldsymbol{X}_t^* - \boldsymbol{X}_{t-\Delta}^*$. Denote $\boldsymbol{f}_t = \int_{t-\Delta}^t \boldsymbol{\sigma}_{fs}\mathsf{d}\boldsymbol{B}_s^F = \boldsymbol{F}_t - \boldsymbol{F}_{t-\Delta}$, $\boldsymbol{g}_t = \boldsymbol{G}_t - \boldsymbol{G}_{t-\Delta}$, $\boldsymbol{z}_t = \boldsymbol{Z}_t - \boldsymbol{Z}_{t-\Delta}$, $\boldsymbol{u}_t = \int_{t-\Delta}^t \boldsymbol{\sigma}_{Us}\mathsf{d}\boldsymbol{B}_s^U = \boldsymbol{U}_t - \boldsymbol{U}_{t-\Delta}$, and $\boldsymbol{v}_t = \boldsymbol{V}_t - \boldsymbol{V}_{t-\Delta}$. Then by (1.2.1)–(1.2.2), we can write the dual factor model as

$$\begin{cases} \boldsymbol{x}_t &= \boldsymbol{x}_t^* + \boldsymbol{z}_t, \\ \boldsymbol{x}_t^* &= \mathbf{\Lambda}_F \boldsymbol{f}_t + \boldsymbol{u}_t, \\ \boldsymbol{z}_t &= \mathbf{\Lambda}_G \mathbf{D}_G \boldsymbol{g}_t + \mathbf{D}_V \boldsymbol{v}_t. \end{cases} \tag{1.2.3}$$

Combining the factor structures for $\boldsymbol{x}_t^*$ and $\boldsymbol{z}_t$, we have

$$\begin{aligned} \boldsymbol{x}_t &= \mathbf{\Lambda}_F \boldsymbol{f}_t + \boldsymbol{u}_t + \mathbf{\Lambda}_G \mathbf{D}_G \boldsymbol{g}_t + \mathbf{D}_V \boldsymbol{v}_t \\ &= \mathbf{\Lambda}_H \mathbf{D}_H \boldsymbol{h}_t + \boldsymbol{w}_t, \end{aligned} \tag{1.2.4}$$

where $\boldsymbol{h}_t = (\boldsymbol{f}_t^\intercal, n^{-1/2}\boldsymbol{g}_t^\intercal)^\intercal$, $\mathbf{\Lambda}_H = (\mathbf{\Lambda}_F, \mathbf{\Lambda}_G)$, $\boldsymbol{w}_t = \boldsymbol{u}_t + \mathbf{D}_V \boldsymbol{v}_t$ and $\mathbf{D}_H = \mathsf{diag}(\mathbf{I}_{K_F}, n^{1/2}\mathbf{D}_G)$. This can be seen as a factor structure for $\boldsymbol{x}_t$ with $\boldsymbol{h}_t = (\boldsymbol{f}_t^\intercal, n^{-1/2}\boldsymbol{g}_t^\intercal)^\intercal$ being the factors and $\mathbf{\Lambda}_H = (\mathbf{\Lambda}_F, \mathbf{\Lambda}_G)$ being the factor loadings. Note that $\boldsymbol{g}_t$ is divided by $n^{1/2}$ in $\boldsymbol{h}_t$ so that both components of $\boldsymbol{h}_t$ are of the same magnitude. Consequently, the magnitude matrix $\mathbf{D}_G$ is multiplied by $n^{1/2}$ in $\mathbf{D}_H$.

Define

$$\bar{\tau}_G^+ = (\bar{\tau}_G)_+, \quad \underline{\tau}_G^- = (\underline{\tau}_G)_-, \quad \text{and } \bar{\tau}_V^+ = (\bar{\tau}_V)_+,$$

with

$$\bar{\tau}_G = 1/2 + \bar{\tau}_G^\diamond, \quad \underline{\tau}_G = 1/2 + \underline{\tau}_G^\diamond, \quad \text{and } \bar{\tau}_V = 1/2 + \bar{\tau}_V^\diamond.$$

Since the factors for efficient returns, $\boldsymbol{f}_t$, are of order $n^{-1/2}$ and the largest component of $\mathbf{D}_G\boldsymbol{g}_t$ is of order $n^{\bar{\tau}_G^\diamond}$, a positive value of $\bar{\tau}_G^+$ indicates that the largest microstructure noise factor dominates the efficient return factors. On the other hand, the smallest component of $\mathbf{D}_G\boldsymbol{g}_t$ is of order $n^{\underline{\tau}_G^\diamond}$, and a negative value of $\underline{\tau}_G^-$ indicates that the efficient return factors dominate the smallest microstructure noise factor. Similarly, a positive value of $\bar{\tau}_V^+$ indicates that the largest idiosyncratic microstructure noise dominates the idiosyncratic efficient returns.

Remark: (Identification) First of all, a gap between the eigenvalues of the covariance matrices of $\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{h}_t$ and $\boldsymbol{w}_t$ (see Assumption 5*) ensures the identifiability of the factor space spanned by $\boldsymbol{h}_t$. Secondly, as in conventional factor models, the factors $\boldsymbol{h}_t$ and factor loadings $\boldsymbol{\Lambda}_H$ are not separately identifiable. For identification up to a rotation matrix, a normalisation condition is imposed on $\boldsymbol{\Lambda}_H$ (see Assumption 4). Lastly, for the identifiability of the factor spaces spanned by $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$, it requires another eigenvalue gap between the covariance matrices of the two types of factors but in the cumulated form. This is implicit in Assumptions 1 and 2, as the types of processes for $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ are different in nature, especially when they are cumulated.

Denote the number of independent factors in the factor model (1.2.4) as K_H. If $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ are collinear, K_H will be less than $K_F + K_G$. In such a case, the efficient prices and the microstructure noise are not separable and the dichotomous structure fails. Thus for identification purposes, we exclude this situation and make the following assumptions.

Assumption 1.C. *(Independence between efficient prices and microstructure noise) The discrete time series $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ are independent of the continuous-time processes $\boldsymbol{F}_t$ and $\boldsymbol{U}_t$.*

Remark: It is prevalent to assume independence between price components due to fundamental security value and noise attributable to market rules and trading mechanisms. The main reason for the independence assumption is modelling simplicity, so that the two components can be identified and interpreted. Recently, there has been some research that allows correlation between the efficient price and the microstructure noise, such as Kunitomo and Kurisu (2021). But we will not pursue this in the current chapter.

Assumption 1.D. *(Factor loadings) The factor loadings matrix $\boldsymbol{\Lambda}_H$ satisfies*

$$\|\boldsymbol{\Lambda}_H\|_{\max} = O(1), \quad \text{and} \quad \|\boldsymbol{\Lambda}_H^\intercal\boldsymbol{\Lambda}_H/d - \mathbf{I}_{K_H}\| = o(1),$$

where $\mathbf{I}_{K_H}$ denotes the $K_H \times K_H$ identify matrix.

As discussed in Aït-Sahalia and Xiu (2017), the condition $\mathscr{h}^\intercal \mathscr{h} = \mathbf{I}_{K_H}$, where $\mathscr{h} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n)^\intercal$, is not appropriate in the high-frequency factor model setting because the probability limit of $\mathscr{h}^\intercal \mathscr{h}$ involves the volatility process $\boldsymbol{\sigma}_{Ft}$, which is path-dependent rather than deterministic. Thus, we adopt the normalisation condition on $\boldsymbol{\Lambda}_H$ in Assumption 1.D. This identification condition looks like a classical strong factor assumption. However, since we have introduced the magnitude matrix $\mathbf{D}_H$, our model allows for 'weak factors' in the sense that the divergence of $\mu_{K_H}(\mathbf{D}_H^\intercal \boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Lambda}_H \mathbf{D}_H)$ can be at a slower rate than d (this is different from the usual weak factors which are defined as being non-pervasive across cross-sections, see Onatski (2010); Fan and Liao (2022); Anatolyev and Mikusheva (2022); Giglio *et al.* (2021); Uematsu and Yamagata (2023b,a); Freyaldenhoven (2022)). Moreover, the theory can be extended to allow more general $\mathbf{D}_H$, especially for the first diagonal block corresponding to $\boldsymbol{f}_t$. But this can introduce an identification issue that requires setting one of the factors to have a unit magnitude. For simplicity, we do not introduce a non-identity magnitude matrix for $\boldsymbol{f}_t$.

For easy reference, we summarise the notation used for variables and factors in Table 1.1. We use different fonts to distinguish between matrices, vectors and scalars. For example, $\mathscr{X} = (\boldsymbol{X}_\Delta, \ldots, \boldsymbol{X}_{n\Delta})^\intercal$ is an $n \times d$ matrix of observed prices, $\boldsymbol{X}_{s\Delta}^\intercal$ is its s-th row, and $X_{i,s\Delta}$ is the (s,i)-entry of the matrix $\mathscr{X}$. Following the same rule, other variables are defined analogously.

Table 1.1: Notations for variables/factors in the dual factor model

Variables	Cumulation Form			First-difference Form		
	Matrix -wise	Row -wise	Element -wise	Matrix -wise	Row -wise	Element -wise
Observed price	$\mathscr{X}$	$\boldsymbol{X}_t$	X_{it}	$\mathscr{x}$	$\boldsymbol{x}_t$	x_{it}
Efficient price (EP)	$\mathscr{X}^*$	$\boldsymbol{X}_t^*$	X_{it}^*	$\mathscr{x}^*$	$\boldsymbol{x}_t^*$	x_{it}^*
Microstructure noise (MN)	$\mathscr{Z}$	$\boldsymbol{Z}_t$	Z_{it}	$\mathscr{z}$	$\boldsymbol{z}_t$	z_{it}
Factors for EP	$\mathscr{F}$	$\boldsymbol{F}_t$	F_{jt}	$\mathscr{f}$	$\boldsymbol{f}_t$	f_{jt}
Factors for MN	$\mathscr{G}$	$\boldsymbol{G}_t$	G_{jt}	$\mathscr{g}$	$\boldsymbol{g}_t$	g_{jt}
Total factors	$\mathscr{H}$	$\boldsymbol{H}_t$	H_{jt}	$\mathscr{h}$	$\boldsymbol{h}_t$	h_{jt}
Idiosyncratic errors for EP	$\mathscr{U}$	$\boldsymbol{U}_t$	U_{it}	$\mathscr{u}$	$\boldsymbol{u}_t$	u_{it}
Idiosyncratic errors for MN	$\mathscr{V}$	$\boldsymbol{V}_t$	V_{it}	$\mathscr{v}$	$\boldsymbol{v}_t$	v_{it}
Total idiosyncratic errors	$\mathscr{W}$	$\boldsymbol{W}_t$	W_{it}	$\mathscr{w}$	$\boldsymbol{w}_t$	w_{it}

[1] The first dimension of the matrices in the table is set as time, while the second dimension is set as a stock or a factor. We have $t = \Delta, \cdots, n\Delta$, $1 \le i \le d$ and $1 \le j \le K$, where $K = K_F$, K_G, or K_H, depending on the circumstance. Note that in the subscript of the element-wise notation, we write the column index first.

[2] Although when $t = 0$, $\boldsymbol{X}_t$ is observable, t starts from Δ in the matrix-wise notation for both cumulation form and first-difference form, for the sake of consistency.

1.2.2 Covariance structure

Denote the integrated covolatility matrices of $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ as $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_U$, respectively. That is $\boldsymbol{\Sigma}_F = \int_0^1 \boldsymbol{\sigma}_{Ft}\boldsymbol{\sigma}_{Ft}^\intercal \mathrm{d}t$ and $\boldsymbol{\Sigma}_U = \int_0^1 \boldsymbol{\sigma}_{Ut}\boldsymbol{\sigma}_{Ut}^\intercal \mathrm{d}t$. Then the factor structure for the efficient prices leads to the following identity,

$$\boldsymbol{\Sigma}_{x^*} = \boldsymbol{\Lambda}_F \boldsymbol{\Sigma}_F \boldsymbol{\Lambda}_F^\intercal + \boldsymbol{\Sigma}_U, \tag{1.2.5}$$

where $\boldsymbol{\Sigma}_{x^*}$ is the integrated covolatility matrix of $\boldsymbol{x}_t^*$. For the microstructure noise, we will assume that its components $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ are stationary. So its covariance matrix has the following identity,

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Lambda}_G \mathbf{D}_G \boldsymbol{\Sigma}_g \mathbf{D}_G \boldsymbol{\Lambda}_G^\intercal + \mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V, \tag{1.2.6}$$

where $\boldsymbol{\Sigma}_z = \mathsf{Var}(\boldsymbol{z}_t)$, $\boldsymbol{\Sigma}_g = \mathsf{Var}(\boldsymbol{g}_t)$, and $\boldsymbol{\Sigma}_v = \mathsf{Var}(\boldsymbol{v}_t)$. Hence, for the integrated covolatility matrix of observed prices, we combine (1.2.5) and (1.2.6) to obtain

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\Sigma}_h \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal + \boldsymbol{\Sigma}_w, \tag{1.2.7}$$

where

$$\begin{cases} \boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_{x^*} + n\boldsymbol{\Sigma}_z, \\ \boldsymbol{\Sigma}_h = \mathsf{diag}(\boldsymbol{\Sigma}_F, \boldsymbol{\Sigma}_g), \\ \boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_U + n\mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V. \end{cases}$$

The matrix $\boldsymbol{\Lambda}_H \mathbf{D}_H \boldsymbol{\Sigma}_h \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal$ has K_H positive eigenvalues, which follows from the positive definiteness of $\boldsymbol{\Sigma}_h$ and the full-column-rankness of $\boldsymbol{\Lambda}_H$ under Assumption 1.D. The reason we multiply $\boldsymbol{\Sigma}_z$ by n is to be consistent with its sample estimates. Note that $\widehat{\boldsymbol{\Sigma}}_z = n^{-1}\sum_{i=1}^n \boldsymbol{z}_t \boldsymbol{z}_t^\intercal$ is an approximation of $\boldsymbol{\Sigma}_z$ while $\widehat{\boldsymbol{\Sigma}}_{x^*} = \sum_{i=1}^n (\boldsymbol{x}_t^*)(\boldsymbol{x}_t^*)^\intercal$ is an approximation of $\boldsymbol{\Sigma}_{x^*}$.

In order to identify the factor structure, we make some sparsity assumptions on the idiosyncratic volatility matrices, as in Fan *et al.* (2013) and Aït-Sahalia and Xiu (2017).

Assumption 1.E*. *(Sparsity of idiosyncratic integrated covolatility matrices) There exist $m_{U,d}$ and $m_{v,d}$, which are bounded away from zero and may diverge as d goes to infinity, such that the integrated covolatility matrices of the idiosyncratic components satisfy*

$$\|\boldsymbol{\Sigma}_U\|_1 = O_P(m_{U,d}), \quad \|\boldsymbol{\Sigma}_v\|_1 = O(m_{v,d}), \quad m_{w,nd}/(dn^{2\underline{\tau}_G}) \to 0 \text{ with } m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V} m_{v,d}.$$

The sparsity conditions are more general than the bounded eigenvalue condition for the idiosyncratic volatility matrix in approximate factor models, c.f., Chamberlain and Rothschild (1983), as $\|\mathbf{\Sigma}_U\|_{\mathsf{O}} \leq \|\mathbf{\Sigma}_U\|_1$, $\|\mathbf{\Sigma}_v\|_{\mathsf{O}} \leq \|\mathbf{\Sigma}_v\|_1$, and $m_{U,d}$ and $m_{v,d}$ may diverge as d goes to infinity. In addition, it is worth pointing out that the condition $m_{w,nd}/(dn^{2\bar{\tau}_G}) \to 0$ is sufficient for the identification of the factors. Specifically, under the condition $m_{w,nd}/(dn^{2\bar{\tau}_G}) \to 0$, the smallest positive eigenvalue of $\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal$ has a lower bound of order $dn^{2\bar{\tau}_G}$ (see Lemma A.1.1), which is larger than $m_{w,nd}$, the order of the largest eigenvalue of $\mathbf{\Sigma}_w$. In order to obtain consistent estimates, we further require $m_{w,nd}/(d^{1/2}n^{2\bar{\tau}_G}) \to 0$. This leads to the following stronger version of Assumption 1.E*.

Assumption 1.E. *(Sparsity of idiosyncratic integrated covolatility matrices) The integrated covolatility matrices of the idiosyncratic components satisfy*

$$\|\mathbf{\Sigma}_U\|_1 = O_P(m_{U,d}), \quad \|\mathbf{\Sigma}_v\|_1 = O(m_{v,d}), \quad m_{w,nd}/(d^{1/2}n^{2\bar{\tau}_G}) \to 0 \text{ with } m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V}m_{v,d}.$$

Assumption 1.E is a stronger version of Assumption **1.E***. That estimation of an approximate factor model requires more strict sparsity conditions than identification is also observed in Aït-Sahalia and Xiu (2017).

We also make the following assumption on the stationarity of the microstructure noise components.

Assumption 1.F. *(Stationary and sub-Weibull microstructure noise)*

(i) The series $\{\boldsymbol{G}_t, \boldsymbol{V}_t\}$ *is strictly stationary. In addition, the eigenvalues of* $\mathbf{\Sigma}_G$ *and* $\mathbf{\Sigma}_V$ *are bounded away from zero, and* $\mathsf{E}[G_{jt}] = \mathsf{E}[V_{it}] = \mathsf{E}[G_{jt}V_{it}] = 0$ *for all* $1 \leq i \leq d$, $1 \leq j \leq K_G$ *and* $t = 0, \Delta, \ldots, n\Delta$.

(ii) There exist positive constants $C_\alpha > 0$ *and* $\gamma_1 > 0$ *such that the strong mixing sequence* $\alpha(\cdot)$ *of the series* $\{\boldsymbol{G}_t, \boldsymbol{V}_t\}$ *satisfies* $\alpha(s\Delta) \leq C_\alpha \exp(-s^{\gamma_1})$.

(iii) There exist $b_1 > 0$, $\gamma_2 > 0$, *with* $\gamma_1^{-1} + 3\gamma_2^{-1} > 1$, *such that for all* $c > 0$, *we have*

$$\max_{1\leq j\leq K_G} \mathsf{P}\left(|G_{jt}| > c\right) \leq \exp(1 - (c/b_1)^{\gamma_2}) \tag{1.2.8}$$

and

$$\max_{1\le i\le d} \mathsf{P}\left(|V_{it}| > c\right) \le \exp(1-(c/b_1)^{\gamma_2}), \tag{1.2.9}$$

for $t = 0, \Delta, \ldots, n\Delta$.

(iv) There exist $b_2 > 0$, $\gamma_3 > 0$, *with* $\gamma_1^{-1} + 3\gamma_3^{-1} > 1$, *such that for all* $c > 0$, *we have*

$$\max_{1\le j\le K_H} \mathsf{P}\left(d^{-1/2}|\boldsymbol{\lambda}_{H,\cdot,j}^{\intercal}\boldsymbol{V}_t| > c\right) \le \exp(1-(c/b_2)^{\gamma_3}), \tag{1.2.10}$$

for $t = 0, \Delta, \ldots, n\Delta$, *where* $\boldsymbol{\lambda}_{H,\cdot,j}$ *is the* j*-th column of* $\boldsymbol{\Lambda}_H$*;*

(v) Let $1/\gamma = 1/\gamma_1 + 3/(\gamma_2 \vee \gamma_3)$. *Then,* $(\log d)^{2/\gamma-1} = o(n)$.

Assumptions 1.F(i) and (ii) are similar to Assumption 3.2(i) and Assumption 3.3 in Fan *et al.* (2013). The strong mixing condition is more general than Assumptions 1 and 3 in Barigozzi *et al.* (2021) in which the common component and the idiosyncratic component of the microstructure noise are both assumed to be linear processes. The series $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$ can be serially correlated, which is more general than the assumption in Kim *et al.* (2016). Moreover, $\boldsymbol{V}_t$ can be cross-sectionally dependent, in which case, (1.2.2) gives an approximate factor model for the microstructure noise. More general assumptions can be found in Bai and Ng (2002), which permits weakly correlated idiosyncratic errors. Assumption 1.F(iii) requires exponential-type tails for the distributions of G_t and V_t, which allows us to apply the large deviation theory (see Lemma A.2.2) to facilitate our proofs. It allows for sub-Gaussian tails, sub-exponential tails and even heavier tails when $\gamma_2, \gamma_3 < 1$, and is commonly used in estimation of large-dimensional volatility matrices, c.f., Fan *et al.* (2013) and Tao *et al.* (2013b). With lengthier proofs of the asymptotic results, this condition can be weakened to a finite moment condition as in Bai and Ng (2002). However, the convergence rates for the estimators would be slower (see Lemma D.2 of Li *et al.* (2023) for an example). On the other hand, if the microstructure noise has a heavy tail that violates Assumption 1.F(iii), then robust estimation can be used to mitigate the influence of heavy tails (see, for example, Fan and Kim (2018)).

Assumption 1.F(iv) is an additional exponential tail condition, which guarantees that

$$\left\| (nd)^{-1}\sum_{s=1}^{n} \boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_v\boldsymbol{\Lambda}_H \right\|_{\max} = O_P((\log d/n)^{1/2}),$$

see Lemma A.2.2(iv). Like Assumption 1.F(iii), it can also be weakened to a finite moment condition

(see Assumption F.3 of Bai (2003) and Assumption 3.4(iii) of Fan *et al.* (2013))

$$\max_{1\leq j\leq K_H} \mathsf{E}[(d^{-1/2}|\boldsymbol{\lambda}_{H,\cdot,j}^{\intercal}\boldsymbol{v}_t|)^4] < C, \tag{1.2.11}$$

for some positive constant C. Assumption 1.F(v) presents a trade-off between the mixing and tail conditions and the dimension d.

1.3 Estimation procedure and asymptotic results

In this section, we develop a two-step estimation procedure for the common factors, $\mathscr{F}$ and $\mathscr{G}$, of the dual factor model. The estimation procedure is based on two PCA procedures and so we call it Double PCA or DPCA. The asymptotic results are presented step by step, so as to provide a better understanding of what the intermediate estimators estimate. In the first step, the factors and factor loadings for the combined factor model (1.2.4) in the first-difference form are estimated. In the second step, we cumulate the normalised factors and separate the efficient price factors and the microstructure noise factors. For the time being, we assume that the number of factors K_F and K_G are known, and $K_H = K_F + K_G$. We will discuss how to determine them in Section 1.4.1.

We also require the following assumption, which restricts the relation between n and d.

Assumption 1.G. *(Relation between n and d) As $n \to \infty$,*

(i)

$$n^{1+4\underline{\tau}_G^- - 4(\bar{\tau}_G^+ \vee \bar{\tau}_V^+)}/(\log d) \to \infty,$$

and

(ii)

$$n^{1+4\underline{\tau}_G^- - 4(\bar{\tau}_G^+ \vee \bar{\tau}_V^+)} m_{w,nd}^2/(d^2 \log d) \to 0.$$

When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = 0$ and $m_{w,nd} = O(1)$, Assumption 1.G degenerates to $n/(\log d) \to \infty$ and $n/(d^2 \log d) \to 0$, which are similar to the corresponding condition in Theorem 3.1 of Fan *et al.* (2013) and assumption A1 of Tao *et al.* (2013b).

1.3.1 First step: PCA estimation in first-difference form

The first step in our estimation procedure is to apply PCA to the first-difference form of the dual factor model to extract estimates of all the factors and factor loadings (for both the efficient price and the microstructure noise).

There are two ways to implement PCA to obtain estimates of the factors and factor loadings. If we normalise the factor loadings, the PCA estimator solves the following optimisation problem,

$$\begin{cases} \min\limits_{\boldsymbol{\Lambda}_H, \hbar} & \|\boldsymbol{x} - \hbar \boldsymbol{\Lambda}_H^\intercal\|_F, \\ \text{s.t.} & \boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Lambda}_H / d = \mathbf{I}_{K_H}. \end{cases} \tag{1.3.1}$$

Computationally, we conduct an eigen-decomposition of the $d \times d$ matrix $\boldsymbol{x}^\intercal \boldsymbol{x}$, and obtain $\widehat{\boldsymbol{\Lambda}}_H = \left(\widehat{\boldsymbol{\lambda}}_{H1}, \ldots, \widehat{\boldsymbol{\lambda}}_{Hd}\right)^\intercal$, which is the $d \times K_H$ matrix consisting of the K_H eigenvectors (multiplied by $\sqrt{d}$) corresponding to the K_H largest eigenvalues of $\boldsymbol{x}^\intercal \boldsymbol{x}$. We then obtain

$$\widehat{\hbar} = \boldsymbol{x} \widehat{\boldsymbol{\Lambda}}_H / d = \left(\widehat{\boldsymbol{h}}_\Delta, \ldots, \widehat{\boldsymbol{h}}_{n\Delta}\right)^\intercal. \tag{1.3.2}$$

The resulting $\widehat{\boldsymbol{\Lambda}}_H$ and $\widehat{\hbar}$ are the estimators of $\boldsymbol{\Lambda}_H$ and $\mathbf{D}_H \hbar$, respectively, up to a rotation.

Alternatively, if we normalise the factors, the PCA estimator solves the following optimisation problem,

$$\begin{cases} \min\limits_{\boldsymbol{\Lambda}_H, \hbar} & \|\boldsymbol{x} - \hbar \boldsymbol{\Lambda}_H^\intercal\|_F, \\ \text{s.t.} & \hbar^\intercal \hbar = \mathbf{I}_{K_H}. \end{cases}$$

Computationally, we conduct PCA* (we add a star to distinguish it from the PCA in (1.3.1), which is based on the normalisation of the factor loadings) on the $n \times n$ matrix $\boldsymbol{x}\boldsymbol{x}^\intercal$, and get $\widehat{\hbar}^\star = \left(\widehat{\boldsymbol{h}}^\star_\Delta, \ldots, \widehat{\boldsymbol{h}}^\star_{n\Delta}\right)^\intercal$ denoting the $n \times K_H$ matrix consisting of the K_H eigenvectors corresponding to the K_H largest eigenvalues of $\boldsymbol{x}\boldsymbol{x}^\intercal$. Then we obtain

$$\widehat{\boldsymbol{\Lambda}}^\star_H = \boldsymbol{x}^\intercal \widehat{\hbar}^\star = \left(\widehat{\boldsymbol{\lambda}}^\star_{H1}, \ldots, \widehat{\boldsymbol{\lambda}}^\star_{Hd}\right)^\intercal. \tag{1.3.3}$$

The resulting $\widehat{\boldsymbol{\Lambda}}^\star_H$ and $\widehat{\hbar}^\star$ are the estimators of $\boldsymbol{\Lambda}_H \mathbf{D}_H (\hbar^\intercal \hbar)^{1/2}$ and $\hbar (\hbar^\intercal \hbar)^{-1/2}$, respectively, up to a rotation.

The two restrictions in the optimisation problems should be distinguished from the identification conditions in Assumption 1.D. The restrictions are imposed on the estimators whereas the identification conditions are imposed on the true parameters. Moreover, it is easy to show that both $\widehat{\hbar}$

and $\widehat{\mathscr{H}}^\star$ are eigenvectors of $\mathscr{X}\mathscr{X}^\intercal$, and thereby, they span the same space. Specifically, we can show the following relation

$$\widehat{\mathscr{H}}^\star = \widehat{\mathscr{H}}(\widehat{\mathscr{H}}^\intercal\widehat{\mathscr{H}})^{-1/2} \quad \text{and} \quad \widehat{\mathbf{\Lambda}}_H^\star = \widehat{\mathbf{\Lambda}}_H(\widehat{\mathscr{H}}^\intercal\widehat{\mathscr{H}})^{1/2}. \tag{1.3.4}$$

In this sense, the two ways of PCA are equivalent. However, when d is much larger than n, it is computationally more convenient to conduct PCA on the $n \times n$ matrix $\mathscr{X}\mathscr{X}^\intercal$, and vice versa. Then the relation in (1.3.4) can be used to get the desired form of estimates. Bai and Li (2012) point out that the analysis of one PCA representation will carry over to the other by switching the roles of n and d and the role of the factor loadings and factors, and thus it is sufficient to carefully examine the asymptotic properties of one representation.

In the second step of our estimation procedure, we will separate the efficient price factors and the microstructure noise factors. The final estimators of $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ based on the above two ways of PCA will not be equivalent. This is because the first-step factor estimators will be fed into the second step in cumulative form for another PCA, and whether the factors from the first step are normalised will affect the final results. In Section 1.4, we will compare the small-sample performance of estimators based on the two different ways of PCA in the first step. We will see that PCA* always outperforms PCA. Hence, we will use PCA* for our first step and derive the asymptotic theory based on it.

The following theorem shows the uniform rate of convergence for $\widehat{\mathbf{\Lambda}}_H^\star$ and $\widehat{\mathscr{H}}^\star$ of the dual factor model. For ease of exposition, we denote

$$a_{nd} = (\log d)^{1/2}\frac{n^{\bar{\tau}_V^+ + \bar{\tau}_G^+ \vee \bar{\tau}_V^+}}{n^{1/2}} + \frac{m_{w,nd}}{d}, \tag{1.3.5}$$

$$\widetilde{a}_{nd} = (\log d)^{1/2}\frac{n^{\bar{\tau}_V^+ + \bar{\tau}_G^+ \vee \bar{\tau}_V^+}}{n^{1/2}} + \frac{m_{w,nd}}{d^{1/2}}, \tag{1.3.6}$$

$$b_{nd} = (\log(nd))^{1/(\gamma_2\wedge 1)} \cdot n^{-2\underline{\tau}_G^-} \cdot a_{nd} + (\log n)^{1/(\gamma_3\wedge 1)} \cdot d^{-1/2}, \tag{1.3.7}$$

and

$$\widetilde{b}_{nd} = (\log(nd))^{1/(\gamma_2\wedge\gamma_3\wedge 1)} \cdot n^{-2\underline{\tau}_G^-} \cdot \widetilde{a}_{nd}. \tag{1.3.8}$$

It is easy to check that $a_{nd} < \widetilde{a}_{nd}$ and $b_{nd} < \widetilde{b}_{nd}$.

Theorem 1.3.1. *Suppose that Assumptions 1.A–1.G are satisfied. We have*

(i)

$$\left\|\widehat{\mathbf{\Lambda}}_H^\star - \mathbf{\Lambda}_H\mathbf{D}_H(\mathscr{H}^\intercal\mathscr{H})^{1/2}\mathbf{R}^\star\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G^-} \cdot a_{nd}\right), \tag{1.3.9}$$

where the rotation matrix $\mathbf{R}^\star$ *is defined by*

$$\mathbf{R}^\star = d^{1/2}(\hbar^\intercal \hbar)^{-1/2}\mathbf{\Lambda}_H^\intercal \widehat{\mathbf{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{-1/2},$$

in which $\widehat{\mathbf{D}}_{x,K_H} = d\widehat{\hbar}^\intercal\widehat{\hbar} = \widehat{\mathbf{\Lambda}}_H^{\star\intercal}\widehat{\mathbf{\Lambda}}_H^\star$ *is a* $K_H \times K_H$ *diagonal matrix with the diagonal elements being the first* K_H *largest eigenvalues of* $x^\intercal x$ *arranged in a descending order;*

(ii)

$$\left\|\widehat{\hbar}^{\star\intercal} - (\mathbf{R}^\star)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_{\mathsf{O}} = O_P\left(n^{-2\underline{\tau}_G^-}\cdot\widetilde{a}_{nd}\right), \tag{1.3.10}$$

and

$$\left\|\widehat{\hbar}^{\star\intercal} - (\mathbf{R}^\star)^{-1}(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal\right\|_{\max} = O_P\left(n^{-1/2+\bar{\tau}_V^+-\underline{\tau}_G^-}\cdot b_{nd}\right); \tag{1.3.11}$$

(iii) $\widehat{\hbar}^\star\widehat{\mathbf{\Lambda}}_H^{\star\intercal} = \widehat{\hbar}\widehat{\mathbf{\Lambda}}_H^\intercal$ *and*

$$\left\|\widehat{\hbar}^\star\widehat{\mathbf{\Lambda}}_H^{\star\intercal} - \hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\right\|_{\max} = O_P\left(n^{-1/2+\bar{\tau}_V^+}\cdot b_{nd}\right); \tag{1.3.12}$$

(iv) $\mathbf{R}^\star$ *is an asymptotically orthogonal matrix, that is*

$$\|\mathbf{R}^{\star\intercal}\mathbf{R}^\star - \mathbf{I}_{K_H}\|_{\mathsf{O}} = O_P(n^{-2\underline{\tau}_G^-}\cdot a_{nd}). \tag{1.3.13}$$

Part (i) of this theorem gives the uniform convergence rate of our factor loadings estimator. Given the definition of a_{nd} in (1.3.5), this convergence rate is close to the classic element-wise convergence rate, $n^{-1/2}+d^{-1}$, of factor loadings estimators (see Theorem 2 in Bai and Ng (2002) for low-frequency factor models and Lemma K.7 in the Appendix of Pelger (2019) for high-frequency factor models). When $\bar{\tau}_G = \underline{\tau}_G = 0 \geq \bar{\tau}_V$, the convergence rate in (i) is $(\log d)^{1/2}n^{-1/2}+m_{w,nd}d^{-1}$, which is faster than the convergence rate, $(\log d)^{1/2}n^{-1/2}+m_{w,nd}d^{-1/2}$, in Theorem 5 of Aït-Sahalia and Xiu (2017).

Part (ii) of Theorem 1.3.1 shows that the estimator $\widehat{\hbar}^\star$ converges to the normalised factors $(\hbar^\intercal\hbar)^{-1/2}\hbar^\intercal$ up to an asymptotically orthogonal matrix. Unlike Aït-Sahalia and Xiu (2017), in which a uniform convergence rate of the factor estimator is given for both the max norm and the spectral norm, we provide separate convergence rates under the two norms. When $\bar{\tau}_G = \underline{\tau}_G =$

$0 \geq \bar{\tau}_V$, the spectral-norm convergence rate is $(\log d)^{1/2} n^{-1/2} + m_{w,nd} d^{-1/2}$, which is the same as that in Theorem 5 of Aït-Sahalia and Xiu (2017). This means that with the introduction of a factor structure for microstructure noise which is of the same magnitude as the factor structure for efficient returns, there is no loss in the convergence rates of factor estimators. Furthermore, when $\bar{\tau}_G = \underline{\tau}_G = 0 \geq \bar{\tau}_V$, the uniform convergence rate in (1.3.12) is close to the element-wise convergence rate, $n^{-1} + d^{-1}$, of the first-differenced factor estimator in high-frequency factor models, provided by Lemma L.4 in the Appendix of Pelger (2019).

The introduction of the magnitude matrices provides insight into how a larger noise-to-signal ratio can worsen the estimation. We can see from Theorem 1.3.1 (and Theorems 1.3.3 and 1.3.5 below) that the convergence rates of the estimators are affected by the magnitudes of both $\mathbf{D}_G$ and $\mathbf{D}_V$. We exemplify this by looking at the uniform convergence rates of the estimators of common components in (1.3.12) under the following three cases with $\gamma_2 = \gamma_3 \geq 1$ and $m_{U,d} = m_{v,d} = O(1)$.

- (i) When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = 0$ (i.e., $\bar{\tau}_G^\diamond = \underline{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.5$), the efficient returns, the microstructure noise common component, and the microstructure idiosyncratic error are of the same magnitude. In this case, the uniform convergence rate in (1.3.12) is $\log(nd) \cdot n^{-1} + (\log n) \cdot (nd)^{-1/2}$.

- (ii) When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = -0.1$ (i.e., $\bar{\tau}_G^\diamond = \underline{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.6$), the microstructure noise common component and idiosyncratic error are of a smaller magnitude than efficient returns, and the convergence rate in (1.3.12) becomes $\log(nd) \cdot n^{-0.8} + (\log n) \cdot (nd)^{-1/2}$, which is slower than the convergence rate in case (i). This may be because the signals of the microstructure noise common component now become weaker than the efficient return idiosyncratic errors, making it more difficult to estimate.

- (iii) When $\bar{\tau}_G = \underline{\tau}_G = \bar{\tau}_V = 0.1$ (i.e., $\bar{\tau}_G^\diamond = \underline{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.4$), the microstructure noise common component and idiosyncratic error are of a larger magnitude than efficient returns, and the convergence rate becomes even slower at $(\log(nd)) \cdot (n^{-0.7} + n^{-0.2} d^{-1}) + n^{0.1} (\log n) \cdot (nd)^{-1/2}$.

We will further compare the performance of our proposed method for cases (ii) and (iii) in the simulation study.

1.3.2 Second step: PCA estimation in cumulative form

Denote $\boldsymbol{\beta} = (\mathbf{O}_{K_G \times K_F},\ \mathbf{I}_{K_G})^\intercal$ and $\boldsymbol{\beta}_\perp = (\mathbf{I}_{K_F},\ \mathbf{O}_{K_F \times K_G})^\intercal$, where $\mathbf{O}_{K_1 \times K_2}$ denotes a $K_1 \times K_2$ matrix of zeros. Then $\mathscr{g} = n^{1/2} \hbar \boldsymbol{\beta}$ is the matrix of true factors for the first difference of microstructure noise, and $\mathscr{f} = \hbar \boldsymbol{\beta}_\perp$ is the matrix of true factors for efficient returns. However, by Theorem 3.1, the estimated factors, $\widehat{\hbar}^\star$, from the first step are consistent only up to a rotation. In other words, each

column of $\widehat{\mathscr{H}}^\star$ may converge to a linear combination of both factors for efficient returns and factors for microstructure noise. Hence, instead of applying $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ directly on $\widehat{\mathscr{H}}^\star$ to obtain estimates of $\mathscr{g}$ and $\mathscr{f}$, we need to find proper rotations of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$ that achieve this goal. This will be the aim of the second step.

There are different ways to estimate such rotations of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$. For example, Barigozzi *et al.* (2021) use Johansen (1995)'s reduced rank estimation in a dynamic factor model to estimate the cointegration coefficients of non-stationary factors (also see Section 1.4.2). However, their method requires the specification of a finite-order vector autoregression (or a vector error correction model) prior to estimation, which is not reasonable in our high-frequency setting. To avoid the (mis)specification issue, we will use a second-step PCA on cumulated factors to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$. This is a high-frequency analogue of Bai and Ng (2004)'s PANIC and is in the same spirit as the methods of Stock and Watson (1988); Harris (1997); Peña and Poncela (2006) and Zhang *et al.* (2019) for identifying nonstationary factors or cointegration by eigenanalysis in the low-frequency setting. The intuition for why the second-step PCA can separate the factors for efficient prices and the factors for microstructure noise is as follows. In the first step we estimate $\boldsymbol{h}_t = (\boldsymbol{f}_t^\intercal, n^{-1/2}\boldsymbol{g}_t^\intercal)^\intercal$, whose components are of the same magnitude. Upon cumulation of $\boldsymbol{h}_t$, we obtain

$$\boldsymbol{H}_{s\Delta} = \sum_{s_1=1}^{s} \boldsymbol{h}_{s_1\Delta} = \left(\left(\int_0^{s\Delta} \boldsymbol{\sigma}_{ft} \mathrm{d}\boldsymbol{B}_t^F \right)^\intercal,\ n^{-1/2}(\boldsymbol{G}_{s\Delta} - \boldsymbol{G}_0)^\intercal \right)^\intercal, \quad 1 < s \leq n.$$

As $\{\boldsymbol{G}_t\}$ is assumed to be a stationary time series, the second component of $\boldsymbol{H}_{s\Delta}$ is dominated by the first in magnitude, which would enable their separation. However, as the factors from the first-step PCA are consistent only up to a rotation, a second PCA on the cumulated factors is needed. The leading eigenvalues in the second PCA will correspond to the efficient price factors, whereas the remaining eigenvalues will correspond to the microstructure noise factors. This separates out the efficient price factors and the microstructure noise factors.

For $1 \leq s \leq n$, let $\widehat{\boldsymbol{H}}_{s\Delta}^\star = \sum_{s_1=1}^{s} \widehat{\boldsymbol{h}}_{s_1\Delta}^\star$ be an estimator of $\boldsymbol{H}_{s\Delta}$. Define the demeaned $\widehat{\boldsymbol{H}}_{s\Delta}^\star$ as $\widehat{\boldsymbol{H}}_{s\Delta}^{\star c} = \widehat{\boldsymbol{H}}_{s\Delta}^\star - \overline{\widehat{\boldsymbol{H}}^\star}$, where $\overline{\widehat{\boldsymbol{H}}^\star} = n^{-1}\sum_{s=1}^{n} \widehat{\boldsymbol{H}}_{s\Delta}^\star$. In the matrix form, this can be written as $\widehat{\mathscr{H}}^{\star c} = \widehat{\mathscr{H}}^\star - \overline{\widehat{\mathscr{H}}^\star}$, with $\widehat{\mathscr{H}}^{\star c} = (\widehat{\boldsymbol{H}}_\Delta^{\star c}, \ldots, \widehat{\boldsymbol{H}}_{n\Delta}^{\star c})^\intercal$, $\widehat{\mathscr{H}}^\star = (\widehat{\boldsymbol{H}}_\Delta^\star, \ldots, \widehat{\boldsymbol{H}}_{n\Delta}^\star)^\intercal$, and $\overline{\widehat{\mathscr{H}}^\star} = \boldsymbol{1}_n \overline{\widehat{\boldsymbol{H}}^\star}^\intercal$. Define the $K_H \times K_H$ matrix

$$\widehat{\mathbf{S}}_{HH} = n^{-1} \widehat{\mathscr{H}}^{\star c\intercal} \widehat{\mathscr{H}}^{\star c}.$$

Let $\widehat{\boldsymbol{\beta}}_\perp$ be the matrix of eigenvectors associated with the largest K_F eigenvalues of $\widehat{\mathbf{S}}_{HH}$ and let $\widehat{\boldsymbol{\beta}}$ be the matrix of eigenvectors associated with the rest of the K_G eigenvalues. Then, $\widehat{\boldsymbol{\beta}}_\perp^\intercal \widehat{\boldsymbol{h}}_t^\star$ is an

estimator of $\boldsymbol{\beta}_\perp^\intercal \boldsymbol{h}_t = \boldsymbol{f}_t$, and $\widehat{\boldsymbol{\beta}}^\intercal \widehat{\boldsymbol{h}}_t^\star$ is an estimator of $\boldsymbol{\beta}^\intercal \boldsymbol{h}_t = n^{-1/2}\boldsymbol{g}_t$.

Lemma 1.3.2. *Under the assumptions of Theorem 1.3.1, $\widehat{\boldsymbol{\beta}}_\perp$ and $\widehat{\boldsymbol{\beta}}$ are consistent in the sense that:*

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\Xi}^\intercal \boldsymbol{\beta}\mathbf{Q}_{\boldsymbol{\beta}} = O_P((\log n)n^{-\bar{\tau}_G} \cdot b_{nd}) \tag{1.3.14}$$

and

$$\widehat{\boldsymbol{\beta}}_\perp - \boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp \mathbf{Q}_{\boldsymbol{\beta}_\perp} = O_P((\log n)n^{-\bar{\tau}_G} \cdot b_{nd}), \tag{1.3.15}$$

where

$$\mathbf{Q}_{\boldsymbol{\beta}} = [\boldsymbol{\beta}^\intercal \boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal \boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal \boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}, \quad \mathbf{Q}_{\boldsymbol{\beta}_\perp} = [\boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal (\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp, \text{ and}$$

$$\boldsymbol{\Xi} = d^{1/2}\boldsymbol{\Lambda}_H^\intercal \widehat{\boldsymbol{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{-1/2} = (\hbar^\intercal \hbar)^{1/2}\mathbf{R}^\star.$$

The lemma shows that $\widehat{\boldsymbol{\beta}}$ estimates a basis for the space spanned by $\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}$. Using Lemma 1.3.2 and Theorem 1.3.1, we can prove the following theorem, which gives the convergence rate of the estimators of the microstructure noise factors, the efficient price factors and their factor loadings. To this end, we define

$$\widehat{\ell}^\star = \widehat{\hbar}^\star\widehat{\boldsymbol{\beta}}_\perp, \quad \widehat{g}^\star = \widehat{\hbar}^\star\widehat{\boldsymbol{\beta}}, \quad \widehat{\boldsymbol{\Lambda}}_F^\star = \widehat{\boldsymbol{\Lambda}}_H^\star\widehat{\boldsymbol{\beta}}_\perp, \quad \widehat{\boldsymbol{\Lambda}}_G^\star = \widehat{\boldsymbol{\Lambda}}_H^\star\widehat{\boldsymbol{\beta}}, \quad \widehat{\mathscr{F}}^\star = \widehat{\mathscr{H}}^\star\widehat{\boldsymbol{\beta}}_\perp, \quad \text{and} \quad \widehat{\mathscr{G}}^\star = \widehat{\mathscr{H}}^\star\widehat{\boldsymbol{\beta}}.$$

Theorem 1.3.3. *Suppose that Assumptions 1.A–1.G are satisfied. We have*

(i)

$$\left\|\widehat{\ell}^\star - \ell\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp\right\|_{\mathsf{O}} = O_P\left((\log n)n^{-\bar{\tau}_G} \cdot \widetilde{b}_{nd}\right) \tag{1.3.16}$$

and

$$\left\|\widehat{\ell}^\star - \ell\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp\right\|_{\max} = O_P\left((\log n)^{2/(\gamma_2\wedge 1)}n^{-1/2+\bar{\tau}_V^+-\bar{\tau}_G^-} \cdot b_{nd}\right); \tag{1.3.17}$$

(ii)

$$\left\|\widehat{g}^\star - n^{-1/2}g\mathbf{Q}_{\boldsymbol{\beta}}\right\|_{\mathsf{O}} = O_P\left((\log n)n^{-\bar{\tau}_G} \cdot \widetilde{b}_{nd}\right) \tag{1.3.18}$$

and

$$\left\|\widehat{\mathscr{g}}^{\star} - n^{-1/2}\mathscr{g}\mathbf{Q}_{\boldsymbol{\beta}}\right\|_{\max} = O_P\left((\log n)^{2/(\gamma_2\wedge 1)} n^{-1/2+\bar{\tau}_V^+-\underline{\tau}_G^-}\cdot b_{nd}\right); \tag{1.3.19}$$

(iii)

$$\left\|\widehat{\boldsymbol{\Lambda}}_F^{\star} - \boldsymbol{\Lambda}_F\mathbf{Q}_{\boldsymbol{\beta}_\perp}\right\|_{\max} = O_P\left(n^{-2\underline{\tau}_G^-}\cdot a_{nd}\right); \tag{1.3.20}$$

(iv)

$$\left\|\widehat{\boldsymbol{\Lambda}}_G^{\star} - n^{1/2}\boldsymbol{\Lambda}_G\mathbf{D}_G\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}\right\|_{\max} = O_P\left(n^{-2\underline{\tau}_G^-}\cdot a_{nd}\right), \tag{1.3.21}$$

where $\mathbf{Q}_{\boldsymbol{\beta}}$, $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ *and* $\boldsymbol{\Xi}$ *are defined in Lemma 1.3.2.*

This theorem shows that $\widehat{\mathscr{f}}^{\star}$ and $\widehat{\mathscr{g}}^{\star}$ are estimators of the factors for the first-differenced efficient prices and the microstructure noise with rotations $\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\widehat{\boldsymbol{\beta}}_\perp$ and $\mathbf{Q}_{\boldsymbol{\beta}}$, respectively, and that $\widehat{\boldsymbol{\Lambda}}_F^{\star}$ and $\widehat{\boldsymbol{\Lambda}}_G^{\star}$ are estimators of the factor loadings for the efficient prices and the microstructure noise with rotations $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ and $\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}$, respectively.

When we estimate the first-differenced common components of the microstructure noise and the efficient prices (i.e., $\mathscr{g}\boldsymbol{\Lambda}_G^{\intercal}$ and $\mathscr{f}\boldsymbol{\Lambda}_F^{\intercal}$), the rotation matrices cancel out. Thus, we have the following corollary.

Corollary 1.3.4. *Suppose that Assumptions 1.A–1.G are satisfied. We have*

(i)

$$\left\|\widehat{\mathscr{f}}^{\star}\widehat{\boldsymbol{\Lambda}}_F^{\star\intercal} - \mathscr{f}\boldsymbol{\Lambda}_F^{\intercal}\right\|_{\max} = O_P\left((\log n)^{2/(\gamma_2\wedge 1)} n^{-1/2+\bar{\tau}_V^+-\underline{\tau}_G^-}\cdot b_{nd}\right); \tag{1.3.22}$$

(ii)

$$\left\|\widehat{\mathscr{g}}^{\star}\widehat{\boldsymbol{\Lambda}}_G^{\star\intercal} - \mathscr{g}\mathbf{D}_G\boldsymbol{\Lambda}_G^{\intercal}\right\|_{\max} = O_P\left((\log n)^{2/(\gamma_2\wedge 1)} n^{-1/2+\bar{\tau}_V^++\bar{\tau}_G-\underline{\tau}_G^-}\cdot b_{nd}\right). \tag{1.3.23}$$

In the following theorem, we provide the uniform convergence rates for the cumulated factors and common components.

Theorem 1.3.5. *Suppose that Assumptions 1.A–1.G are satisfied. Then,*

(i) for factors, we have

$$\left\|\widehat{\mathscr{H}}^{\star} - (\mathscr{H} - \mathbf{1}_n\boldsymbol{H}_0^{\intercal})(\boldsymbol{\Xi}^{\intercal})^{-1}\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G^-}\cdot b_{nd}\right), \tag{1.3.24}$$

$$\left\|\widehat{\mathscr{G}}^{\star}-n^{-1/2}(\mathscr{G}-\mathbf{1}_n\boldsymbol{G}_0^{\intercal})\mathbf{Q}_{\boldsymbol{\beta}}\right\|_{\max}=O_P\left((\log n)^2n^{-\underline{\tau}_G^-}\cdot b_{nd}\right), \tag{1.3.25}$$

and

$$\left\|\widehat{\mathscr{F}}^{\star}-(\mathscr{F}-\mathbf{1}_n\boldsymbol{F}_0^{\intercal})\boldsymbol{\beta}_{\perp}^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\widehat{\boldsymbol{\beta}}_{\perp}\right\|_{\max}=O_P\left((\log n)^2n^{-\underline{\tau}_G^-}\cdot b_{nd}\right); \tag{1.3.26}$$

(ii) for common components, we have

$$\left\|\widehat{\mathscr{H}^{\star}}\widehat{\boldsymbol{\Lambda}}_H^{\star\intercal}-(\mathscr{H}-\mathbf{1}_n\boldsymbol{H}_0^{\intercal})\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\right\|_{\max}=O_P\left(n^{-\underline{\tau}_G^-+\bar{\tau}_G^+}\cdot b_{nd}\right), \tag{1.3.27}$$

$$\left\|\widehat{\mathscr{G}}^{\star}\widehat{\boldsymbol{\Lambda}}_G^{\star\intercal}-(\mathscr{G}-\mathbf{1}_n\boldsymbol{G}_0^{\intercal})\mathbf{D}_G\boldsymbol{\Lambda}_G^{\intercal}\right\|_{\max}=O_P\left((\log n)^2n^{-\underline{\tau}_G^-+\bar{\tau}_G}\cdot b_{nd}\right), \tag{1.3.28}$$

and

$$\left\|\widehat{\mathscr{F}}^{\star}\widehat{\boldsymbol{\Lambda}}_F^{\star\intercal}-(\mathscr{F}-\mathbf{1}_n\boldsymbol{F}_0^{\intercal})\boldsymbol{\Lambda}_F^{\intercal}\right\|_{\max}=O_P\left((\log n)^2n^{-\underline{\tau}_G^-}\cdot b_{nd}\right), \tag{1.3.29}$$

where $\mathbf{Q}_{\boldsymbol{\beta}}$, $\mathbf{Q}_{\boldsymbol{\beta}_\perp}$ *and* $\boldsymbol{\Xi}$ *are defined in Lemma 1.3.2.*

When $\bar{\tau}_G=\underline{\tau}_G=0\geq\bar{\tau}_V$, $\gamma_2=\gamma_3\geq 1$, and $m_{w,nd}=O(1)$, the uniform convergence rate of $\widehat{\mathscr{H}^{\star}}$ in Theorem 1.3.5(i) becomes $(\log(nd))(n^{-1/2}+d^{-1/2})$. In comparison, Pelger (2019) obtains an element-wise convergence rate of $n^{-1/2}+d^{-1/2}$ for the cumulated factor estimator (see Lemma L.3 in the Appendix of Pelger (2019)). On the other hand, in a low-frequency factor model setting, Bai and Ng (2004) obtain a uniform convergence rate of $n^{-3/4}+d^{-1/2}$ (see Lemma 2 in Bai and Ng (2004)). Our estimators have a slower convergence rate, which is mainly due to the different settings.

Although in this chapter we do not consider estimation of the integrated covolatility, $\boldsymbol{\Sigma}_{x^*}=\boldsymbol{\Lambda}_F\boldsymbol{\Sigma}_F\boldsymbol{\Lambda}_F^{\intercal}+\boldsymbol{\Sigma}_U$, we can use the estimated common component of efficient price to estimate $\boldsymbol{\Lambda}_F\boldsymbol{\Sigma}_F\boldsymbol{\Lambda}_F^{\intercal}$ without the need to use noise-robust estimation methods (such as pre-averaging). This may lead to efficiency gain. However, for the estimation of $\boldsymbol{\Sigma}_U$, since we do not separate the idiosyncratic errors for the efficient price and the microstructure noise, methods like pre-averaging may still be needed.

1.4 Monte-Carlo simulation

1.4.1 Number of factors

In this chapter, we apply the commonly-used information criterion proposed by Bai and Ng (2002) to estimate the total number of factors. Then, we use the PANIC test procedure proposed by Bai and Ng (2004) to determine the number of factors for the efficient prices.

To introduce Bai and Ng (2002)'s information criterion, we denote by $\overline{Q}$ a finite positive integer that is no smaller than K_H. For any $1 \leq q_H \leq \overline{Q}$, we let $\widehat{\mathscr{h}}^{\star}(q_H) = \left(\widehat{\boldsymbol{h}}^{\star}_{\Delta}(q_H), \ldots, \widehat{\boldsymbol{h}}^{\star}_{n\Delta}(q_H)\right)^{\intercal}$ be the matrix of estimated factors when the total number of factors is assumed to be q_H, and denote by $\widehat{\boldsymbol{\Lambda}}^{\star}_H(q_H)$ the corresponding loadings matrix. The information criterion is defined as

$$\mathsf{IC}_1(q_H) = \log\left[\mathsf{V}_n(q_H)\right] + q_H \cdot \frac{n+d}{nd} \cdot \log(\frac{nd}{n+d}), \tag{1.4.1}$$

where $\mathsf{V}_n(q_H) = \|x - \widehat{\mathscr{h}}^{\star}(q_H)(\widehat{\boldsymbol{\Lambda}}^{\star}_H(q_H))^{\intercal}\|^2_F$. The total number of factors is then estimated as

$$\widehat{K}_H = \underset{0 \leq q_H \leq \overline{Q}}{\arg\min}\, \mathsf{IC}_1(q_H), \tag{1.4.2}$$

with $\mathsf{IC}_1(0) = \|x\|^2_F$ for convention. For consistency of $\widehat{K}_H$, we refer to the asymptotic results given in Theorem 2 of Bai and Ng (2002).

Bai and Ng (2004) propose two tests to determine the number of nonstationary factors. We adopt the one that does not specify a finite order VAR representation. For any $1 \leq q_F \leq \widehat{K}_H$, we let $\widehat{\mathscr{F}}^{\star}(q_F) = \widehat{\mathscr{H}}^{\star}\widehat{\boldsymbol{\beta}}_{\perp}(q_F)$, when the number of factors for efficient prices is assumed to be q_F. Let $\widehat{\xi}^F_t(q_F)$ be the residuals from estimating a first-order VAR of $\widehat{\mathscr{F}}^{\star}(q_F)$, $\widehat{\boldsymbol{\Sigma}}_{S,\xi}(q_F)$ be the sample covariance matrix of $\widehat{\xi}^F_t$, and $\widehat{\boldsymbol{\Sigma}}_{L,\xi}(q_F)$ be the estimated long-run covariance matrix of $\widehat{\xi}^F_t$. The test statistic for $\mathsf{H}_0 : K_F = q_F$ is defined by

$$MQ(q_F) = n(\nu(q_F) - 1), \tag{1.4.3}$$

where $\nu(q_F)$ is the smallest eigenvalue of

$$\Big[\sum_{s=2}^{n} \frac{1}{2}\left(\widehat{\boldsymbol{F}}^{\star}_{s\Delta}(q_F)\widehat{\boldsymbol{F}}^{\star\intercal}_{(s-1)\Delta}(q_F) + \widehat{\boldsymbol{F}}^{\star}_{(s-1)\Delta}(q_F)\widehat{\boldsymbol{F}}^{\star\intercal}_{s\Delta}(q_F)\right) - n(\widehat{\boldsymbol{\Sigma}}_{L,\xi}(q_F) - \widehat{\boldsymbol{\Sigma}}_{S,\xi}(q_F))\Big] \cdot \Big(\sum_{s=2}^{n} \widehat{\boldsymbol{F}}^{\star}_{s\Delta}(q_F)\widehat{\boldsymbol{F}}^{\star\intercal}_{s\Delta}(q_F)\Big)^{-1}. \tag{1.4.4}$$

For a given significance level α, define

$$\widehat{K}_F = \max_{\substack{0\leq q_F\leq \widehat{K}_H,\\ MQ(q_F)<c_{\alpha,MQ}(q_F)}} q_F, \tag{1.4.5}$$

where $c_{\alpha,MQ}(q_F)$ is the critical value of the test statistic $MQ(q_F)$ at α significance level. For convention, we define $c_{\alpha,MQ}(0) = +\infty$. Following Bai and Ng (2004), we use simulation to calculate the critical values. Specifically, we use a $q_F \times 1$ vector of demeaned standard Brownian motions to replace $\widehat{\mathbf{F}}^{\star}_{s\Delta}(q_F)$ and a q_F-dimensional identity matrix to replace $\widehat{\mathbf{\Sigma}}_{L,F}(q_F)$ in (1.4.4), and then use (1.4.3) to obtain a value of the MQ statistic. With 100,000 replications, we calculate the critical values at 1%, 5%, and 10% significance levels. We refer the reader to Theorem 1 of Bai and Ng (2004) for the asymptotic distribution of the test statistic.

1.4.2 Alternative approaches for comparison

We consider two alternative approaches for comparison. The first approach, denoted as DPCA, estimates the factors as $\widehat{\hbar}$ in (1.3.2) instead of $\widehat{\hbar}^{\star}$, and uses $\widehat{\hbar}$ for the second-step PCA while keeping the rest of the steps exactly the same. The second approach is proposed by Barigozzi *et al.* (2020) and Barigozzi *et al.* (2021) and is denoted as PCA*-VECM. The PCA*-VECM uses $\widehat{\boldsymbol{h}}^{\star}_t$ from the first-step PCA* to construct a Vector Error Correction Model (VECM) in order to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\perp}$, while in our method, we use a second-step PCA to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\perp}$. We set the lag of VECM to 1 for simplicity, and for a lag length larger than 1, we refer the reader to Chapter 6 of Johansen (1995). Specifically, the PCA*-VECM uses Johansen (1995)'s reduced rank regression method to estimate a VECM for $\widehat{\boldsymbol{h}}^{\star}_t$:

Step 1: Implement OLS of $\widehat{\boldsymbol{h}}^{\star}_t$ and $\widehat{\boldsymbol{H}}^{\star}_{t-\Delta}$ on $\widehat{\boldsymbol{h}}^{\star}_{t-\Delta}$ to get residuals $\widehat{\boldsymbol{e}}_{0,t}$ and $\widehat{\boldsymbol{e}}_{1,t}$, respectively.

Step 2: Let $\widehat{\mathbf{S}}_{ij} = n^{-1}\sum_{s=1}^{n}\widehat{\boldsymbol{e}}_{i,s\Delta}\widehat{\boldsymbol{e}}^{\intercal}_{j,s\Delta}$ for $i,j = 0,1$. Then let $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1,\ldots,\widehat{\boldsymbol{\beta}}_{K_G})$, where $\widehat{\boldsymbol{\beta}}_l$, $l = 1,\ldots,K_G$, is the eigenvector belonging to the l-th largest eigenvalue of the matrix $(\widehat{\mathbf{S}}_{11} - \widehat{\mathbf{S}}_{10}\widehat{\mathbf{S}}^{-1}_{00}\widehat{\mathbf{S}}_{01})$. Define $\widehat{\boldsymbol{\beta}}_{\perp}$ as the orthogonal complement matrix of $\widehat{\boldsymbol{\beta}}$ such that $\widehat{\boldsymbol{\beta}}^{\intercal}_{\perp}\widehat{\boldsymbol{\beta}} = \mathbf{O}_{K_F\times K_G}$ and $\widehat{\boldsymbol{\beta}}^{\intercal}_{\perp}\widehat{\boldsymbol{\beta}}_{\perp} = \mathbf{I}_{K_F}$.

Note that even if the estimated factors are not normalised in Step 1, the residuals $\widehat{\boldsymbol{e}}_{0,t}$ and $\widehat{\boldsymbol{e}}_{1,t}$ will not change and therefore the estimates of $\widehat{\boldsymbol{\beta}}_{\perp}$ and $\widehat{\boldsymbol{\beta}}$ will not be affected. Also note that the factors for the efficient prices follow a diffusion model and hence, are heteroskedastic. One can implement more efficient estimation of VECM under heteroskedasticity (e.g., generalised least squares estimation in Seo (2007) and Herwartz and Lütkepohl (2011)). But we do not pursue this in our chapter.

1.4.3 Data generating processes

The data generating process in our simulation is similar to that in Bollerslev *et al.* (2019). The observable prices are the sum of the efficient prices and microstructure noise. The former has two orthogonal factors and the latter has one factor.

The two factors in the efficient prices are independently generated from a GARCH diffusion model as in Andersen and Bollerslev (1998),

$$\begin{cases} \mathsf{d}F_{it} = \sigma_{f,it}\mathsf{d}B_{it}^F, \\ \mathsf{d}\sigma_{f,it}^2 = \kappa_{fi}(\theta_{fi} - \sigma_{f,it}^2)\mathsf{d}t + \iota_{fi}\sigma_{f,it}^2\mathsf{d}W_{it}^F, \end{cases} \tag{1.4.6}$$

for $i = 1, 2$, where B_{it}^F and W_{it}^F are dependent Brownian motions with $\mathsf{corr}(B_{it}^F, W_{it}^F) = -0.5$. The parameters are set as $\kappa_{f1} = \kappa_{f2} = 0.035$, $\theta_{f1} = 0.636$, $\theta_{f2} = 0.3$, $\iota_{fi} = \sqrt{2\kappa_{fi}\phi_{fi}}$, $\phi_{f1} = \phi_{f2} = 0.296$, and initial value $(f_{i0}, \sigma_{f,i0}^2) = (0, \theta_{fi})$. Then we draw the factor loadings of the efficient prices independently from a normal distribution with mean zero and unit variance.

The idiosyncratic components of the efficient prices are generated as $\mathsf{d}U_{it} = \sigma_{u,it}\mathsf{d}\mathsf{B}_{it}^U$, where B_{it}^U is a Brownian motion, and $\sigma_{u,it}$ is generated by three different volatility processes for different stocks to allow for heterogeneity.

- For $1 \leq i \leq \lfloor d/3 \rfloor$, the volatility process is generated by an exponential ARCH diffusion limit model as in Nelson (1990):

$$\mathsf{d}\log(\sigma_{u,it}^2) = -0.6(0.157 - \log(\sigma_{u,it}^2))\mathsf{d}t + 0.25\mathsf{d}W_{it}^U$$

  with initial value $\log(\sigma_{u,i0}^2) = 0.157$, where B_{it}^U is a Brownian motion with $\mathsf{corr}(B_{it}^U, W_{it}^U) = -0.3$.

- For $\lfloor d/3 \rfloor + 1 \leq i \leq \lfloor 2d/3 \rfloor$, the volatility process is generated by a GARCH-M diffusion limit model as in Nelson (1990),

$$\mathsf{d}(\sigma_{u,it}^2) = (0.1 - \sigma_{u,it}^2)\mathsf{d}t + 0.2\sigma_{u,it}^2\mathsf{d}W_{it}^U$$

  with initial value $\sigma_{u,i0}^2 = 0.1$, where B_{it}^U is a Brownian motion with $\mathsf{corr}(B_{it}^U, W_{it}^U) = -0.3$.

- For $\lfloor 2d/3 \rfloor + 1 \leq i \leq d$, the volatility process is generated by a GARCH diffusion model as in

Andersen and Bollerslev (1998),

$$\mathsf{d}(\sigma^2_{u,it}) = 0.035(0.636 - \sigma^2_{u,it})\mathsf{d}t + 0.2\sigma^2_{u,it}\mathsf{d}W^U_{it},$$

with initial value $\sigma^2_{u,i0} = 0.636$, where B^U_{it} is a Brownian motion with $\mathsf{corr}(B^U_{it}, W^U_{it}) = -0.3$.

The two dimensional Brownian motion (B^U_{it}, W^U_{it}) is independent over $1 \leq i \leq d$ and also independent with the driving Brownian motions (B^F_{1t}, W^F_{1t}) and (B^F_{2t}, W^F_{2t}) for the efficient price factors.

As for microstructure noise, we draw the factor loadings, λ_{Gi}, $i = 1, ..., d$, independently from a normal distribution with mean one and unit variance. Then we introduce the noise-to-signal ξ^2_G and ξ^2_V as in Bollerslev *et al.* (2019), which take values $n^{2\bar{\tau}^\diamond_G}$ and $n^{2\bar{\tau}^\diamond_V}$, respectively, with $\bar{\tau}^\diamond_G \in \{-0.4, -0.6\}$ and $\bar{\tau}^\diamond_V \in \{-0.4, -0.6\}$. The variance of the factor for the microstructure noise satisfies $\mathsf{Var}(G_t) = 0.5\xi^2_G(\frac{1}{nd}\sum_{i=1}^d \sum_{t=1}^n \sigma^4_{*,it})^{1/2}/\bar{c}$, and is thus time-invariant, where $\sigma_{*,it}$ is the spot volatility of the efficient price process of asset i at time t, that is, $\sigma^2_{*,it} = (\lambda_{F,i1}\sigma_{f,1t})^2 + (\lambda_{F,i2}\sigma_{f,2t})^2 + \sigma^2_{u,it}$, and $\bar{c} = \frac{1}{d}\sum_{i=1}^d \lambda^2_{Gi}$. The variance of idiosyncratic component V_{it} makes up $0.1\xi^2_V$ of the total variance, that is $\mathsf{Var}(V_{it}) = 0.1\xi^2_V(\frac{1}{n}\sum_{t=1}^n \sigma^4_{*,it})^{1/2}$. We draw the factor G_t independently from a normal distribution with mean zero and variance $\mathsf{Var}(G_t)$, and draw V_{it} independently from a normal distribution with mean zero and variance $\mathsf{Var}(V_{it})$.

We assume that the prices are synchronously recorded once every one or five minutes during 6.5 trading hours, that is $n = 390$ or 78. The number of assets is assumed to be $d = 50, 100, 300$ and 500. We present the simulation results based on 1000 Monte Carlo replications.

1.4.4 Simulation results

Firstly, we provide simulation results for the estimation of number of factors by the information criteria described in Section 1.4.1, as well as by the PANIC test with different significance levels, 1%, 5%, and 10%. More specifically, we use Bai and Ng (2002)'s information criterion, IC_1, to estimate the total number of factors and then use the PANIC test to identify the number of efficient price factors. For the PANIC test in (1.4.5), $\widehat{K}_H$ is determined by IC_1.

Table 1.2 presents the average number of factors determined by each method (over 1000 replications) for $n = 78$ and $n = 390$, respectively. It can be seen that IC_1 has excellent performance in estimating the total number of factors in all scenarios. The PANIC tests using the MQ statistic at different significance levels are denoted as $MQ_{1\%}$, $MQ_{5\%}$, and $MQ_{10\%}$. They are used to determine the number of efficient price factors and perform satisfactorily in all scenarios, in particular

Table 1.2: Average estimated number of factors and standard deviation (in parentheses) for sampling frequency=5 mins (n=78) or 1 min (n=390), true number of factors=2 (efficient prices)+1 (microstructure noise)

	IC_1	$MQ_{1\%}$	$MQ_{5\%}$	$MQ_{10\%}$		IC_1	$MQ_{1\%}$	$MQ_{5\%}$	$MQ_{10\%}$
$n=78$, $\bar{\tau}_G^\diamond=-0.4$, $\bar{\tau}_V^\diamond=-0.4$					$n=390$, $\bar{\tau}_G^\diamond=-0.4$, $\bar{\tau}_V^\diamond=-0.4$				
d=50	3.026	1.998	1.931	1.855	d=50	3.035	1.989	1.901	1.815
	(0.025)	(0.030)	(0.086)	(0.150)		(0.036)	(0.027)	(0.101)	(0.177)
d=100	3.001	1.983	1.931	1.858	d=100	3.004	1.987	1.917	1.827
	(0.001)	(0.017)	(0.072)	(0.134)		(0.004)	(0.015)	(0.080)	(0.165)
d=300	3.000	1.986	1.942	1.871	d=300	3.000	1.987	1.920	1.850
	(0.000)	(0.014)	(0.059)	(0.122)		(0.000)	(0.013)	(0.080)	(0.144)
d=500	3.000	1.988	1.940	1.869	d=500	3.000	1.988	1.923	1.850
	(0.000)	(0.012)	(0.062)	(0.130)		(0.000)	(0.012)	(0.071)	(0.146)
$n=78$, $\bar{\tau}_G^\diamond=-0.6$, $\bar{\tau}_V^\diamond=-0.4$					$n=390$, $\bar{\tau}_G^\diamond=-0.6$, $\bar{\tau}_V^\diamond=-0.4$				
d=50	3.026	2.015	1.939	1.861	d=50	3.035	1.991	1.909	1.818
	(0.025)	(0.047)	(0.091)	(0.156)		(0.036)	(0.027)	(0.101)	(0.175)
d=100	3.001	1.984	1.930	1.858	d=100	3.004	1.990	1.920	1.827
	(0.001)	(0.018)	(0.073)	(0.134)		(0.004)	(0.010)	(0.078)	(0.165)
d=300	3.000	1.987	1.943	1.873	d=300	3.000	1.989	1.920	1.851
	(0.000)	(0.013)	(0.058)	(0.121)		(0.000)	(0.011)	(0.080)	(0.143)
d=500	3.000	1.988	1.939	1.869	d=500	3.000	1.988	1.923	1.850
	(0.000)	(0.012)	(0.063)	(0.130)		(0.000)	(0.012)	(0.071)	(0.146)
$n=78$, $\bar{\tau}_G^\diamond=-0.4$, $\bar{\tau}_V^\diamond=-0.6$					$n=390$, $\bar{\tau}_G^\diamond=-0.4$, $\bar{\tau}_V^\diamond=-0.6$				
d=50	3.001	1.991	1.949	1.884	d=50	3.000	1.993	1.931	1.847
	(0.001)	(0.011)	(0.056)	(0.117)		(0.000)	(0.007)	(0.072)	(0.154)
d=100	3.000	1.986	1.942	1.881	d=100	3.000	1.991	1.932	1.848
	(0.000)	(0.014)	(0.063)	(0.115)		(0.000)	(0.009)	(0.067)	(0.147)
d=300	3.000	1.991	1.951	1.890	d=300	3.000	1.989	1.926	1.863
	(0.000)	(0.009)	(0.051)	(0.108)		(0.000)	(0.011)	(0.073)	(0.132)
d=500	3.000	1.988	1.948	1.887	d=500	3.000	1.990	1.931	1.860
	(0.000)	(0.012)	(0.055)	(0.114)		(0.000)	(0.010)	(0.064)	(0.137)
$n=78$, $\bar{\tau}_G^\diamond=-0.6$, $\bar{\tau}_V^\diamond=-0.6$					$n=390$, $\bar{\tau}_G^\diamond=-0.6$, $\bar{\tau}_V^\diamond=-0.6$				
d=50	3.001	2.008	1.956	1.892	d=50	3.000	1.995	1.935	1.852
	(0.001)	(0.028)	(0.054)	(0.114)		(0.000)	(0.007)	(0.071)	(0.150)
d=100	3.000	1.987	1.945	1.881	d=100	3.000	1.991	1.935	1.849
	(0.000)	(0.015)	(0.060)	(0.115)		(0.000)	(0.009)	(0.065)	(0.146)
d=300	3.000	1.992	1.951	1.890	d=300	3.000	1.990	1.927	1.863
	(0.000)	(0.008)	(0.051)	(0.108)		(0.000)	(0.010)	(0.072)	(0.132)
d=500	3.000	1.988	1.949	1.887	d=500	3.000	1.989	1.931	1.861
	(0.000)	(0.012)	(0.054)	(0.114)		(0.000)	(0.011)	(0.064)	(0.136)

for $MQ_{1\%}$. In summary, IC_1 is very satisfactory in determining the number of total factors, and the PANIC test with 1% significance level is the most robust method to determine the number of efficient price factors.

Next, we compare the estimation of common components in the dual factor model. Our method is denoted as DPCA*. For the alternative methods discussed in Section 1.4.2, we denote them as DPCA and PCA*-VECM, respectively. We use the relative estimation error (REE) to measure the performance of different methods. It is defined as

$$\text{REE} = \|\mathbf{M} - \widehat{\mathbf{M}}\|/\|\mathbf{M}\|,$$

where $\|\cdot\|$ can be the Frobenius norm, the max norm or the spectral norm, and $\widehat{\mathbf{M}}$ is an estimate of the matrix $\mathbf{M}$, which varies from place to place (depending on the quantity being estimated). In Tables 1.3–1.6, "diff.total" refers to the total common component, $\mathscr{h}\mathbf{\Lambda}_H^\intercal$. "EP" and "diff.EP" refer to the common components of efficient prices and efficient returns. That is, EP $= \mathscr{F}\mathbf{\Lambda}_F^\intercal$ and diff.EP $= \mathscr{f}\mathbf{\Lambda}_F^\intercal$, respectively. Similarly, "MN" and "diff.MN" refer to the common components of the microstructure noise and its first difference, i.e., MN $= \mathscr{G}\mathbf{\Lambda}_G^\intercal$ and diff.MN $= \mathscr{g}\mathbf{\Lambda}_G^\intercal$, respectively. To focus on the comparison of different methods in estimating the common components, we assume that the true numbers of factors for the efficient price factors and the microstructure noise are known. Otherwise, one can obtain correct estimates of the numbers of factors by IC_1 and the PANIC test in most cases (as can be seen from Table 1.2).

In the following study, we will only consider the cases $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.6$ and $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.4$. To give a sense of the magnitudes of the different components in these two cases, we calculate the average volatilities (averaged over stocks) of the common and idiosyncratic components of efficient returns and the average sample variances of first differenced microstructure noise. We only provide the results for $d = 500$, as the results for other values of d are similar. When $n = 78$, the average volatilities of the common and idiosyncratic components for efficient returns are 0.01200 and 0.00684, respectively, while the average variances of the common and idiosyncratic components for differenced microstructure noise are 0.04523 and 0.00904 when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.4$ and decrease to 0.00788 and 0.00157 when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.6$. Similarly, when $n = 390$, the average volatilities of the common and idiosyncratic components for efficient returns are 0.00240 and 0.00137, respectively, while the average variances of the common and idiosyncratic components for differenced microstructure noise are 0.01250 and 0.00249 when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.4$ and decrease to 0.00115 and 0.00023 when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.6$. Hence, when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.4$, the common and idiosyncratic components of the differenced microstructure noise dominate those of the efficient returns, while the reverse is true when $\bar{\tau}_G^\diamond = \bar{\tau}_V^\diamond = -0.6$.

Table 1.3 gives the simulation results for $n = 78$, $\bar{\tau}_G^\diamond = -0.6$, and $\bar{\tau}_V^\diamond = -0.6$. In this case, the factors for the first differenced microstructure noise are smaller than the efficient return factors and idiosyncratic errors in magnitude and hence, are more difficult to detect and estimate. We can see that the REE's for "diff.MN" are generally larger than those for "diff.EP" for all the three methods. The estimates of the total common component, $\mathscr{h}\mathbf{\Lambda}_H^\intercal$, are the same for all the three methods, and the corresponding REE values are listed under "diff.total". Among the three estimation methods, it can be seen that DPCA* outperforms the rest of the methods. It can also be seen that the REE's for "MN" are always larger than those for "diff.MN". On the contrary, the REE's for "EP" are always smaller than those for "diff.EP". This may be due to the fact that the REE (calculated as

Table 1.3: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 78$, $\bar{\tau}_G^\diamond = -0.6$, $\bar{\tau}_V^\diamond = -0.6$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.209	0.283	0.284	0.261	0.690	0.339	0.359	0.266	0.731	0.306	0.319	0.278	1.006
	(0.021)	(0.059)	(0.092)	(0.111)	(0.383)	(0.106)	(0.134)	(0.116)	(0.371)	(0.080)	(0.121)	(0.113)	(0.381)
d=100	0.173	0.236	0.243	0.207	0.527	0.266	0.282	0.211	0.557	0.270	0.293	0.231	0.899
	(0.014)	(0.046)	(0.067)	(0.065)	(0.255)	(0.086)	(0.109)	(0.272)	(0.068)	(0.074)	(0.111)	(0.075)	(0.292)
d=300	0.143	0.199	0.209	0.166	0.389	0.202	0.208	0.168	0.403	0.245	0.273	0.200	0.859
	(0.010)	(0.041)	(0.061)	(0.038)	(0.173)	(0.075)	(0.091)	(0.042)	(0.216)	(0.084)	(0.117)	(0.066)	(0.354)
d=500	0.136	0.189	0.198	0.157	0.351	0.186	0.190	0.159	0.362	0.239	0.267	0.193	0.832
	(0.009)	(0.037)	(0.055)	(0.030)	(0.152)	(0.073)	(0.084)	(0.034)	(0.182)	(0.087)	(0.121)	(0.065)	(0.337)
REE under max norm													
d=50	0.210	0.271	0.284	0.253	0.583	0.338	0.363	0.259	0.620	0.293	0.317	0.266	0.854
	(0.049)	(0.082)	(0.113)	(0.100)	(0.276)	(0.142)	(0.154)	(0.103)	(0.276)	(0.103)	(0.141)	(0.104)	(0.356)
d=100	0.176	0.225	0.253	0.208	0.443	0.261	0.295	0.211	0.472	0.257	0.296	0.229	0.741
	(0.034)	(0.061)	(0.083)	(0.064)	(0.180)	(0.113)	(0.119)	(0.068)	(0.202)	(0.085)	(0.122)	(0.079)	(0.290)
d=300	0.158	0.196	0.240	0.176	0.332	0.200	0.235	0.177	0.338	0.236	0.290	0.203	0.684
	(0.029)	(0.054)	(0.078)	(0.043)	(0.119)	(0.081)	(0.099)	(0.046)	(0.162)	(0.088)	(0.122)	(0.070)	(0.325)
d=500	0.157	0.190	0.235	0.170	0.305	0.188	0.219	0.172	0.305	0.232	0.288	0.200	0.659
	(0.027)	(0.050)	(0.074)	(0.039)	(0.107)	(0.076)	(0.085)	(0.042)	(0.137)	(0.086)	(0.122)	(0.065)	(0.309)
REE under spectral norm													
d=50	0.161	0.235	0.230	0.257	0.660	0.315	0.302	0.261	0.677	0.259	0.258	0.274	0.983
	(0.022)	(0.067)	(0.079)	(0.126)	(0.389)	(0.131)	(0.122)	(0.130)	(0.375)	(0.087)	(0.099)	(0.129)	(0.379)
d=100	0.126	0.194	0.203	0.201	0.495	0.235	0.240	0.204	0.513	0.230	0.239	0.226	0.874
	(0.013)	(0.057)	(0.063)	(0.072)	(0.262)	(0.103)	(0.099)	(0.076)	(0.272)	(0.082)	(0.094)	(0.084)	(0.294)
d=300	0.115	0.166	0.188	0.161	0.356	0.167	0.184	0.163	0.368	0.215	0.228	0.197	0.834
	(0.010)	(0.053)	(0.060)	(0.041)	(0.176)	(0.082)	(0.078)	(0.045)	(0.212)	(0.088)	(0.095)	(0.074)	(0.353)
d=500	0.113	0.157	0.182	0.153	0.317	0.154	0.169	0.155	0.329	0.210	0.224	0.191	0.806
	(0.010)	(0.046)	(0.054)	(0.033)	(0.155)	(0.078)	(0.072)	(0.037)	(0.180)	(0.089)	(0.096)	(0.072)	(0.340)

Table 1.4: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 78$, $\bar{\tau}_G^\diamond = -0.4$, $\bar{\tau}_V^\diamond = -0.4$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.185	0.391	0.151	0.306	0.360	1.562	0.778	0.514	1.088	0.494	0.216	0.403	0.847
	(0.021)	(0.058)	(0.039)	(0.107)	(0.201)	(0.595)	(0.280)	(0.250)	(0.427)	(0.153)	(0.096)	(0.160)	(0.323)
d=100	0.151	0.322	0.131	0.257	0.295	1.478	0.742	0.476	1.049	0.434	0.199	0.364	0.803
	(0.014)	(0.050)	(0.031)	(0.069)	(0.140)	(0.652)	(0.313)	(0.235)	(0.484)	(0.144)	(0.085)	(0.155)	(0.282)
d=300	0.123	0.269	0.116	0.221	0.251	1.444	0.722	0.457	1.030	0.396	0.189	0.337	0.802
	(0.010)	(0.042)	(0.027)	(0.044)	(0.126)	(0.706)	(0.339)	(0.241)	(0.508)	(0.165)	(0.093)	(0.143)	(0.292)
d=500	0.117	0.256	0.112	0.214	0.246	1.435	0.717	0.450	1.042	0.387	0.186	0.333	0.789
	(0.010)	(0.043)	(0.026)	(0.041)	(0.134)	(0.723)	(0.346)	(0.241)	(0.524)	(0.179)	(0.098)	(0.158)	(0.322)
REE under max norm													
d=50	0.196	0.398	0.158	0.316	0.310	1.644	0.795	0.527	0.964	0.487	0.216	0.402	0.683
	(0.070)	(0.106)	(0.084)	(0.115)	(0.172)	(0.708)	(0.280)	(0.239)	(0.379)	(0.193)	(0.111)	(0.169)	(0.299)
d=100	0.157	0.322	0.139	0.273	0.247	1.548	0.759	0.489	0.929	0.422	0.196	0.367	0.636
	(0.041)	(0.080)	(0.043)	(0.080)	(0.102)	(0.753)	(0.303)	(0.231)	(0.431)	(0.167)	(0.087)	(0.159)	(0.267)
d=300	0.138	0.276	0.138	0.244	0.212	1.485	0.739	0.469	0.896	0.392	0.194	0.342	0.624
	(0.030)	(0.064)	(0.038)	(0.060)	(0.084)	(0.783)	(0.320)	(0.231)	(0.423)	(0.178)	(0.094)	(0.144)	(0.271)
d=500	0.134	0.269	0.138	0.241	0.208	1.487	0.734	0.465	0.909	0.384	0.192	0.338	0.616
	(0.028)	(0.062)	(0.038)	(0.059)	(0.087)	(0.785)	(0.325)	(0.234)	(0.448)	(0.179)	(0.096)	(0.150)	(0.298)
REE under spectral norm													
d=50	0.119	0.317	0.122	0.297	0.346	1.775	0.739	0.485	0.950	0.444	0.184	0.400	0.839
	(0.020)	(0.068)	(0.036)	(0.119)	(0.203)	(0.711)	(0.271)	(0.276)	(0.384)	(0.191)	(0.090)	(0.179)	(0.323)
d=100	0.089	0.254	0.106	0.250	0.280	1.683	0.705	0.448	0.914	0.400	0.167	0.363	0.796
	(0.012)	(0.061)	(0.027)	(0.076)	(0.144)	(0.784)	(0.300)	(0.256)	(0.438)	(0.183)	(0.080)	(0.175)	(0.284)
d=300	0.078	0.219	0.099	0.215	0.235	1.644	0.687	0.432	0.898	0.386	0.159	0.341	0.795
	(0.008)	(0.052)	(0.023)	(0.049)	(0.130)	(0.842)	(0.324)	(0.264)	(0.455)	(0.205)	(0.088)	(0.162)	(0.291)
d=500	0.077	0.212	0.097	0.210	0.229	1.635	0.681	0.425	0.910	0.382	0.157	0.339	0.782
	(0.008)	(0.051)	(0.021)	(0.047)	(0.138)	(0.865)	(0.330)	(0.264)	(0.473)	(0.219)	(0.092)	(0.178)	(0.323)

$\|\mathbf{M} - \widehat{\mathbf{M}}\|/\|\mathbf{M}\|$) for "EP" has a larger denominator value than the REE for "diff.EP".

Table 1.4 presents the simulation results for $n = 78$, $\bar{\tau}_G^{\diamond} = -0.4$, and $\bar{\tau}_V^{\diamond} = -0.4$. In this case, the first differenced microstructure noise factors and idiosyncratic errors are larger than the efficient return factors in magnitude. The REE's for "diff.MN" are now smaller than those for "diff.EP" for all the three methods. DPCA* still outperforms the other two methods, and DPCA performs the poorest. The results for EP and diff.EP in Table 1.3 are better than their counterparts in Table 1.4, while the results for MN and diff.MN in Table 1.3 are worse than their counterparts in Table 1.4. This suggests that factors with larger magnitudes are estimated with smaller relative errors.

Table 1.5 gives the simulation results when $n = 390$, $\bar{\tau}_G^{\diamond} = -0.6$, and $\bar{\tau}_V^{\diamond} = -0.6$. When we increase the sample size from 78 to 390, the REEs are smaller for "diff.total", "diff.EP", "diff.MN", and "EP". However, all three methods have larger REEs for "MN". When $d = 50$, the REEs are even larger than 1, but they eventually decay when d becomes larger. Table 1.6 provides the simulation results when $n = 390$, $\bar{\tau}_G^{\diamond} = -0.4$, and $\bar{\tau}_V^{\diamond} = -0.4$. The pattern is similar to that in Table 1.4, i.e., DPCA* outperforms DPCA and PCA*-VECM.

In summary, the simulation results show that DPCA* always outperforms DPCA, which means that the eigen-decomposition of $\boldsymbol{x}\boldsymbol{x}^\intercal$ gives better estimates than the eigen-decomposition of $\boldsymbol{x}^\intercal\boldsymbol{x}$ in finite samples. In addition, we can find that the difference between the average REEs of DPCA* and DPCA is larger when the magnitude of the factors for microstructure noise is larger (see Tables 1.4 and 1.6) but smaller when the magnitude of the factors for microstructure noise is smaller (see Tables 1.3 and 1.5). This may be due to the fact that the factors corresponding to the leading eigenvalues in the second-step PCA may come from the microstructure noise when the factors are not normalised in the first-step PCA, especially when the magnitude of the factors for microstructure noise is larger. Misidentification of the two types of factors leads to large REEs. Even when the magnitude of the factors for microstructure noise is smaller, estimation error in the first-step PCA may still lead to the same problem, but with a lower probability.

1.5 An empirical application

We now apply the proposed method to 1-min and 5-min intraday returns of S&P 500 Index constituents (505 stocks in total). The data were collected from the Thomson Reuters Eikon database and cover a period from 29 March 2021 to 30 June 2021. For each day, the observed prices constitute an $(n+1)$-by-d matrix, $\mathcal{X}$, with $n \leq 78$ (for 5-min returns) or $n \leq 390$ (for 1-min returns) and $d \leq 505$. The value of n (i.e., the number of observations) and d (i.e., the number of stocks) may vary from day to day due to contemporaneous missing values at a time or suspension of trading in

Table 1.5: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 390$, $\bar{\tau}_G^\diamond = -0.6$, $\bar{\tau}_V^\diamond = -0.6$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.177	0.222	0.257	0.214	1.524	0.258	0.317	0.214	1.500	0.197	0.213	0.200	1.548
	(0.014)	(0.050)	(0.080)	(0.102)	(0.933)	(0.066)	(0.102)	(0.102)	(0.893)	(0.022)	(0.032)	(0.097)	(0.609)
d=100	0.131	0.166	0.193	0.156	1.117	0.184	0.156	0.223	1.113	0.153	0.170	0.150	1.238
	(0.008)	(0.034)	(0.058)	(0.069)	(0.663)	(0.043)	(0.069)	(0.069)	(0.654)	(0.019)	(0.031)	(0.067)	(0.425)
d=300	0.089	0.112	0.133	0.100	0.630	0.116	0.140	0.100	0.628	0.115	0.138	0.102	0.943
	(0.004)	(0.020)	(0.035)	(0.034)	(0.362)	(0.022)	(0.037)	(0.034)	(0.359)	(0.021)	(0.035)	(0.034)	(0.288)
d=500	0.078	0.098	0.117	0.087	0.494	0.097	0.114	0.087	0.491	0.106	0.131	0.092	0.863
	(0.003)	(0.018)	(0.030)	(0.026)	(0.285)	(0.016)	(0.026)	(0.026)	(0.284)	(0.021)	(0.035)	(0.026)	(0.253)
REE under max norm													
d=50	0.173	0.210	0.243	0.204	1.147	0.255	0.309	0.205	1.139	0.188	0.203	0.192	1.223
	(0.031)	(0.062)	(0.086)	(0.094)	(0.520)	(0.090)	(0.124)	(0.094)	(0.514)	(0.037)	(0.046)	(0.088)	(0.416)
d=100	0.125	0.154	0.182	0.147	0.833	0.178	0.215	0.148	0.831	0.141	0.162	0.142	0.947
	(0.020)	(0.041)	(0.063)	(0.060)	(0.375)	(0.058)	(0.081)	(0.060)	(0.368)	(0.030)	(0.041)	(0.058)	(0.302)
d=300	0.084	0.104	0.134	0.097	0.470	0.109	0.142	0.098	0.468	0.106	0.139	0.099	0.687
	(0.012)	(0.027)	(0.045)	(0.029)	(0.202)	(0.032)	(0.047)	(0.029)	(0.201)	(0.030)	(0.046)	(0.030)	(0.210)
d=500	0.077	0.092	0.125	0.087	0.369	0.090	0.121	0.087	0.366	0.100	0.136	0.091	0.629
	(0.011)	(0.024)	(0.041)	(0.025)	(0.155)	(0.022)	(0.033)	(0.025)	(0.152)	(0.028)	(0.043)	(0.025)	(0.201)
REE under spectral norm													
d=50	0.151	0.191	0.214	0.214	1.516	0.244	0.262	0.214	1.477	0.170	0.183	0.201	1.543
	(0.016)	(0.050)	(0.057)	(0.116)	(0.934)	(0.080)	(0.091)	(0.116)	(0.900)	(0.026)	(0.022)	(0.113)	(0.609)
d=100	0.107	0.141	0.155	0.155	1.109	0.169	0.182	0.155	1.098	0.129	0.138	0.149	1.233
	(0.010)	(0.036)	(0.044)	(0.079)	(0.665)	(0.053)	(0.062)	(0.079)	(0.659)	(0.024)	(0.023)	(0.078)	(0.425)
d=300	0.062	0.093	0.107	0.098	0.620	0.099	0.114	0.098	0.617	0.095	0.111	0.100	0.937
	(0.004)	(0.028)	(0.034)	(0.040)	(0.365)	(0.031)	(0.036)	(0.040)	(0.362)	(0.028)	(0.034)	(0.039)	(0.289)
d=500	0.053	0.080	0.098	0.085	0.484	0.077	0.094	0.085	0.481	0.089	0.108	0.090	0.857
	(0.003)	(0.026)	(0.031)	(0.030)	(0.288)	(0.022)	(0.026)	(0.030)	(0.288)	(0.028)	(0.034)	(0.029)	(0.254)

Table 1.6: Average REE and standard deviation (in parentheses) for estimation of common components by DPCA*, DPCA and PCA*-VECM, when $n = 390$, $\bar{\tau}_G^\diamond = -0.4$, $\bar{\tau}_V^\diamond = -0.4$.

		DPCA*				DPCA				PCA*-VECM			
	diff.total	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN	diff.EP	diff.MN	EP	MN
REE under Frobenius norm													
d=50	0.152	0.342	0.101	0.234	0.480	0.825	0.339	0.258	0.637	0.358	0.112	0.257	0.864
	(0.014)	(0.035)	(0.024)	(0.096)	(0.287)	(0.414)	(0.186)	(0.117)	(0.328)	(0.039)	(0.026)	(0.099)	(0.292)
d=100	0.111	0.250	0.074	0.177	0.353	0.593	0.242	0.196	0.477	0.278	0.090	0.211	0.805
	(0.007)	(0.021)	(0.012)	(0.064)	(0.204)	(0.357)	(0.160)	(0.088)	(0.273)	(0.040)	(0.022)	(0.081)	(0.273)
d=300	0.074	0.167	0.054	0.126	0.211	0.392	0.161	0.139	0.310	0.209	0.076	0.172	0.764
	(0.003)	(0.013)	(0.007)	(0.032)	(0.114)	(0.308)	(0.135)	(0.053)	(0.209)	(0.042)	(0.022)	(0.059)	(0.257)
d=500	0.064	0.145	0.048	0.115	0.176	0.332	0.136	0.128	0.264	0.192	0.073	0.163	0.752
	(0.003)	(0.012)	(0.024)	(0.006)	(0.095)	(0.302)	(0.132)	(0.053)	(0.211)	(0.045)	(0.022)	(0.055)	(0.266)
REE under max norm													
d=50	0.162	0.356	0.106	0.243	0.382	0.943	0.379	0.278	0.533	0.363	0.117	0.259	0.648
	(0.070)	(0.072)	(0.069)	(0.105)	(0.202)	(0.525)	(0.219)	(0.128)	(0.258)	(0.070)	(0.069)	(0.103)	(0.242)
d=100	0.110	0.253	0.073	0.184	0.269	0.670	0.270	0.209	0.386	0.269	0.088	0.212	0.574
	(0.034)	(0.047)	(0.033)	(0.063)	(0.133)	(0.436)	(0.185)	(0.086)	(0.223)	(0.057)	(0.037)	(0.079)	(0.216)
d=300	0.073	0.166	0.058	0.137	0.160	0.436	0.184	0.153	0.249	0.197	0.075	0.175	0.519
	(0.014)	(0.030)	(0.014)	(0.036)	(0.062)	(0.356)	(0.147)	(0.053)	(0.168)	(0.050)	(0.023)	(0.059)	(0.197)
d=500	0.066	0.147	0.057	0.131	0.132	0.365	0.158	0.144	0.212	0.183	0.073	0.168	0.514
	(0.013)	(0.028)	(0.014)	(0.035)	(0.051)	(0.333)	(0.140)	(0.053)	(0.167)	(0.049)	(0.022)	(0.056)	(0.208)
REE under spectral norm													
d=50	0.104	0.291	0.091	0.231	0.476	0.925	0.321	0.250	0.572	0.299	0.102	0.254	0.860
	(0.015)	(0.037)	(0.021)	(0.110)	(0.286)	(0.502)	(0.188)	(0.128)	(0.309)	(0.039)	(0.022)	(0.113)	(0.291)
d=100	0.071	0.206	0.063	0.174	0.350	0.661	0.229	0.190	0.431	0.228	0.080	0.209	0.804
	(0.007)	(0.022)	(0.012)	(0.073)	(0.205)	(0.430)	(0.160)	(0.093)	(0.253)	(0.046)	(0.021)	(0.090)	(0.273)
d=300	0.041	0.127	0.041	0.122	0.207	0.430	0.153	0.134	0.280	0.178	0.064	0.171	0.763
	(0.003)	(0.017)	(0.007)	(0.036)	(0.115)	(0.369)	(0.134)	(0.054)	(0.191)	(0.058)	(0.022)	(0.067)	(0.257)
d=500	0.033	0.106	0.039	0.112	0.171	0.359	0.130	0.123	0.239	0.170	0.060	0.162	0.751
	(0.002)	(0.018)	(0.007)	(0.027)	(0.097)	(0.361)	(0.129)	(0.055)	(0.191)	(0.063)	(0.022)	(0.062)	(0.267)

one day. For asynchronous missing data, we fill them using Next Observation Carried Backwards (NOCB) if the missing data are at the beginning of the series, Last Observation Carried Forward (LOCF) if they are at the end of the series, and linear interpolation otherwise.

Firstly, we determine the number of factors using IC_1 and the PANIC tests for each of the 66 trading days within the sampling period. Figure 1.1 illustrates that the estimated number of factors varies from day to day. Table 1.7 shows some summary statistics for the estimated numbers of factors. The PANIC tests with different significance levels give similar estimates of the numbers of factors for the microstructure noise and the efficient prices in each day, with a difference less than 1.5 on average. Since the PANIC test at 1% performs best in the simulation, it will be used as the default PANIC test hereafter, unless specifically stated otherwise.

Table 1.7: Summary statistics for estimated numbers of factors over the sampling period

1-min data							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
K_H (IC_1)	13.045	13	12	14	8	17	1.818
K_F (PANIC 1%)	8.894	9	8	10	4	12	1.590
K_F (PANIC 5%)	8.106	8	7	9	4	12	1.656
K_F (PANIC 10%)	7.561	8	6	9	3	12	1.890
K_G (PANIC 1%)	4.152	4	3	5	1	8	1.552
K_G (PANIC 5%)	4.939	5	4	6	2	9	1.626
K_G (PANIC 10%)	5.485	5	4	6.75	2	10	1.629
5-min data							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
K_H (IC_1)	7.697	8	7	8.75	5	12	1.488
K_F (PANIC 1%)	6.758	7	6	8	3	12	1.710
K_F (PANIC 5%)	6.076	6	5	7	2	12	1.892
K_F (PANIC 10%	5.500	5	4	7	2	10	1.774
K_G (PANIC 1%)	0.939	1	0	2	0	3	0.990
K_G (PANIC 5%)	1.621	2	0	2	0	5	1.274
K_G (PANIC 10%)	2.197	2	2	3	0	6	1.268

We can see from Table 1.7 that the numbers of factors are larger for 1-min data than those for 5-min data. The reason might be twofold. Firstly, the signal of the efficient returns increases when 1-min data are cumulated to 5-min data. Thus the factors for the microstructure noise are more difficult to detect. Secondly, due to the Epps effect (as evident in Table 1.9), the correlation between stocks decreases as the sampling frequency increases, resulting in higher numbers of factors for both the microstructure noise and the efficient prices of higher frequency data.

To see how the number of factors change during the sample period, we looked at whether there is a relation between the number of factors and the following variables: the market excess return (MKT), the size factor (SMB), the value factor (HML), the Momentum factor (MOM), the short-

term Reversal factor (STREV), the long-term Reversal factor (LTREV), the CBOE Volatility Index (VIX), the high low volatility on the S&P500 index (HLVOL) and the implied Roll measure of bid-ask spread (ROLL, computed as $2\sqrt{-\left[Cov(r_t, r_{t-1}) \wedge 0\right]}$ using 5-min price changes r_t and averaged over all stocks with equal weights). The first 6 risk factors were downloaded from Kenneth R. French's data library[2] and the VIX and S&P500 index were downloaded from Thomson Reuters Eikon database.

We find from Table 1.8 that the contemporaneous correlations between the number of factors and the above-mentioned risk variables are insignificant. However, some of these variables can partially explain changes in the number of factors for efficient prices in the next day. Specifically, the highest (in absolute value) correlation is between lag-1 SMB and 1-min K_F, which is 0.329 with a p-value of 0.007. Significant correlations are also found, at 10% significant level, between 1-min K_F and HLVOL at -0.240, between 1-min K_F and MOM at 0.211, and between 1-min K_F and LTREV at 0.213. What is more, we also find a negative relation between the number of factors and VIX (or HLVOL). Figure 1.2 shows this relation in a time series plot: when the S&P500 high-low volatility peaks on 12 May 2021, the 1-min K_H and 1-min K_F drop in the next day. This indicates that the co-movement of stocks increases during High VIX (or HLVOL) period, confirming the old adage that diversification disappears when needed most.

Table 1.8: Correlation between number of factors and 9 risk variables. Panel A shows the contemporaneous correlation and Panel B shows the lagged correlation (the 9 risk variables are lagged by 1 day). P-values less than 0.1, 0.05 or 0.01 are flagged with one, two or three stars (*, **, ***), respectively.

(Panel A)	VIX	HLVOL	MKT	SMB	HML	MOM	STREV	LTREV	ROLL
K_H(1-min)	−0.193	−0.121	0.043	0.185	−0.038	0.178	0.141	0.153	−0.034
K_F(1-min)	−0.134	−0.033	0.004	0.159	−0.054	0.065	−0.010	0.094	−0.113
K_G(1-min)	−0.089	−0.108	0.047	0.055	0.010	0.142	0.175	0.083	0.075
K_H(5-min)	−0.008	0.123	0.042	0.107	0.115	0.086	0.090	0.205	0.083
K_F(5-min)	−0.010	0.093	0.009	0.004	−0.061	−0.011	−0.051	−0.005	0.010
K_G(5-min)	0.004	0.025	0.048	0.155	0.279	0.149	0.223	0.317	0.106
(Panel B)	VIX	HLVOL	MKT	SMB	HML	MOM	STREV	LTREV	ROLL
K_H(1-min)	−0.206*	−0.191	0.157	0.292**	0.129	0.231*	−0.184	0.254**	−0.132
K_F(1-min)	−0.131	−0.240*	0.144	0.329***	0.050	0.211*	−0.027	0.213*	0.005
K_G(1-min)	−0.105	0.023	0.036	0.002	0.099	0.052	−0.187	0.078	−0.159
K_H(5-min)	0.003	−0.088	0.110	0.026	0.041	−0.084	−0.072	−0.006	0.033
K_F(5-min)	−0.013	−0.076	0.183	−0.019	0.055	−0.085	−0.150	−0.032	−0.056
K_G(5-min)	0.027	−0.003	−0.146	0.071	−0.032	0.019	0.146	0.046	0.144

For illustration purposes, we consider the results for 30 June 2021, i.e., the last day in the sample period. For 1-min data, the estimated total number of factors is 13, among which 6 are

[2]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 1.9: Summary statistics for pairwise correlations between stock returns on 30 June 2021

Correlation							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
1-min	0.083	0.075	-0.025	0.179	-0.471	0.925	0.160
5-min	0.097	0.097	-0.062	0.253	-0.674	0.960	0.229
Absolute value of correlation							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
1-min	0.139	0.112	0.053	0.195	0.000	0.925	0.114
5-min	0.200	0.171	0.082	0.288	0.000	0.960	0.148

Table 1.10: Summary statistics for the variance ratio of the common component of the microstructure noise to that of the efficient price for each stock on 30 June 2021. Note that we define both variance ratios based on $\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Fi}^{\star\intercal}\widehat{\boldsymbol{f}}_{s\Delta}^{\star})$, as the variance $\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Fi}^{\star\intercal}\widehat{\boldsymbol{F}}_{s\Delta}^{\star})$ is not meaningful.

Variance ratio of common components: $\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Gi}^{\star\intercal}\widehat{\boldsymbol{G}}_{s\Delta}^{\star})/\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Fi}^{\star\intercal}\widehat{\boldsymbol{f}}_{s\Delta}^{\star})$							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
1-min	4.116	2.541	1.302	5.451	0.030	48.788	4.875
5-min	0.148	0.048	0.013	0.138	0.000	4.228	0.325
Variance ratio of differenced common components: $\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Gi}^{\star\intercal}\widehat{\boldsymbol{g}}_{s\Delta}^{\star})/\widehat{\text{Var}}(\widehat{\boldsymbol{\lambda}}_{Fi}^{\star\intercal}\widehat{\boldsymbol{f}}_{s\Delta}^{\star})$							
	mean	median	1st quartile	3rd quartile	min	max	s.d.
1-min	2.633	1.604	0.812	3.492	0.017	35.849	3.229
5-min	0.069	0.020	0.006	0.062	0.000	2.209	0.156

identified as factors for efficient prices by the PANIC test. Figure 1.3 shows the 13 estimated factors in cumulative form, i.e., $\left(\widehat{\boldsymbol{\beta}}_{\perp}, \widehat{\boldsymbol{\beta}}\right)^{\intercal}\widehat{\boldsymbol{H}}_{s\Delta}^{\star c}$, where $\left(\widehat{\boldsymbol{\beta}}_{\perp}, \widehat{\boldsymbol{\beta}}\right)$ is the matrix of eigenvectors of the matrix $n^{-1}\widehat{\mathscr{H}}^{\star c\intercal}\widehat{\mathscr{H}}^{\star c}$, arranged in descending order of their corresponding eigenvalues. The first 6 factors appear to be more variable than the last 7 factors.

We cannot tell from Figure 1.3 whether the microstructure noise factors dominate the efficient price factors, as estimated factors have been standardised. Instead, we calculate the variance ratio of the common component of the microstructure noise to that of efficient price for each stock, to take the magnitude of factor loadings into consideration. We give the summary statistics in Table 1.10. We can see that on average, the common components for the microstructure noise dominate the common components of efficient prices at 1-min frequency, while the relation reverses at 5-min frequency. For individual stocks, however, the contribution of the common component for the microstructure noise may still be small even at the 1-min frequency. To show this, we look at the decomposition of prices (cumulative returns) into three components: a common component of efficient prices, $\boldsymbol{\Lambda}_F\mathbf{F}_t$ (CC.EP), a common component of microstructure noise, $\boldsymbol{\Lambda}_G\mathbf{D}_G\mathbf{G}_t$ (CC.MN), and an idiosyncratic error component (Residuals). We illustrate with five randomly selected stocks that have the stock ticker symbols — POOL, CHRW, AJG, CNP, and WM. Figure 1.4 shows the decomposition for the cumulative 1-min returns of the five stocks on 30 June 2021. The correspond-

ing numbers of factors are $\widehat{K}_F = 6$ and $\widehat{K}_G = 7$. Figure 1.5 shows the decomposition for 5-min data with $\widehat{K}_F = 5$ and $\widehat{K}_G = 1$. The two figures show that the common component of the microstructure noise can explain only a small amount of the variability of the prices.

In summary, our analysis finds existence of common components for the microstructure noise of S&P 500 stocks, although their magnitude is small. The small magnitude is also consistent with the expectation that there are very few arbitrage opportunities in a frictional market.

Figure 1.1: Estimated number of factors for each trading day from 29 March 2021 to 30 June 2021. The y-axis represents the number of factors and the x-axis represents the dates (given in the format `mmdd`). The y-coordinate of the top of each grey bar gives the estimated total number of factors, $\widehat{K}_H$, from IC_1. The length of each grey bar represents the difference between $\widehat{K}_H$ and $\widehat{K}_F$, which is obtained from the PANIC test using 1% significance level. The length of each red bar represents the difference between $\widehat{K}_F$'s obtained from the PANIC tests using 1% and 5% significance levels. The length of each blue bar represents the difference between $\widehat{K}_F$'s obtained from the PANIC tests using 5% and 10% significance levels. The y-coordinate of the bottom of each blue bar gives the value of $\widehat{K}_F$ obtained from a 10% PANIC test.

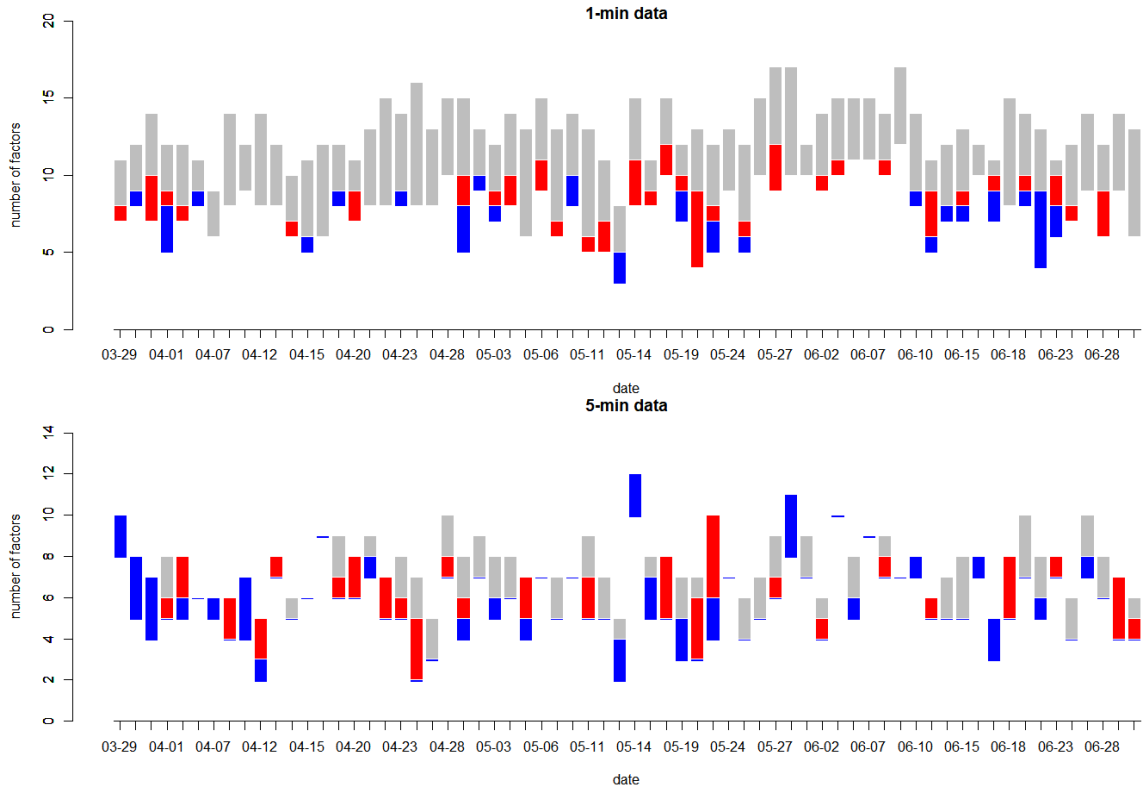
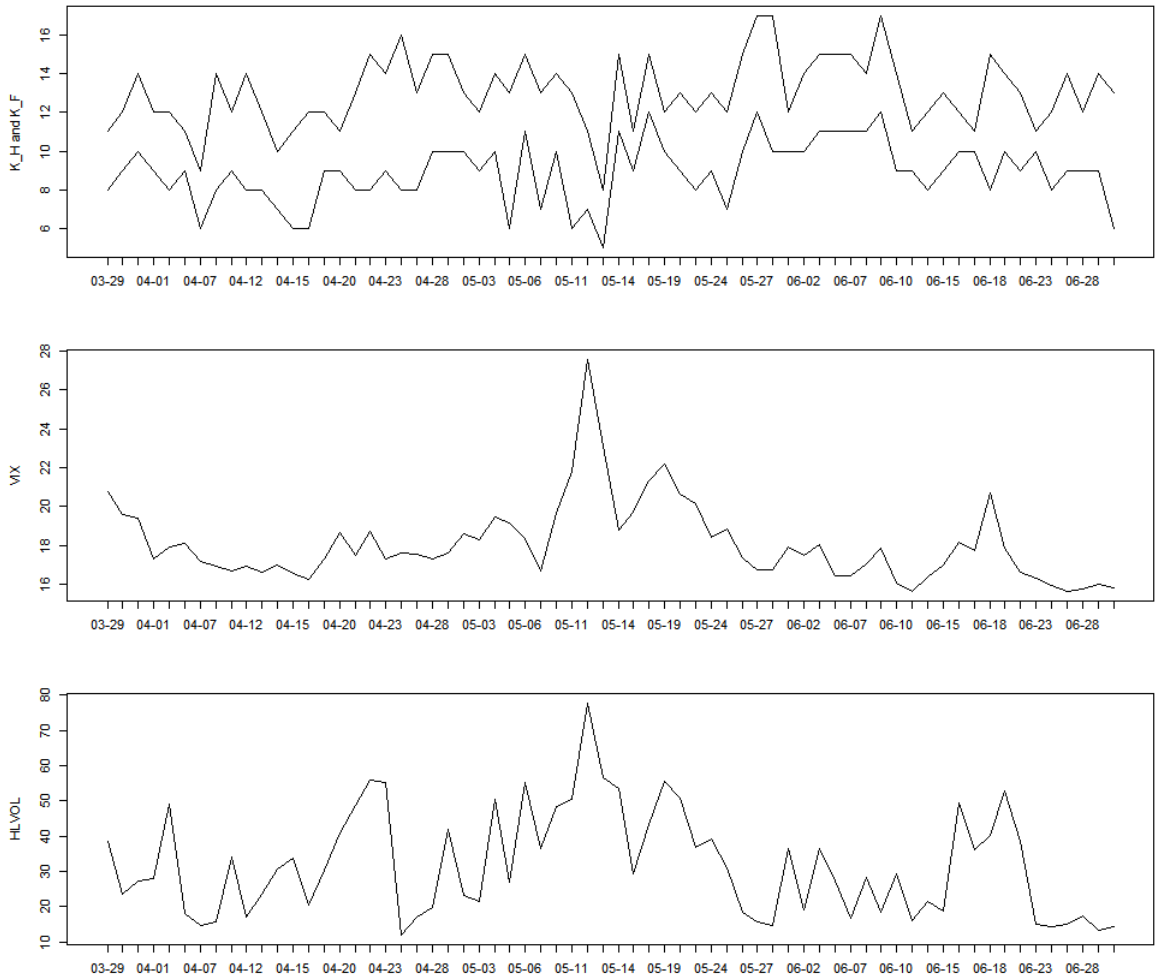


Figure 1.2: Time series plots of K_H(1-min) and K_F(1-min), VIX, and HLVOL over the sampling period



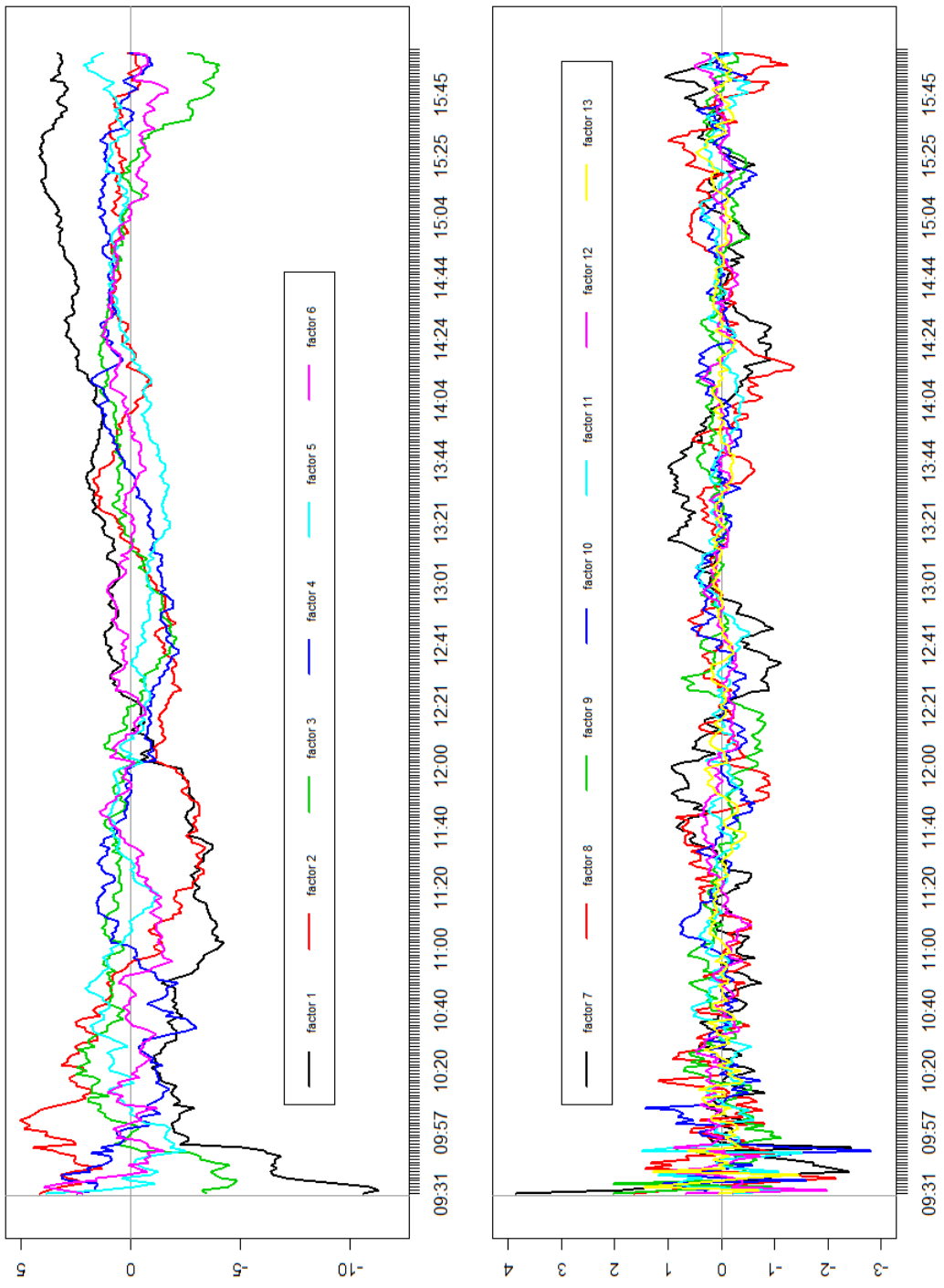


Figure 1.3: Plot of the 13 cumulative factors estimated from the DPCA*, i.e., plot of the 13 components of $\left(\widehat{\boldsymbol{\beta}}_\perp, \widehat{\boldsymbol{\beta}}\right)^\intercal \widehat{H}^{\star c}_{s\Delta}$, $s = 1, \ldots, 376$, for 1-min data on 30 June 2021

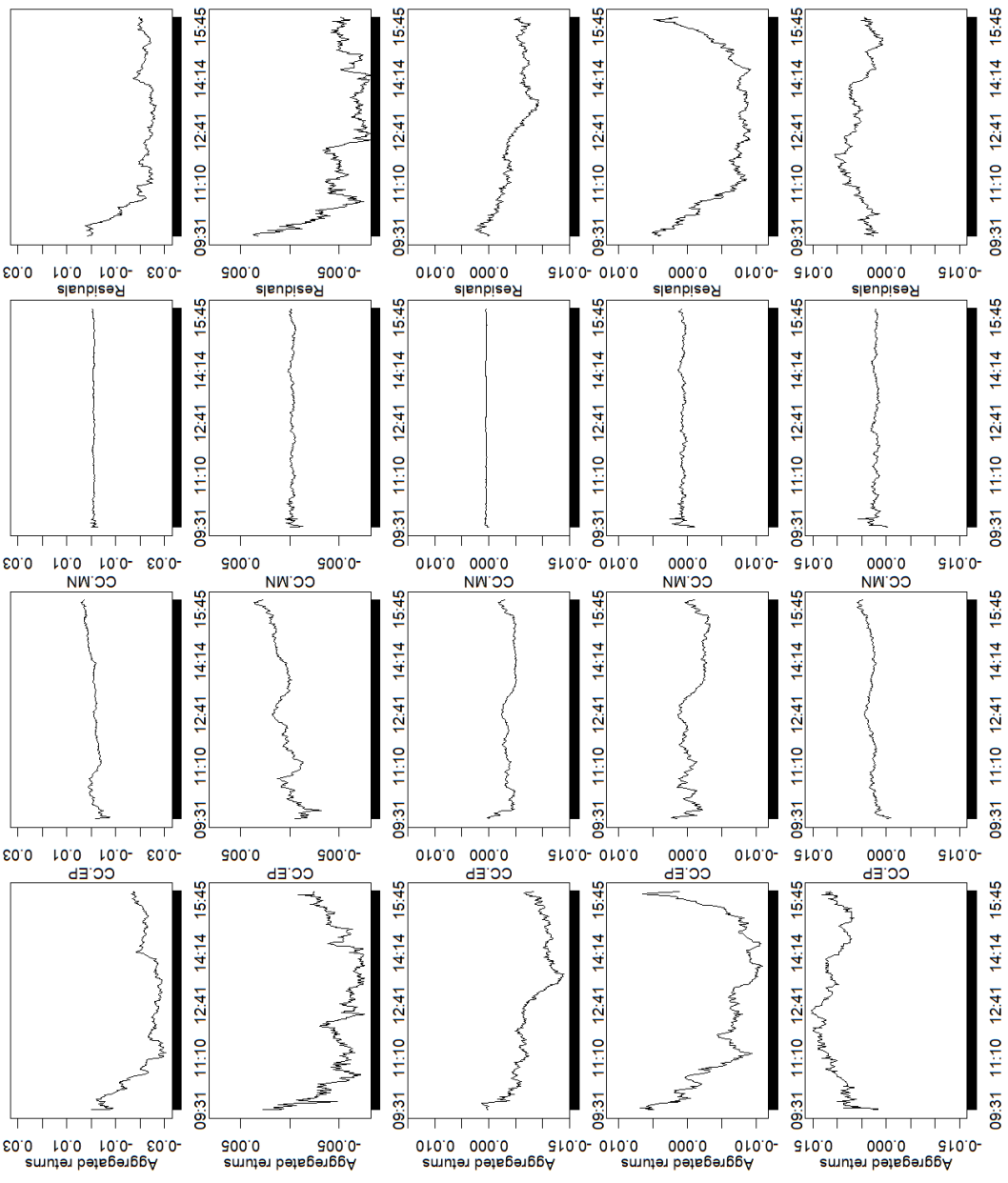


Figure 1.4: Decomposition of cumulative 1-min returns on 30 June 2021 into the common components of efficient prices (CC.EP), the common components of the microstructure noise (CC.MN), and the idiosyncratic errors (Residuals) for the stocks POOL, CHRW, AJG, CNP, and WM, with $\widehat{K}_F = 6$ and $\widehat{K}_G = 7$. Each row gives the decomposition for each stock, with the first diagram giving the cumulative returns, followed by CC.EP, CC.MN, and Residuals. The variance ratio $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\lambda}}^{\star\intercal}_{Gi}\widehat{\boldsymbol{G}}^{\star}_{s\Delta})/\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\lambda}}^{\star\intercal}_{Fi}\widehat{\boldsymbol{f}}^{\star}_{s\Delta})$ for each stock is 0.462, 0.980, 0.125, 1.282, and 6.818, respectively.

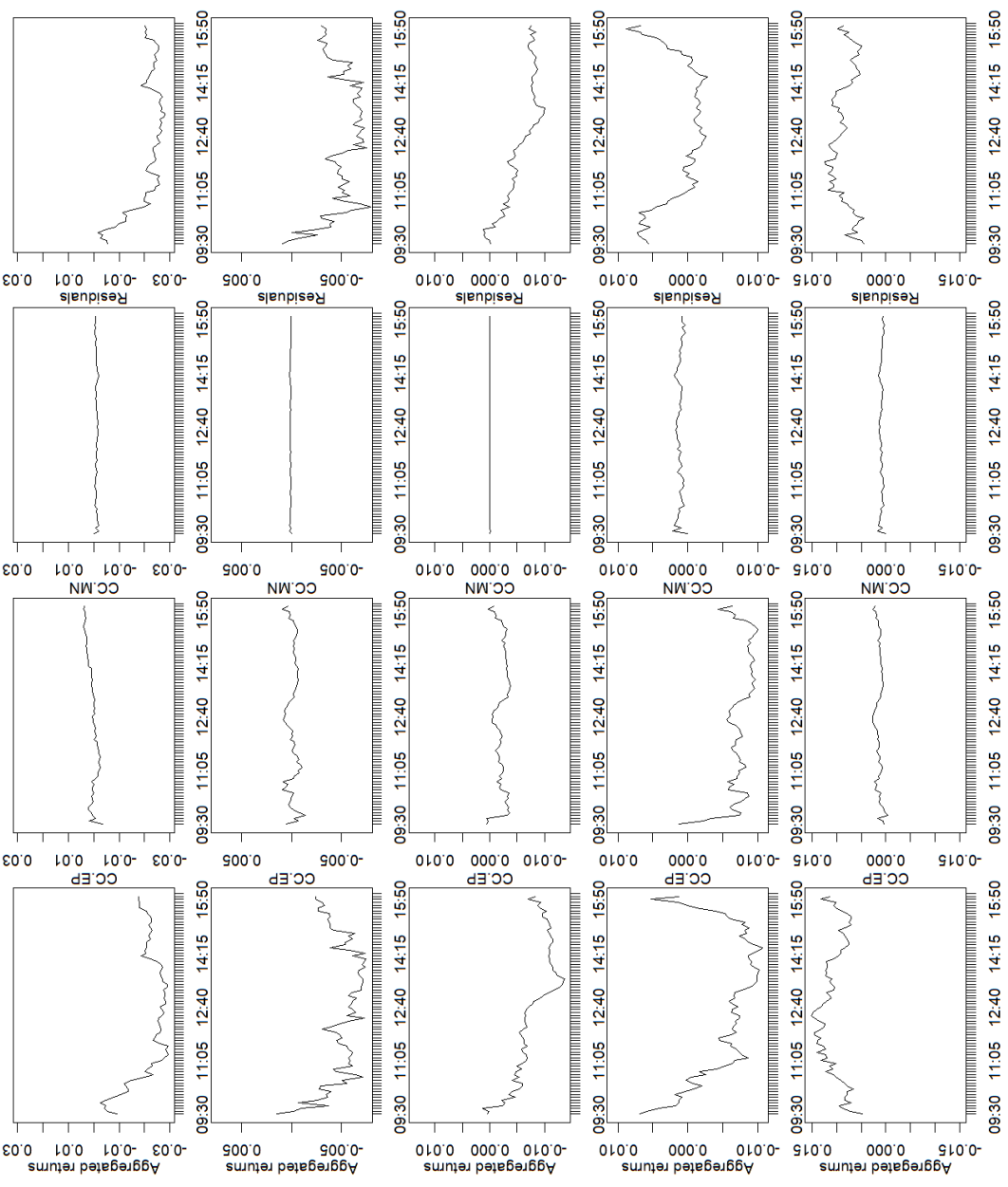


Figure 1.5: Decomposition of cumulative 5-min returns on 30 June 2021 into the common components of efficient prices (CC.EP), the common components of the microstructure noise (CC.MN), and the idiosyncratic errors (Residuals) for the stocks POOL, CHRW, AJG, CNP, and WM, with $\widehat{K}_F = 5$ and $\widehat{K}_G = 1$. Each row gives the decomposition for each stock, with the first diagram giving the cumulative returns, followed by CC.EP, CC.MN, and Residuals. The variance ratio $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\lambda}}_{Gi}^{\star\intercal}\widehat{\boldsymbol{G}}_{s\Delta}^{\star})/\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\lambda}}_{Fi}^{\star\intercal}\widehat{\boldsymbol{f}}_{s\Delta}^{\star})$ for each stock is 0.183, 0.015, 0.001, 0.180, and 0.550, respectively.

1.6 Conclusion

We consider a dual factor model for high-frequency stock prices contaminated with microstructure noise. We develop the Double Principle Component Analysis (DPCA*) to estimate the separate factor structures for efficient prices and microstructure noise. When comparing with the PCA-VECM approach, the DPCA* approach is free from the need to impose strong parametric assumptions on the microstructure noise and applies instead to a broad class of stationary processes. The estimators are proven to be consistent and perform well in simulations. The empirical analysis of intraday returns of S&P 500 constituents provides some evidence of co-movement in the microstructure noise, apart from co-movement of prices caused by common systematic risk factors.

Identifying and separating out the common component of microstructure noise from observed prices are very useful for the study of properties of the microstructure noise and efficient price processes. For example, once the common component of the microstructure noise are separated out, the common component of the efficient prices are no longer contaminated by microstructure noise and hence, can be used to obtain a more accurate estimate of the common part of realised volatility. For the idiosyncratic part of realised volatility, one can use the estimated idiosyncratic errors and apply the pre-averaging method of Jacod *et al.* (2009). Adding these two parts together, we get an estimator of the realised volatility matrix. We may introduce sparsity or block structure into idiosyncratic components like Dai *et al.* (2019) and Aït-Sahalia and Xiu (2017), respectively. However, since our main interests are the identification of common factors, we avoid introducing these structures and leave the estimation of the realised volatility matrix to the future work.

The estimated common factors and loadings for microstructure noise provide useful tools for portfolio management. With such estimates, portfolio managers can construct a new factor mimicking portfolio which is only exposed to microstructure noise factors. Such a portfolio can be used to hedge risks from microstructure noise. Since the value of the portfolio is stationary, one can apply the mean-reverting strategy to earn profits from the portfolio, once its volatility is large enough to cover the cost. Even if its volatility is small, one can still time the market according to it, e.g., when adjusting the position of another portfolio, to lower the cost.

Chapter 2

Estimating Time-Varying Networks for High-Dimensional Time Series

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.

Key Words: factor model, Granger causality, partial correlation, time-varying network, VAR.

2.1 Introduction

In recent years, network analysis has become an effective tool to explore inter-connections among a large number of variables, with applications to various disciplines such as: epidemiology, economics, finance, and social networks (e.g., Newman, 2002; Burt *et al.*, 2013; Diebold and Yılmaz, 2014, 2015; Hautsch *et al.*, 2014; Serrat, 2017; Barigozzi and Brownlees, 2019; Zhu *et al.*, 2019). The so-called graphical model is commonly used in network analysis to visualise the connectedness of a large panel with vertices representing variables in the panel and the presence of an edge indicating appropriate (conditional) dependence between the variables. In the past decades, most of the existing literature on statistical estimation and inference of network data limits attention to the *static* network, which is assumed to be invariant over time (e.g., Yuan and Lin, 2007; Fan *et al.*, 2009; Loh and Wainwright, 2013; Basu *et al.*, 2015; Zhao *et al.*, 2022). However, such an assumption may be too restrictive and often fails in practical applications where the underlying data generating mechanism is dynamic. There have been some attempts in the recent literature to relax the static network assumption, allowing the connectivity structure to exhibit time-varying features. For example, Kolar *et al.* (2010) and Zhou *et al.* (2010) study dynamic network models with smooth time-varying structural changes; whereas Wang *et al.* (2021a) 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, 2005). In recent years, there has been increasing interest in extending the finite-dimensional VAR to the high-dimensional setting. Under appropriate sparsity restrictions on the transition (or autoregressive coefficient) matrices, various regularised methods have been proposed to estimate high-dimensional VAR models and identify non-zero entries in the transition matrices (e.g., Basu and Michailidis, 2015; Han *et al.*, 2015; Kock and Callot, 2015; Davis *et al.*, 2016). Zhu *et al.* (2017) introduce a network VAR model by incorporating the adjacency matrix to capture the network effect and estimate the model via ordinary least squares. More recently, Chen *et al.* (2023) and Miao *et al.* (2023) further study high-dimensional VAR and network VAR with latent common factors, allowing strong cross-sectional dependence in large panel time series. The methodology and theory developed in these papers heavily rely on the stationarity assumption with both transition and volatility matrices being time-invariant.

The stable VAR model cannot capture smooth structural changes and breaks in the underlying data generating process, two typical dynamic features in time series data collected over a long time span. To address this problem, Ding *et al.* (2017) consider a time-varying VAR model for high-

dimensional time series (allowing the number of variables to diverge at a sub-exponential rate of the sample size), and estimate the time-varying transition matrices by combining the kernel smoothing with ℓ_1-regularisation, whereas Safikhani and Shojaie (2022) simultaneously detect breaks and estimate transition matrices in high-dimensional VAR via a three-stage procedure using the total variation penalty. Xu *et al.* (2020) detect structural breaks and estimate smooth changes (between breaks) in the covariance and precision matrices of high-dimensional time series (covering VAR as a special case). In this chapter, 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 *et al.* (2020) develop a kernel estimation method and theory for time-varying vector moving average models. This chapter complements the locally stationary time series literature by further exploring the high-dimensional dynamic network structure.

We study the time-varying VAR and network models for large-scale time series, allowing the number of variables to be much larger than the time series length. Under the sparsity assumption on the transition and precision matrices with smooth structural changes, we introduce a three-stage estimation procedure: (i) preliminary local linear estimation of the transition matrices and their derivatives with time-varying LASSO; (ii) joint local linear estimation and feature selection of the time-varying transition matrices with weighted group LASSO; (iii) estimation of the precision matrix via time-varying CLIME. To guarantee the oracle property, the weights of LASSO in the second estimation stage are constructed via a local linear approximation to the SCAD penalty (e.g., Zou and Li, 2008) using the consistent preliminary estimates obtained in the first stage. Our penalised estimation methodology for the time-varying transition matrices is connected to various nonparametric

screening and shrinkage methods developed for high-dimensional functional-coefficient models (e.g., Wang and Xia, 2009; Lian, 2012; Fan *et al.*, 2014a; Liu *et al.*, 2014; Li *et al.*, 2015a), whereas the time-varying CLIME is a natural extension of the conventional CLIME for static precision matrix estimation (e.g., Cai *et al.*, 2011). The theoretical properties of the techniques developed in the aforementioned literature (such as the oracle property and minimax optimal convergence rates) rely on the independent data assumption. Extension of the methodology and theory to the high-dimensional locally stationary time series is non-trivial, requiring new technical tools such as the concentration inequality for time-varying VAR. Under some regularity conditions, we show that the proposed local linear estimates with weighted group LASSO equal to the infeasible oracle estimates with prior information on the significant entries of time-varying transition matrices, and the precision matrix estimate with time-varying CLIME is uniformly consistent with sensible convergence rates under various matrix norms. The estimated transition matrices are used to consistently estimate the uniform network structure with directed Granger causality linkages, whereas the estimated precision matrix is used to construct the network structure with undirected partial correlation linkages.

We further consider highly-correlated large-scale time series, for which the sparsity model assumption is no longer valid 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. This chapter thus substantially extends the recent work on the factor-adjusted stable VAR model estimation (e.g., Fan *et al.*, 2021; Barigozzi *et al.*, 2022; Krampe and Margaritella, 2021).

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 this chapter is organised as follows. Section 2.2 introduces the time-varying VAR and network model structures. Section 2.3 presents the estimation procedures for the time-varying transition and precision matrices and Section 2.4 gives the asymptotic properties of the developed estimates. Section 2.5 considers the factor-adjusted time-varying VAR model and network estimation. Sections 2.6 and 2.7 report simulation studies and an empirical application, respectively. Section 2.8 concludes this chapter. Appendix B contains proofs of the main theorems, some technical lemmas with proofs, verification of a key assumption and discussions on tuning parameter selection. Throughout this chapter, 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 $\|\mathbf{W}\|_O=\lambda_{\max}^{1/2}\left(\mathbf{W}^\intercal\mathbf{W}\right)$ be the operator norm, $\|\mathbf{W}\|_F=\left[\mathsf{Tr}\left(\mathbf{W}^\intercal\mathbf{W}\right)\right]^{1/2}$ the Frobenius norm, $\|\mathbf{W}\|_1=\max_{1\le j\le d}\sum_{i=1}^d|w_{ij}|$, $\|\mathbf{W}\|_{\max}=\max_{1\le i\le d}\max_{1\le j\le d}|w_{ij}|$, 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 $a_n/b_n\to 1$, $0<\underline{c}\le a_n/b_n\le\overline{c}<\infty$ and $b_n/a_n\to 0$, respectively.

2.2 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 2.5 will further generalise them to the factor-adjusted time-varying VAR and network setting.

2.2.1 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:

$$X_t=\sum_{k=1}^p\mathbf{A}_{t,k}X_{t-k}+e_t\quad\text{with}\quad e_t=\mathbf{\Sigma}_t^{1/2}\varepsilon_t,\quad t=1,\ldots,n,\tag{2.2.1}$$

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 (2.2.1) 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, (2.2.1) becomes the same model as that considered by Ding *et al.* (2017). Furthermore, when both $\mathbf{A}_{t,k}$, $k = 1, \ldots, p$, and $\mathbf{\Sigma}_t$ are time-invariant constant matrices, (2.2.1) becomes the high-dimensional stable VAR:

$$X_t = \sum_{k=1}^{p} \mathbf{A}_k X_{t-k} + \mathbf{\Sigma}^{1/2} \varepsilon_t, \tag{2.2.2}$$

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

Assumption 2.A. (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 $\|u\| = 1$, $\mathsf{E}\left[\exp\left\{\iota_1 (u^\intercal \varepsilon_t)^2\right\}\right] \leq C_0 < \infty$, where ι_1 and C_0 are positive constants.

The first condition in Assumption 2.A(i) is a natural extension of the stability assumption imposed on the constant transition matrices (e.g., Lütkepohl, 2005), indicating that the time-varying VAR process is locally stationary/stable and leading to the following Wold representation

$$X_t = \sum_{k=0}^{\infty} \mathbf{\Phi}_{t,k} e_{t-k}, \tag{2.2.3}$$

with the coefficient matrices $\mathbf{\Phi}_{t,k}$ being absolutely summable (in an 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 $k \geq 1$. Assume that, for k sufficiently large,

$$\max_{0 \leq t \leq n} \|\mathbf{\Phi}_{t,k}\|_O \leq C_1 \rho^k, \tag{2.2.4}$$

where C_1 is a positive constant and $0 < \rho < 1$. A similar assumption can be found in Ding *et al.* (2017). In some special model settings, (2.2.4) may be violated, and we refer interested readers to the discussions in Basu and Michailidis (2015) and Liu and Zhang (2021). In fact, the condition (2.2.4) may be removed by imposing some high-level conditions (e.g., the sub-Gaussian condition on $x_{t,i}$ proved in Lemma B.2.1). The smoothness conditions in Assumption 2.A(i)(ii) are common in kernel-based local estimation method and theory. The sub-Gaussian moment condition in Assumption 2.A(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).

2.2.2 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 $1 \leq t \leq n$ and $0 \leq \tau \leq 1$. We define the network structure via a time-varying graph $\mathbb{G}_t = (\mathbb{V}, \mathbb{E}_t)$, where $\mathbb{V} = \{1, 2, \ldots, d\}$ 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 *et al.* (2010) for independent network data. Following the stable network analysis in Barigozzi and Brownlees (2019) and Barigozzi *et al.* (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 $\mathbf{A}_k = (a_{k,ij})_{d\times d}$ in (2.2.2) 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 (2.2.1). 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

$$\mathbb{E}_\tau^G = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, k \in \{1,2,\ldots,p\},\ a_{k,ij}(\tau) \neq 0\}\,. \tag{2.2.5}$$

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 (2.2.1). 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

$$\mathbb{E}_\tau^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \omega_{ij}(\tau) \neq 0,\ i \neq j\}\,. \tag{2.2.6}$$

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

$$\mathbb{E}^G = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, k \in \{1,2,\ldots,p\} \text{ and } \tau \in (0,1),\ a_{k,ij}(\tau) \neq 0\} \tag{2.2.7}$$

and

$$\mathbb{E}^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, \tau \in (0,1),\ \omega_{ij}(\tau) \neq 0,\ i \neq j\}\,. \tag{2.2.8}$$

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 2.3.4 below defines the discrete versions of the above uniform networks and provide their estimates.

2.3 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,

$$\boldsymbol{\alpha}_{i\bullet}(\cdot) = \left[A_{1,i}^\intercal(\cdot), \ldots, A_{p,i}^\intercal(\cdot)\right]^\intercal, \quad \mathbf{X}_t = \left(X_t^\intercal, \ldots, X_{t-p+1}^\intercal\right)^\intercal, \tag{2.3.1}$$

and $\tau_t = t/n$. The time-varying VAR model (2.2.1) can be equivalently written as

$$x_{t,i} = \boldsymbol{\alpha}_{i\bullet}^{\intercal}(\tau_t)\mathbf{X}_{t-1} + e_{t,i} \quad \text{with} \quad e_{t,i} = C_i^{\intercal}(\tau_t)\varepsilon_t, \quad i = 1, \ldots, d, \tag{2.3.2}$$

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 *et al.*, 2008; Wang and Xia, 2009; Lian, 2012; Cheng *et al.*, 2014; Fan *et al.*, 2014a; Liu *et al.*, 2014; Li *et al.*, 2015a). 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.

2.3.1 Preliminary time-varying LASSO estimation

For $\tau \in (0, 1)$, under the smoothness condition on the transition matrices in Assumption 2.A(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):

$$\mathscr{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) = \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\intercal}\mathbf{X}_{t-1}\right\}^2 K_h(\tau_t - \tau), \tag{2.3.3}$$

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 $\mathscr{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$ 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 *et al.*, 2011). In our high-dimensional setting, as the number of predictors may exceed n, it is challenging to obtain satisfactory estimation by directly minimising $\mathscr{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$. 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 (2.3.3).

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

$$\mathscr{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) = \mathscr{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) + \lambda_1 \left(|\boldsymbol{\alpha}|_1 + h|\boldsymbol{\beta}|_1\right), \tag{2.3.4}$$

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 $\mathscr{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$ 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.

2.3.2 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 $\mathscr{J}_i = \bigcup_{t=1}^n \mathscr{J}_i(\tau_t)$ and $\mathscr{J}'_i = \bigcup_{t=1}^n \mathscr{J}'_i(\tau_t)$, where

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

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,

$1 \leq j \leq pd$, such that $D_{i,j} = 0$, where

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

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

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

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:

$$\mathcal{Q}_i(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^{n} \mathcal{L}_i(\boldsymbol{\alpha}_{\bullet t}, \boldsymbol{\beta}_{\bullet t} \mid \tau_t) + \sum_{j=1}^{pd} p'_{\lambda_2}\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \|\boldsymbol{\alpha}_j\| + \sum_{j=1}^{pd} p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right) \|h\boldsymbol{\beta}_j\|, \tag{2.3.5}$$

where

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

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

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

with $a_0 = 3.7$ as suggested in Fan and Li (2001) and $I(\cdot)$ being the indicator function. The penalty terms in (2.3.5) are motivated by the local linear approximation to the SCAD penalty function (Zou and Li, 2008). The terms $p'_{\lambda_2}\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right)$ and $p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right)$ in (2.3.5) serve as the weights for the group LASSO, and their values are determined by the preliminary estimates in Section 2.3.1, 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 $\mathcal{Q}_i(\mathbf{A}, \mathbf{B})$ with respect to $\mathbf{A}$ and $\mathbf{B}$, where

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

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

The index set $\mathscr{J}_i$ is estimated by $\widehat{\mathscr{J}}_i = \{j : \ \widehat{\boldsymbol{\alpha}}_{i,j} \neq \mathbf{0}_n\}$, and $\mathscr{J}'_i$ is estimated by $\widehat{\mathscr{J}}'_i = \left\{j : \ \widehat{\boldsymbol{\alpha}}'_{i,j} \neq \mathbf{0}_n\right\}$, where $\mathbf{0}_k$ is a k-dimensional vector of zeros. A similar shrinkage estimation method is used by Li *et al.* (2015a) and Chen *et al.* (2021a) 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.

2.3.3 Estimation of the time-varying precision matrix

In this section, we study the estimation of $\boldsymbol{\Omega}(\cdot)$ in model (2.2.1), 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 *et al.*, 2011). Xu *et al.* (2020) further introduce a time-varying CLIME method for high-dimensional locally stationary time series which are observable. Note that in this chapter, $\boldsymbol{\Omega}(\cdot)$ 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 2.3.2, 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

$$\widehat{e}_t = (\widehat{e}_{t,1}, \ldots, \widehat{e}_{t,d})^{\intercal} = X_t - \sum_{k=1}^{p} \widehat{\mathbf{A}}_k(\tau_t) X_{t-k}, \quad t = 1, \ldots, n. \tag{2.3.6}$$

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

$$\begin{aligned} \mathscr{S}(q, \xi_d) = \Big\{ & \mathbf{W}(\tau) = [w_{ij}(\tau)]_{d \times d}, 0 \leq \tau \leq 1 : \ \mathbf{W}(\tau) \succ 0, \sup_{0 \leq \tau \leq 1} \|\mathbf{W}(\tau)\|_1 \leq C_2, \\ & \sup_{0 \leq \tau \leq 1} \max_{1 \leq i \leq d} \sum_{j=1}^{d} |w_{ij}(\tau)|^q \leq \xi_d \Big\}, \end{aligned} \tag{2.3.7}$$

where $0 \leq q < 1$, 0^0 is defined as 0, "$\mathbf{W} \succ 0$" denotes that $\mathbf{W}$ is positive definite, and C_2 is a

bounded positive constant. Define

$$\widehat{\mathbf{\Sigma}}(\tau) = [\widehat{\sigma}_{ij}(\tau)]_{d\times d} \quad \text{with} \quad \widehat{\sigma}_{ij}(\tau) = \sum_{t=1}^{n} \varpi_{n,t}(\tau)\widehat{e}_{t,i}\widehat{e}_{t,j} / \sum_{t=1}^{n} \varpi_{n,t}(\tau), \tag{2.3.8}$$

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

$$\widetilde{\mathbf{\Omega}}(\tau) = [\widetilde{\omega}_{ij}(\tau)]_{d\times d} = \underset{\mathbf{\Omega}}{\operatorname{argmin}} |\mathbf{\Omega}|_1 \quad \text{subject to} \quad \left\| \widehat{\mathbf{\Sigma}}(\tau)\mathbf{\Omega} - \mathbf{I}_d \right\|_{\max} \leq \lambda_3, \tag{2.3.9}$$

where λ_3 is a tuning parameter. As the underlying time-varying precision matrix is symmetric, the matrix estimate obtained from (2.3.9) needs to be symmetrised to obtain the final estimate, denoted as $\widehat{\mathbf{\Omega}}(\tau) = [\widehat{\omega}_{ij}(\tau)]_{d\times d}$, where

$$\widehat{\omega}_{ij}(\tau) = \widehat{\omega}_{ji}(\tau) = \widetilde{\omega}_{ij}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| \leq |\widetilde{\omega}_{ji}(\tau)|\right) + \widetilde{\omega}_{ji}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| > |\widetilde{\omega}_{ji}(\tau)|\right). \tag{2.3.10}$$

2.3.4 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:

$$\mathbb{E}_n^G = \{(i,j) \in \mathbb{V}\times\mathbb{V} : \ \exists\, k \in \{1,2,\ldots,p\} \ \text{ and } \ t \in \{1,\ldots,n\},\ a_{k,ij}(\tau_t) \neq 0\} \tag{2.3.11}$$

and

$$\mathbb{E}_n^P = \{(i,j) \in \mathbb{V}\times\mathbb{V} : \ \exists\, t \in \{1,\ldots,n\},\ \omega_{ij}(\tau_t) \neq 0,\ i \neq j\}. \tag{2.3.12}$$

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 2.3.2 and 2.3.3, we can estimate $\mathbb{E}_n^G$ by

$$\widehat{\mathbb{E}}_n^G = \left\{(i,j) \in \mathbb{V}\times\mathbb{V} : \ \exists\, k \in \{1,2,\ldots,p\},\ \sum_{t=1}^{n} \widehat{a}_{k,ij}^2(\tau_t) > 0\right\}, \tag{2.3.13}$$

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

$$\widehat{\mathbb{E}}_n^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, t \in \{1,\ldots,n\},\ |\widehat{\omega}_{ij}(\tau_t)| \geq \lambda_3,\ i \neq j\}\,, \tag{2.3.14}$$

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

2.4 Main theoretical results

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

$$X_t = \mathbf{A}(\tau_t)X_{t-1} + \mathbf{\Sigma}_t^{1/2}\varepsilon_t, \tag{2.4.1}$$

where $\mathbf{A}(\tau) = [\alpha_{ij}(\tau)]_{d\times d}$. For a general time-varying VAR(p) model (2.2.1), 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 (2.3.1), $\mathbf{e}_t = \left(e_t^\intercal, 0_d^\intercal, \ldots, 0_d^\intercal\right)^\intercal$, and $\mathbf{A}_t^*$ is a $(pd)\times(pd)$ time-varying transition matrix:

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

2.4.1 Uniform consistency of the time-varying LASSO estimates

Define

$$\mathbf{\Psi}(\tau) = \begin{bmatrix} \mathbf{\Psi}_0(\tau) & \mathbf{\Psi}_1(\tau) \\ \mathbf{\Psi}_1(\tau) & \mathbf{\Psi}_2(\tau) \end{bmatrix} \quad \text{with} \quad \mathbf{\Psi}_k(\tau) = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\tau_t - \tau}{h}\right)^k X_{t-1}X_{t-1}^\intercal K_h(\tau_t - \tau), \quad k = 0,1,2, \tag{2.4.2}$$

and

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

where $\mathscr{I}_i(\tau)$ and $\mathscr{I}_i'(\tau)$ are defined as in Section 2.3.2 but with $p = 1$. To derive the uniform consistency property of the preliminary time-varying LASSO estimates defined in Section 2.3.1, we need the following assumptions, some of which may be weakened at the cost of lengthier proofs.

Assumption 2.B. (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

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

where $s = \max_{1 \le i \le d} s_i$ with s_i being the cardinality of the index set $\mathscr{I}_i$.

Assumption 2.C. (i) The tuning parameter λ_1 satisfies

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

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

$$\min_{1 \le i \le d} \min_{1 \le t \le n} \inf_{u \in \mathscr{B}_i(\tau_t)} u^{\intercal} \mathbf{\Psi}(\tau_t) u \ge \kappa_0. \tag{2.4.3}$$

Assumption 2.B(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.B(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 2.C(ii) can be seen as a uniform version of the so-called restricted eigenvalue condition widely used in high-dimensional linear regression models (e.g., Bickel *et al.*, 2009; Basu and Michailidis, 2015). Appendix B.4 provides sufficient conditions for the high-dimensional locally stationary Gaussian time series to satisfy Assumption 2.C(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 $\max_{1 \le t \le n} \|\mathbf{\Psi}(\tau_t) - \mathsf{E}[\mathbf{\Psi}(\tau_t)]\|_{\max} = O_P\left(\sqrt{\log(n \vee d)/(nh)}\right)$. Then, using Lemma B.4.1

in Appendix B.4 and assuming $s\sqrt{\log(n \vee d)/(nh)} = o(1)$, a sufficient condition for (2.4.3) is

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

Theorem 2.4.1. *Suppose that Assumptions 2.A–2.C are satisfied. Then we have*

$$\max_{1\leq i\leq d}\max_{1\leq t\leq n}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t) - \boldsymbol{\alpha}_{i\bullet}(\tau_t)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{2.4.4}$$

Theorem 2.4.1 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 (2.4.4) can be strengthened to

$$\max_{1\leq i\leq d}\sup_{0\leq\tau\leq 1}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau) - \boldsymbol{\alpha}_{i\bullet}(\tau)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{2.4.5}$$

A similar uniform convergence property holds for the first-order derivative function estimates, see (B.1.1) in the proof of Theorem 2.4.1.

2.4.2 The oracle property of the weighted group LASSO estimates

Denote the complement of $\mathscr{J}_i$ and $\mathscr{J}_i'$ as $\overline{\mathscr{J}}_i$ and $\overline{\mathscr{J}}_i'$, respectively, i.e., $\overline{\mathscr{J}}_i = \bigcap_{t=1}^n \{j:\ \alpha_{i,j}(\tau_t) = 0\}$ and $\overline{\mathscr{J}}_i' = \bigcap_{t=1}^n \left\{j:\ \alpha_{i,j}'(\tau_t) = 0\right\}$. 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 $j \in \overline{\mathscr{J}}_i$ and $\boldsymbol{\beta}_{\bullet t}^o = (\beta_{1|t}^o, \ldots, \beta_{d|t}^o)^{\intercal}$ with $\beta_{j|t}^o = 0$ for $j \in \overline{\mathscr{J}}_i'$. Define the (infeasible) oracle estimates:

$$\widehat{\mathbf{A}}_i^o = \left(\widehat{\boldsymbol{\alpha}}_{i,1}^o, \ldots, \widehat{\boldsymbol{\alpha}}_{i,d}^o\right) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}_{i,j}^o = \left[\widehat{\alpha}_{i,j}^o(\tau_1), \ldots, \widehat{\alpha}_{i,j}^o(\tau_n)\right]^{\intercal}, \tag{2.4.6}$$

$$\widehat{\mathbf{B}}_i^o = \left(\widehat{\boldsymbol{\alpha}}_{i,1}^{\prime o}, \ldots, \widehat{\boldsymbol{\alpha}}_{i,d}^{\prime o}\right) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}_{i,j}^{\prime o} = \left[\widehat{\alpha}_{i,j}^{\prime o}(\tau_1), \ldots, \widehat{\alpha}_{i,j}^{\prime o}(\tau_n)\right]^{\intercal}, \tag{2.4.7}$$

as the values of $\mathbf{A}^o$ and $\mathbf{B}^o$ that minimise $\mathscr{Q}_i(\mathbf{A}^o, \mathbf{B}^o)$. 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 2.D. (i) The tuning parameter λ_2 satisfies

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

where $\zeta_{n,d}$ is defined in Assumption 2.C(i).

(ii) It holds that

$$\min_{1\le i\le d}\min_{j\in\mathscr{I}_i}\left(\sum_{t=1}^{n}\alpha_{i,j}^2(\tau_t)\right)^{\frac{1}{2}} \ge (a_0+1)\lambda_2 \quad \text{and} \quad \min_{1\le i\le d}\min_{j\in\mathscr{I}_i'} D_{i,j} \ge (a_0+1)\lambda_2,$$

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

Theorem 2.4.2. *Suppose that Assumptions 2.A–2.D are satisfied. The minimiser to the objective function of the weighted group LASSO, $\mathcal{Q}_i(\mathbf{A},\mathbf{B})$, exists and equals the oracle estimates defined in (2.4.6) and (2.4.7) w.p.a.1. In addition, we have the following mean squared convergence result:*

$$\max_{1\le i\le d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left[\widehat{\alpha}_{ij}(\tau_t) - \alpha_{ij}(\tau_t)\right]^2 = O_P\left(s\zeta_{n,d}^2\right), \tag{2.4.8}$$

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

Since the penalised local linear estimates are identical to the infeasible oracle estimates defined in (2.4.6) and (2.4.7) *w.p.a.1*, the sparsity property holds for the global model selection procedures proposed in Section 2.3.2, i.e., the zero elements in the time-varying transition matrix can be estimated exactly as zeros. Following the proof of Theorem 2.4.2, we may verify properties (i)–(iv) for the folded concave penalty function discussed in Fan *et al.* (2014b) *w.p.a.1*. Hence, Theorem 2.4.2 may be regarded as a generalisation of Theorem 1 in Fan *et al.* (2014b) and Theorem 3.1 in Li *et al.* (2015a) to high-dimensional locally stationary time series.

With the oracle property in Theorem 2.4.2, it is straightforward to derive the following consistency property of the network estimates for the directed edges of Granger causality linkages.

Corollary 2.4.3. Under the assumptions of Theorem 2.4.2, we have

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^G = \mathbb{E}_n^G\right) \to 1. \tag{2.4.9}$$

2.4.3 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 2.E. (i) The bandwidth b satisfies

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

In addition, $s\zeta_{n,d}\sqrt{\log(n \vee d)} \to 0$, where $\zeta_{n,d}$ is defined in Assumption 2.C(i).

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

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

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

Theorem 2.4.4. *Suppose Assumptions 2.A–2.E are satisfied and* $\{\mathbf{\Omega}(\tau) : 0 \le \tau \le 1\} \in \mathcal{S}(q, \xi_d)$. *Then we have*

$$\sup_{0\le\tau\le1} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{\max} = O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right), \tag{2.4.10}$$

$$\sup_{0\le\tau\le1} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{O} = O_P\left(\xi_d(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*})^{1-q}\right), \tag{2.4.11}$$

$$\sup_{0\le\tau\le1} \frac{1}{d}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{F}^2 = O_P\left(\xi_d(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*})^{2-q}\right), \tag{2.4.12}$$

where ξ_d is defined in (2.3.7), $\nu_{n,d}^{\diamond}$ and $\nu_{n,d}^{*}$ are defined in Assumption 2.E(ii).

The uniform convergence rates in Theorem 2.4.4 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 2.4.4 can be seen as an extension of Theorem 1 in Cai *et al.* (2011) to the high-dimensional locally stationary time series setting.

From Theorem 2.4.4, we readily have the following consistency property for the network estimates of the undirected edges of partial correlation linkages.

Corollary 2.4.5. Under the assumptions of Theorem 2.4.4, if $\min_{(i,j)\in\mathbb{E}^P} \min_{1\le t\le n} |\omega_{ij}(\tau_t)| \gg \lambda_3$, we have

$$\mathsf{P}\left(\widehat{\mathbb{E}}^P_n = \mathbb{E}^P_n\right) \to 1. \tag{2.4.13}$$

2.5 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:

$$Z_t = \mathbf{\Lambda} F_t + X_t, \quad t = 1, \ldots, n, \tag{2.5.1}$$

where $\mathbf{\Lambda} = (\Lambda_1, \ldots, \Lambda_d)^{\intercal}$ 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 (2.2.1). More generally, we may assume the following time-varying factor model structure:

$$Z_t = \mathbf{\Lambda}_t F_t + X_t, \quad t = 1, \ldots, n, \tag{2.5.2}$$

where $\mathbf{\Lambda}_t = \mathbf{\Lambda}(t/n)$ 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 *et al.*, 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 2.3. However, this is non-trivial as neither the common components ($\mathbf{\Lambda} F_t$ or $\mathbf{\Lambda}_t F_t$) nor the idiosyncratic error components are observable. Motivated by recent work on bridging factor and sparse models for high-dimensional data (e.g., Fan *et al.*, 2021; Krampe and Margaritella, 2021), 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 (2.5.1), 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

$$\widehat{X}_t = Z_t - \widehat{\mathbf{\Lambda}}\widehat{F}_t, \quad t = 1, \ldots, n. \tag{2.5.3}$$

For the time-varying factor model (2.5.2), 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 2.3.1, 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

$$\widehat{X}_t = Z_t - \widehat{\mathbf{\Lambda}}(\tau_t)\widehat{F}(\tau_t), \quad t = 1, \ldots, n, \tag{2.5.4}$$

where we've kept the same notation $\widehat{X}_t$ as in (2.5.3) to avoid notational burden.

As in Section 2.4, 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 2.3. 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 (2.3.13) and (2.3.14), 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 2.C–2.E to incorporate the approximation error of the idiosyncratic error components.

Assumption 2.F. (i) Denote $\delta_X = \max_{1\le t\le n}\left|\widehat{X}_t - X_t\right|_{\max}$. It holds that $[\log(n\vee d)]^{1/2}s\delta_X = o_P(1)$.

(ii) Assumption 2.C(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 2.D(i) holds when $\zeta_{n,d}$ is replaced by $\zeta_{n,d}^{\dagger}$.

(iv) Assumption 2.E 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 2.F(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 *et al.* (2013), a typical rate for the approximation error from PCA estimation of the conventional factor model (2.5.1) is

$$\delta_X = O_P\left((\log n)^{1/2}\left[(\log d)^{1/2}n^{-1/2} + n^{1/\upsilon}d^{-1/2}\right]\right), \tag{2.5.5}$$

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 (2.5.2) when the local PCA estimation is used. In Assumption 2.F(ii)–(iv), we amend Assumptions 2.C(i), 2.D(i) and 2.E(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 (2.5.5) can be simplified to $\delta_X = O_P\left((\log d)n^{-1/2}\right) = o_P(h^2)$ and thus $\zeta_{n,d} \propto \zeta_{n,d}^{\dagger}$. Consequently, we may remove Assumption 2.F(ii)–(iv) and δ_X would not be involved in the estimation convergence rates under model (2.5.1).

The following two propositions extend the theoretical results in Section 2.4 to the factor-adjusted time-varying VAR and networks.

Proposition 2.5.1. Suppose that the factor model (2.5.1) or (2.5.2), and Assumptions 2.A, 2.B and 2.C(ii) are satisfied.

(i) Under Assumption 2.F(i)(ii), we have

$$\max_{1\leq i\leq d}\max_{1\leq t\leq n}\sum_{j=1}^{d}\left[\widetilde{\alpha}_{ij}^{\dagger}(\tau_t)-\alpha_{ij}(\tau_t)\right]^2=O_P\left(s\lambda_1^2\right). \tag{2.5.6}$$

(ii) Under Assumption 2.F(i)–(iii), the oracle property holds for the second-stage weighted group LASSO estimates and furthermore,

$$\max_{1\leq i\leq d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left[\widehat{\alpha}_{ij}^{\dagger}(\tau_t)-\alpha_{ij}(\tau_t)\right]^2=O_P\left(s\left(\zeta_{n,d}^{\dagger}\right)^2\right). \tag{2.5.7}$$

(iii) Under Assumption 2.F and the sparsity condition that $\{\mathbf{\Omega}(\tau):0\leq\tau\leq 1\}\in\mathscr{S}(q,\xi_d)$, we have

$$\sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau)-\mathbf{\Omega}(\tau)\right\|_{\max}=O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{\dagger}\right), \tag{2.5.8}$$

$$\sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau)-\mathbf{\Omega}(\tau)\right\|_{O}=O_P\left(\xi_d(\nu_{n,d}^{\diamond}+\nu_{n,d}^{\dagger})^{1-q}\right), \tag{2.5.9}$$

$$\sup_{0\leq\tau\leq 1}\frac{1}{d}\left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau)-\mathbf{\Omega}(\tau)\right\|_{F}^{2}=O_P\left(\xi_d(\nu_{n,d}^{\diamond}+\nu_{n,d}^{\dagger})^{2-q}\right). \tag{2.5.10}$$

Proposition 2.5.2. (i) Under the assumptions of Proposition 2.5.1(ii), we have

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^{G,\dagger}=\mathbb{E}_n^{G}\right)\to 1. \tag{2.5.11}$$

(ii) Under the assumptions of Proposition 2.5.1(iii) and $\min_{(i,j)\in\mathbb{E}^P}\min_{1\leq t\leq n}|\omega_{ij}(\tau_t)|\gg\lambda_3$, we have

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^{P,\dagger}=\mathbb{E}_n^{P}\right)\to 1. \tag{2.5.12}$$

2.6 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 *et al.* (2015a). 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 (2.3.3) is approximately $2nh \approx 140$, which is smaller than the combined number of unknown coefficient functions and their derivatives, $2d = 200$. 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 B.5 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 B.5. 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,

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

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

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

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

Example 2.1. 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 2.1 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 2.2 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

Table 2.1: Transition matrix and Granger network estimation in Example 2.1.

		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.

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 a negligible impact on the precision matrix and partial correlation network estimation.

Example 2.2. 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 2.3 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 2.1, which affects the network estimation accuracy. When $d = 100$ and $n = 200$,

Table 2.2: Precision matrix and partial correlation network estimation in Example 2.1.

		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

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 1.1, 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 2.4 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 2.3 that the VAR error approximation is poorer than that in Example 2.1. 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 2.3. The data is generated from a VAR(1) model with $\mathbf{A}_1(\tau) = [a_{ij}(\tau)]_{d\times d}$ being a Toeplitz matrix and $a_{ij}(\tau) = (0.4 - 0.1\tau)^{|i-j|+1}$. 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 2.5 reports the estimation errors of the various methods considered. In this example, the

Table 2.3: Transition matrix and Granger network estimation in Example 2.2.

		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 2.4: Precision matrix and partial correlation network estimation in Example 2.2.

		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 2.5: Estimation accuracy of dual networks in Example 2.3.

		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

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 2.1. 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 2.4. The data is generated from a factor-adjusted time-varying VAR model in the form of (2.5.2). The idiosyncratic errors of the time-varying factor model are generated from a VAR(1) model in Example 1.2. The two factors in $F_t = (F_{t,1}, F_{t,2})^\intercal$ are generated from two univariate 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 2.6 reports the estimation results of the time-varying transition matrices and Granger networks for the idiosyncratic errors, and Table 2.7 reports the estimation results of the time-varying precision matrices and partial correlation networks. Comparing with the results in Tables 2.3 and 2.4, 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

Table 2.6: Factor-adjusted transition matrix and Granger network estimation in Example 2.4.

		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 2.7: Factor-adjusted precision matrix and partial correlation network estimation in Example 2.4.

		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

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 ($n = 400$).

2.7 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. A 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

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

missing values. Each variable is standardised to have zero mean and unit variance. We consider the two factor modelling methods in Section 2.5 to accommodate strong cross-sectional dependence: the approximate factor model (2.5.1) with constant factor loadings, and the time-varying factor model (2.5.2) 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 B.5 for a description of the criteria). Seven factors are selected for the factor model with constant loadings, whereas only four are selected for the time-varying factor model. Since the latter provides a more parsimonious model specification, we hereafter report network estimation results only for this model. The estimated idiosyncratic errors, denoted as $\widehat{x}_{t,i}$, $i = 1, \ldots, 127$, $t = 1, \ldots, 763$, are then used for our empirical analysis. Miao *et al.* (2023) 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 B.5 for detail) and subsequently select the time-varying VAR(1) model for $\widehat{X}_t = (\widehat{x}_{t,1}, \ldots, \widehat{x}_{t,127})^\intercal$.

Figure 2.1 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 2.2 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

[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.

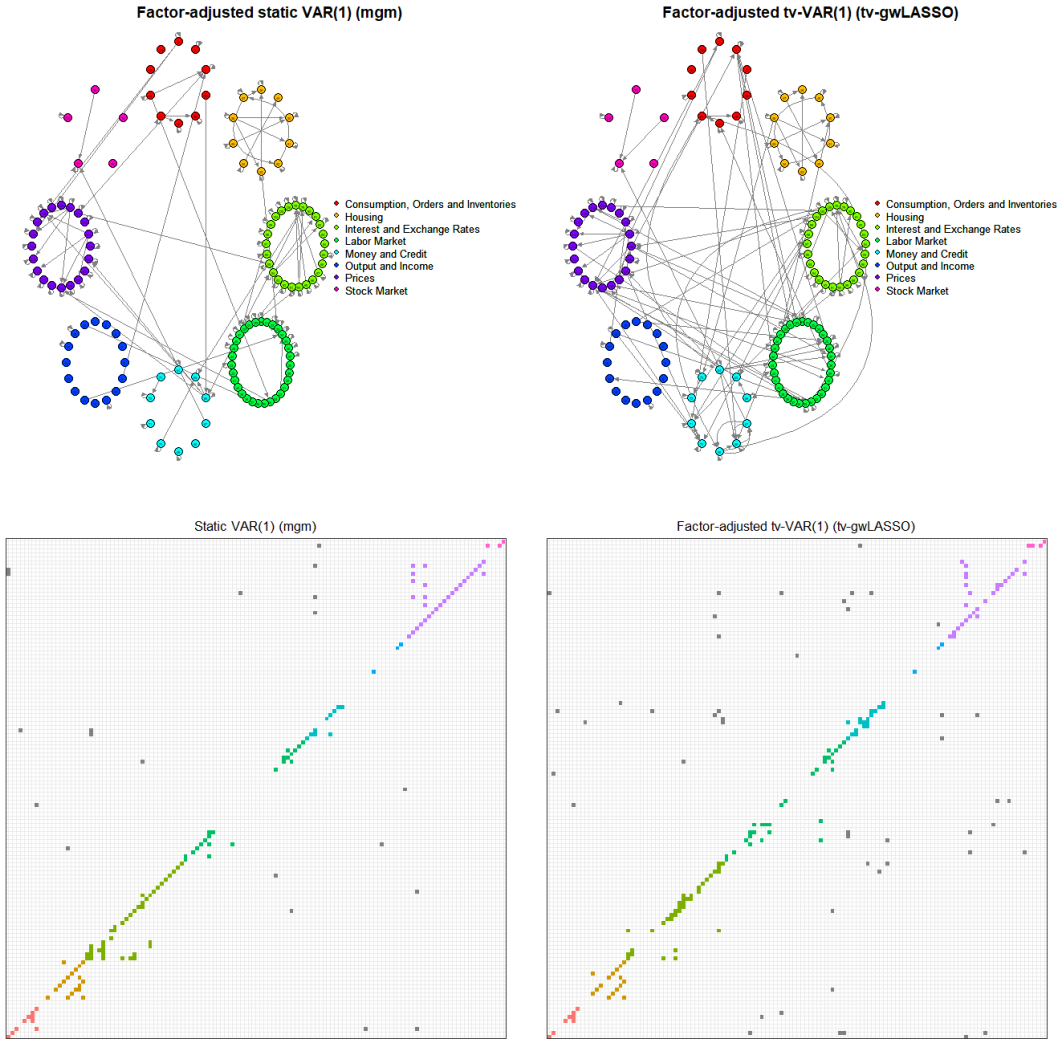


Figure 2.1: The estimated Granger causality networks using the factor-adjusted static VAR(1) model (left) and time-varying VAR(1) model (right).

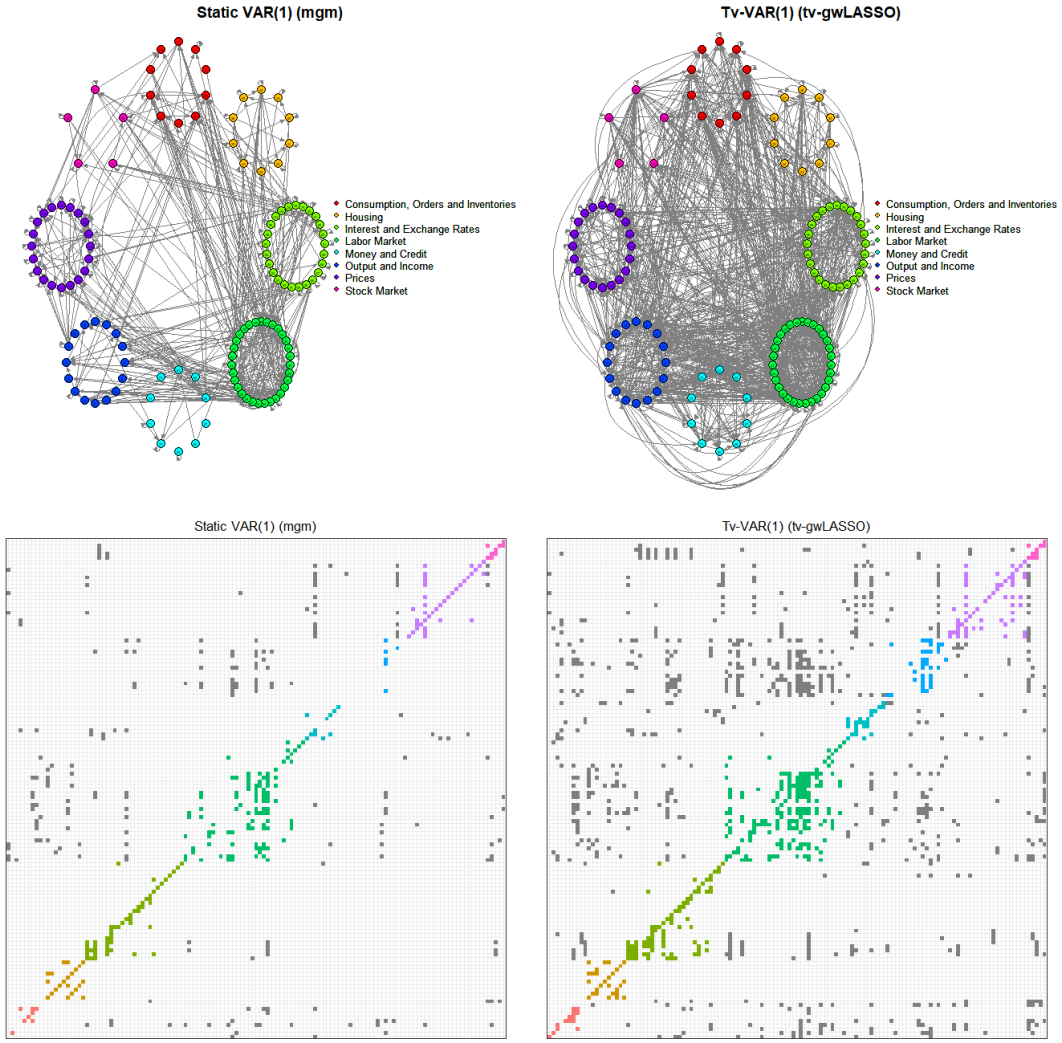


Figure 2.2: The estimated Granger causality networks using the static VAR(1) model (left) and time-varying VAR(1) model (right) without factor-adjustment.

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 2.3 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 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.

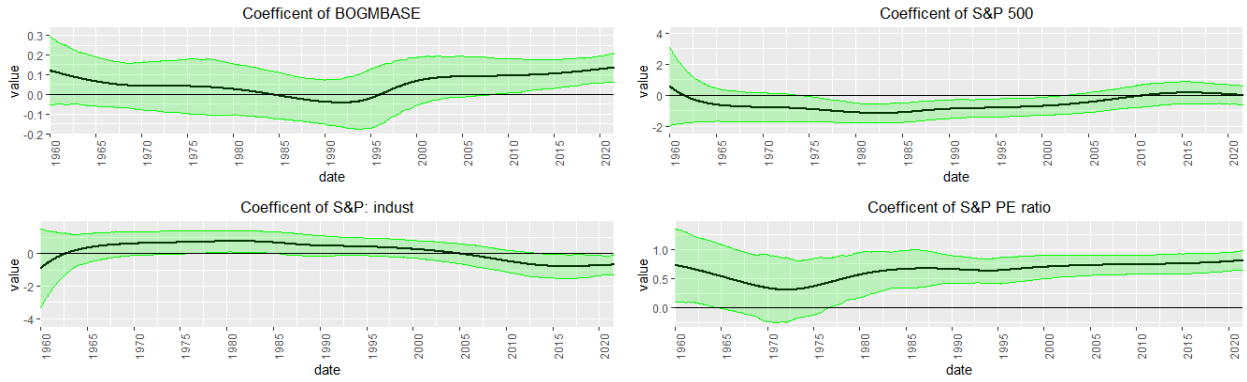


Figure 2.3: The estimated time-varying coefficients linked to S&P PE ratio with 90% confident bands.

We plot the estimated partial correlation networks in Figure 2.4, 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 (Janková and van de Geer, 2015), and draw the 90% confidence bands using R package "SILGGM" in Figure 2.5. 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 2.5 suggest that time-invariant partial correlation linkages are inappropriate to describe the network structure of the FRED-MD data.

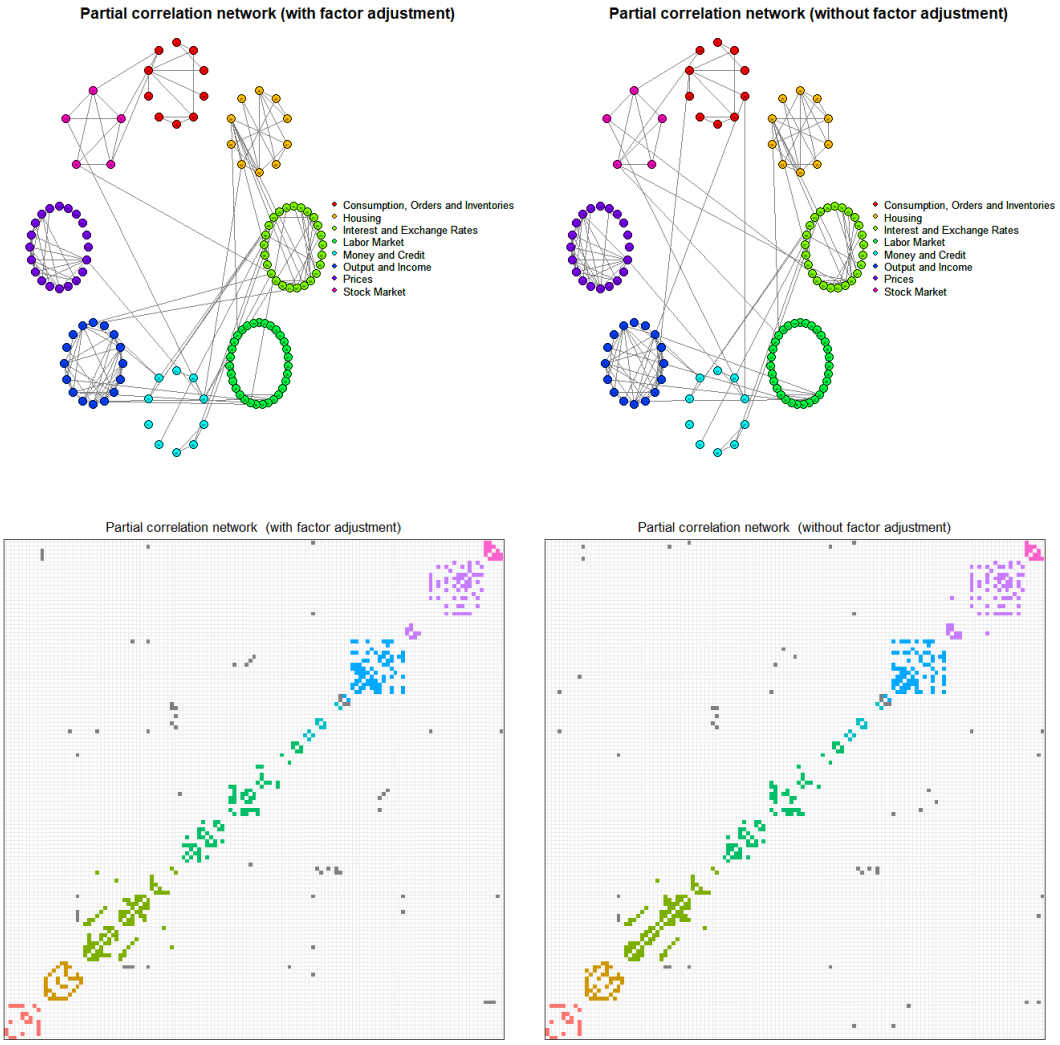


Figure 2.4: The estimated partial correlation networks with (left) and without (right) factor adjustment.

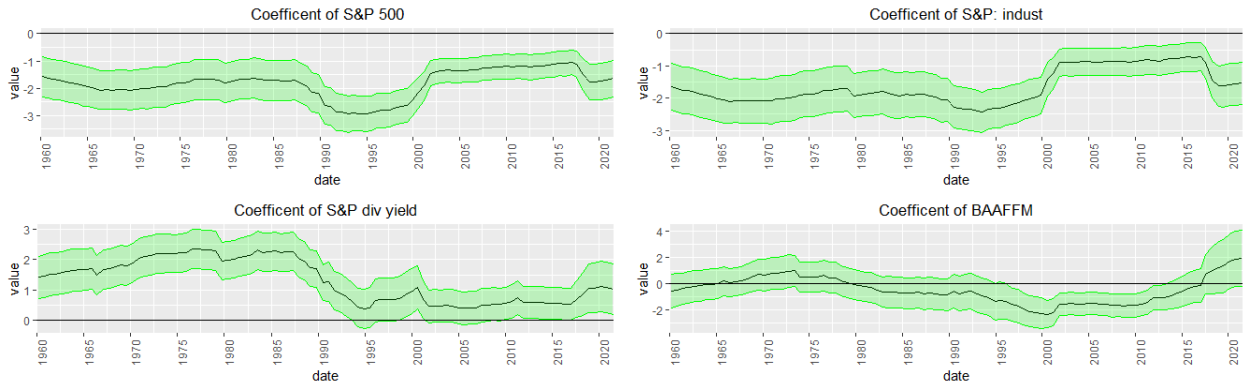


Figure 2.5: The estimated time-varying elements in the precision matrix linked to S&P PE ratio with 90% confident bands.

2.8 Conclusion

In this chapter, 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 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.

Chapter 3

Estimation of Large Dynamic Precision Matrices with a Latent Semiparametric Structure

Abstract We estimate large dynamic precision matrices for high-dimensional time series data where the conditioning random variables are multivariate. To overcome the challenges posed by the curse of dimensionality, we introduce the approximate factor structure and employ the semiparametric MAMAR approximation to estimate the underlying dynamic covariance matrix of the factors and the idiosyncratic components. By using the Sherman-Morrison-Woodbury formula, we obtain the dynamic precision matrix for the time series. Under some mild conditions such as the approximate sparsity assumption, the proposed precision matrix estimation is proved to be uniformly consistent. The simulation highlights the advantage of utilising the factor structure when estimating large dynamic precision matrices. In the empirical analysis, we apply the proposed method to the returns of S&P 500 constituents. The results indicate that our method performs well in the portfolio selection problem.

Keywords: Approximate factor model, Conditional sparsity, Large precision matrix, MAMAR, Semiparametric estimation.

3.1 Introduction

There has been increasing interest in recent decades on estimation of large precision matrices, which have applications in various fields including discriminant analysis, network or graphical model estimation and optimal portfolio choice. Existing literature often assumes that the precision matrix of a high-dimensional random vector (with dimension possibly much larger than the sample size) is static, satisfying an approximate sparsity condition similar to that often imposed on large covariance matrices (e.g., Bickel and Levina, 2008), and then uses various techniques, such as penalised likelihood (Lam and Fan, 2009), graphical Danzig selector (Yuan, 2010) and constrained ℓ_1-minimisation for inverse matrix estimation (CLIME) (Cai *et al.*, 2011), to estimate it. Under a high-dimensional semiparametric Gaussian copula model framework, Liu *et al.* (2012) and Xue and Zou (2012) estimate the inverse of a correlation matrix by combining Spearman's rho or Kendall's tau with the aforementioned techniques for large precision matrices. A comprehensive review of recent developments in large precision matrix estimation can be found in Cai *et al.* (2016) and Fan *et al.* (2016c).

The static and sparsity assumptions for precision matrices may be too restrictive to be realistic for many practical applications. When considering large precision matrices over a long time span, the static assumption is likely to be violated and hence, it is important to explore their dynamic/time-evolving pattern to avoid misleading results from subsequent analysis. The sparsity assumption is also often violated in reality. Hence, there have been some attempts to relax either of the assumptions in recent years. For example, Kolar *et al.* (2010) and Zhou *et al.* (2010) relax the static assumption by allowing smooth time-varying changes in large precision matrices of dynamic network or graphical models, where observations are assumed to be either serially independent or stationary weakly dependent; Qiu *et al.* (2016) estimate a large precision matrix which depends on a subject-specific variable; and Xu *et al.* (2020) test structural breaks on a large precision matrix and then estimate its smooth time-varying structure between breaks. For the sparsity assumption, a popular approach to circumvent it is to make use of the fact that many economic and financial time series variables often exhibit co-movements, possibly driven by some latent factors. Chandrasekaran *et al.* (2012) is among the first to study a large precision matrix with latent variables involved. They provide the graphical model identification conditions and propose a penalised likelihood estimation method under the joint Gaussian assumption. By decomposing the large precision matrix into a "low-rank plus sparse" structure under an approximate factor model assumption, Wu *et al.* (2017) and Tang *et al.* (2020) introduce the IPOD (Inverting Principal Orthogonal Decomposition) and LVD (Latent Variables graphical models via ℓ_1 and penalised D-trace loss) approaches, respectively. A similar technique is also used by Cai *et al.* (2020) to estimate large precision matrices for high-dimensional and high-frequency financial data.

In this chapter, we aim to estimate a large dynamic precision matrix with a latent factor structure, avoiding both the static and the sparsity assumptions. Specifically, suppose that $\mathbf{X}_t = (X_{t,1}, \ldots, X_{t,N})^\intercal$ is an N-dimensional random vector generated from the following approximate factor model:

$$\mathbf{X}_t = \boldsymbol{\chi}_t + \boldsymbol{\varepsilon}_t \quad \text{with} \quad \boldsymbol{\chi}_t = \boldsymbol{\Lambda}\mathbf{F}_t, \quad t = 1, \ldots, T, \tag{3.1.1}$$

where $\boldsymbol{\Lambda} = (\lambda_{ij})_{N\times K}$ is an $N \times K$ matrix of factor loadings, $\mathbf{F}_t = (F_{t,1}, \ldots, F_{t,K})^\intercal$ is a K-dimensional vector of latent factors, and $\boldsymbol{\epsilon}_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,N})^\intercal$ is an N-dimensional vector of idiosyncratic errors.

The approximate factor model has become an effective tool in analysing high-dimensional economic and financial time series (e.g., Chamberlain and Rothschild, 1983; Fama and French, 1992; Stock and Watson, 2002; Bai and Ng, 2002). As in the large panel literature, we consider the setting that both N and T diverge to infinity, making it feasible to consistently estimate the factor space and factor loadings (up to an appropriate rotation). Let $\mathbf{U}_t = (U_{t,1}, \cdots, U_{t,d})^\intercal$ be a d-dimensional vector of conditioning variables which may be chosen as the lagged terms of some components of $\mathbf{X}_t$ or some low-dimensional observed factors (such as the Fama-French three factors). We assume that K and d are fixed, $\mathbf{\Lambda}$ is deterministic, and $\mathbf{F}_t$ and $\boldsymbol{\varepsilon}_t$ are conditionally uncorrelated given the past conditioning variables $\mathbf{U}_s$, $s \leq t-1$. The dynamic covariance matrix of $\mathbf{X}_t$ is defined and computed as

$$\mathbf{\Sigma}_X(\mathbf{u}) = \mathsf{Var}\left(\mathbf{X}_{t+1}|\mathbf{U}_t = \mathbf{u}\right) = \mathbf{\Lambda}\mathbf{\Sigma}_F(\mathbf{u})\mathbf{\Lambda}^\intercal + \mathbf{\Sigma}_\varepsilon(\mathbf{u}), \tag{3.1.2}$$

where $\mathbf{\Sigma}_F(\mathbf{u}) = \mathsf{Var}\left(\mathbf{F}_{t+1}|\mathbf{U}_t = \mathbf{u}\right)$ and $\mathbf{\Sigma}_\varepsilon(\mathbf{u}) = \mathsf{Var}\left(\boldsymbol{\varepsilon}_{t+1}|\mathbf{U}_t = \mathbf{u}\right)$. It is worth noting that $\mathbf{\Sigma}_F(\mathbf{u})$ is a $K \times K$ dynamic covariance matrix, whereas $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$ is a large covariance matrix with size $N \times N$. As in Fan *et al.* (2013) and Wang *et al.* (2021b), we impose a sparsity assumption on $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$, resulting in a "low-rank plus sparse" or conditionally sparse structure in (3.1.2), which makes it possible to develop sensible estimation theory. Also note that it is not unreasonable to assume that $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$ is sparse, as all the systematic and common pattern in $\mathbf{X}_t$ should have been accounted for by $\boldsymbol{\chi}_t$, leaving only individual specific errors in $\boldsymbol{\varepsilon}_t$.

The primary interest of this chapter is to estimate the dynamic precision matrix, denoted as $\mathbf{\Omega}_X(\mathbf{u})$, which is the inverse of $\mathbf{\Sigma}_X(\mathbf{u})$. Letting $\mathbf{\Lambda}_F(\mathbf{u}) = \mathbf{\Lambda}\mathbf{\Sigma}_F^{1/2}(\mathbf{u})$, by (3.1.2) and the Sherman-Morrison-Woodbury formula, we readily have

$$\mathbf{\Omega}_X(\mathbf{u}) = \mathbf{\Omega}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}_\varepsilon(\mathbf{u})\mathbf{\Lambda}_F(\mathbf{u})\left[\mathbf{I}_K + \mathbf{\Lambda}_F(\mathbf{u})^\intercal\mathbf{\Omega}_\varepsilon(\mathbf{u})\mathbf{\Lambda}_F(\mathbf{u})\right]^{-1}\mathbf{\Lambda}_F(\mathbf{u})^\intercal\mathbf{\Omega}_\varepsilon(\mathbf{u}), \tag{3.1.3}$$

where $\mathbf{\Omega}_\varepsilon(\mathbf{u}) = \mathbf{\Sigma}_\varepsilon^{-1}(\mathbf{u})$ and $\mathbf{I}_K$ is the $K \times K$ identity matrix. As in Wu *et al.* (2017) and Tang *et al.* (2020), we may impose a sparsity assumption on $\mathbf{\Omega}_\varepsilon(\mathbf{u})$, which leads to a "low-rank plus sparse" structure for the precision matrix, $\mathbf{\Omega}_X(\mathbf{u})$.

The rest of this chapter is organised as follows. Section 3.2 first introduces a semiparametric Model Averaging MArginal Regression (MAMAR) approximation for $\mathbf{\Omega}_X(\mathbf{u})$ and then discusses the estimation of this MAMAR approximation. It also discusses the construction of minimum-variance portfolios using the MAMAR estimates of dynamic precision matrices. Section 3.3 provides uniform consistency results for the MAMAR estimators of the precision matrices. Sections 3.4 and 3.5 present Monte-Carlo simulation results and an empirical application showcasing the usefulness of our proposed method. Section 3.6 concludes. Proofs of the asymptotic results in Section 3.3 are relegated to Appendix C. Throughout this chapter, we use $\|\cdot\|$ to denote the Euclidean norm of a vector; and $\lambda_{\max}(\cdot)$, $\lambda_{\min}(\cdot)$ and $\mathsf{Tr}(\cdot)$ to denote the maximum eigenvalue, minimum eigenvalue and trace of a square matrix, respectively. For a $p \times p$ matrix $\mathbf{W} = (w_{ij})_{p\times p}$, we let $\|\mathbf{W}\|_O = \lambda_{\max}^{1/2}\left(\mathbf{W}^\intercal\mathbf{W}\right)$ be its operator (or spectral) norm, $\|\mathbf{W}\|_F = \mathsf{Tr}^{1/2}\left(\mathbf{W}^\intercal\mathbf{W}\right)$ its Frobenius norm, $|\mathbf{W}|_1 = \sum_{i=1}^p\sum_{j=1}^p |w_{ij}|$, $\|\mathbf{W}\|_1 = \max_{1\leq j\leq p}\sum_{i=1}^p |w_{ij}|$, $\|\mathbf{W}\|_\infty = \max_{1\leq i\leq p}\sum_{j=1}^p |w_{ij}|$, and $\|\mathbf{W}\|_{\max} = \max_{1\leq i\leq p}\max_{1\leq j\leq p}|w_{ij}|$.

3.2 Methodology

3.2.1 Semiparametric MAMAR approximation

It is clear from the decomposition (3.1.3) that two key dynamic components of $\mathbf{\Omega}_X(\mathbf{u})$ are $\mathbf{\Omega}_\varepsilon(\mathbf{u}) = \mathbf{\Sigma}_\varepsilon^{-1}(\mathbf{u})$ and $\mathbf{\Sigma}_F(\mathbf{u})$. Let $\sigma_{\varepsilon,ij}(\mathbf{u})$, $1 \leq i,j \leq N$, and $\sigma_{F,ij}(\mathbf{u})$, $1 \leq i,j \leq K$, be the (i,j)-entry of the matrices $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$ and $\mathbf{\Sigma}_F(\mathbf{u})$, respectively. Throughout this chapter, we do not impose any pre-specified parametric form on either $\sigma_{\varepsilon,ij}(\mathbf{u})$ or $\sigma_{F,ij}(\mathbf{u})$. As $\mathbf{U}_t$ is assumed to be a multivariate vector of conditioning variables, a direct application of classic nonparametric methods to estimate $\sigma_{\varepsilon,ij}(\mathbf{u})$ and $\sigma_{F,ij}(\mathbf{u})$ would suffer from the "curse of dimensionality" issue when the dimension of $\mathbf{u}$, i.e., d, is larger than three. To address this problem, we use a semiparametric approximation via Model Averaging MArginal Regression (MAMAR). It is detailed below.

When $\mathsf{E}(\varepsilon_{t+1,i}|\mathbf{U}_t) = 0$, $1 \leq i \leq N$, the MAMAR approximation for $\sigma_{\varepsilon,ij}(\mathbf{u}) = \mathsf{E}(\varepsilon_{t+1,i}\varepsilon_{t+1,j}|\mathbf{U}_t = \mathbf{u})$, $1 \leq i,j \leq N$, is defined as

$$\sigma_{\varepsilon,ij}(\mathbf{u}) \approx a_{0,ij} + \sum_{k=1}^{d} a_{k,ij}\mathsf{E}(\varepsilon_{t+1,i}\varepsilon_{t+1,j}|U_{t,k} = u_k) =: a_{0,ij} + \sum_{k=1}^{d} a_{k,ij}\sigma_{\varepsilon,k,ij}(u_k), \tag{3.2.1}$$

where $a_{k,ij}$, $0 \leq k \leq d$, are unknown weights and $\sigma_{\varepsilon,k,ij}(u_k)$, $1 \leq k \leq d$, are univariate nonparametric functions which can be easily estimated from commonly-used nonparametric methods (such as kernel smoothing) without incurring the "curse of dimensionality". The MAMAR approximation is first introduced by Li *et al.* (2015b) in the context of semiparametric time series regression estimation and forecasting, and is further generalised by Chen *et al.* (2018) to the ultra large time series regression setting. This idea has been applied to semiparametric dynamic portfolio choice (Chen *et al.*, 2016), high-dimensional classification (Fan *et al.*, 2016a) and high-dimensional dynamic covariance matrix estimation (Chen *et al.*, 2019). To write (3.2.1) in matrix form, denote $\mathbf{A}_k = (a_{k,ij})_{N\times N}$, $k = 0,1,\ldots,d$, and $\mathbf{\Sigma}_{\varepsilon,k}(u_k) = [\sigma_{\varepsilon,k,ij}(u_k)]_{N\times N}$, $k = 1,\ldots,d$. Then, the MAMAR approximation for $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$ can be written as

$$\mathbf{\Sigma}_\varepsilon(\mathbf{u}) \approx \mathbf{A}_0 + \sum_{k=1}^{d} \mathbf{A}_k \odot \mathbf{\Sigma}_{\varepsilon,k}(u_k), \tag{3.2.2}$$

where $\odot$ denotes the Hadamard product.

The weighting parameters $a_{k,ij}$ play a crucial role in the MAMAR approximation. We next derive the theoretically optimal weights for the optimal MAMAR approximation of $\mathbf{\Sigma}_\varepsilon(\mathbf{u})$. For $1 \leq i,j \leq N$, the optimal weights $a_{k,ij}^o$, $k = 0,1,\ldots,d$, are obtained by minimising

$$Q(a_{0,ij}, a_{1,ij}, \ldots, a_{d,ij}) = \mathsf{E}\left[\varepsilon_{t+1,i}\varepsilon_{t+1,j} - a_{0,ij} - \sum_{k=1}^{d} a_{k,ij}\sigma_{\varepsilon,k,ij}(U_{t,k})\right]^2$$

with respect to $a_{k,ij}$, $k = 0,1,\ldots,d$. Following standard calculations as in Li *et al.* (2015b), we have

the following solution to the above minimisation problem:

$$\left(a^o_{1,ij},\ldots,a^o_{d,ij}\right)^\intercal := \boldsymbol{\Delta}^{*-1}_{\varepsilon,ij}\mathbf{W}^*_{\varepsilon,ij}, \qquad a^o_{0,ij} := \left(1-\sum_{k=1}^{d} a^o_{k,ij}\right)\mathsf{E}(\varepsilon_{t,i}\varepsilon_{t,j}), \tag{3.2.3}$$

where $\boldsymbol{\Delta}^*_{\varepsilon,ij}$ is a $d\times d$ matrix with the (k,l)-entry being $\delta^*_{\varepsilon,ij,kl} = \mathsf{Cov}\left[\sigma_{\varepsilon,k,ij}(U_{t,k}),\sigma_{\varepsilon,l,ij}(U_{t,l})\right]$, and $\mathbf{W}^*_{\varepsilon,ij}$ is a d-dimensional vector with the k-th element being $w^*_{\varepsilon,ij,k} = \mathsf{Cov}\left[\sigma_{\varepsilon,k,ij}(U_{t,k}),\varepsilon_{t+1,i}\varepsilon_{t+1,j}\right] = \mathsf{Var}\left[\sigma_{\varepsilon,k,ij}(U_{t,k})\right]$. Letting $\mathbf{A}^o_k = \left(a^o_{k,ij}\right)_{N\times N}$ and replacing $\mathbf{A}_k$ by $\mathbf{A}^o_k$ in (3.2.2), we obtain the (theoretically) optimal MAMAR approximation to $\boldsymbol{\Sigma}_\varepsilon(\mathbf{u})$:

$$\boldsymbol{\Sigma}^o_\varepsilon(\mathbf{u}) := \mathbf{A}^o_0 + \sum_{k=1}^{d}\mathbf{A}^o_k \odot \boldsymbol{\Sigma}_{\varepsilon,k}(u_k). \tag{3.2.4}$$

Consequently, taking the inverse of $\boldsymbol{\Sigma}^o_\varepsilon(\mathbf{u})$, we have the following optimal MAMAR approximation to the error precision matrix, $\boldsymbol{\Omega}_\varepsilon(\mathbf{u})$,

$$\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u}) := \left[\mathbf{A}^o_0 + \sum_{k=1}^{d}\mathbf{A}^o_k \odot \boldsymbol{\Sigma}_{\varepsilon,k}(u_k)\right]^{-1}. \tag{3.2.5}$$

When $\mathsf{E}(\varepsilon_{t+1,i}|\mathbf{U}_t)\neq 0$, $1\leq i\leq N$, note that

$$\boldsymbol{\Sigma}_\varepsilon(\mathbf{u}) = \mathsf{E}\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}^\intercal_{t+1}|\mathbf{U}_t=\mathbf{u}\right) - \mathsf{E}\left(\boldsymbol{\varepsilon}_{t+1}|\mathbf{U}_t=\mathbf{u}\right)\left[\mathsf{E}\left(\boldsymbol{\varepsilon}_{t+1}|\mathbf{U}_t=\mathbf{u}\right)\right]^\intercal =: \mathbf{C}_\varepsilon(\mathbf{u}) - \mathbf{M}_\varepsilon(\mathbf{u})\mathbf{M}_\varepsilon(\mathbf{u})^\intercal.$$

We can similarly apply MAMAR to both $\mathbf{C}_F(\mathbf{u})$ and $\mathbf{M}_F(\mathbf{u})$ and obtain their approximations, denoted as $\mathbf{C}^o_F(\mathbf{u})$ and $\mathbf{M}^o_F(\mathbf{u})$, respectively, and then obtain the optimal MAMAR approximation to $\boldsymbol{\Sigma}_\varepsilon(\mathbf{u})$:

$$\boldsymbol{\Sigma}^o_\varepsilon(\mathbf{u}) := \mathbf{C}^o_\varepsilon(\mathbf{u}) - \mathbf{M}^o_\varepsilon(\mathbf{u})\mathbf{M}^o_\varepsilon(\mathbf{u})^\intercal.$$

For the component $\boldsymbol{\Sigma}_F(\mathbf{u})$, we can similarly write

$$\boldsymbol{\Sigma}_F(\mathbf{u}) = \mathsf{E}\left(\mathbf{F}_{t+1}\mathbf{F}^\intercal_{t+1}|\mathbf{U}_t=\mathbf{u}\right) - \mathsf{E}\left(\mathbf{F}_{t+1}|\mathbf{U}_t=\mathbf{u}\right)\left[\mathsf{E}\left(\mathbf{F}_{t+1}|\mathbf{U}_t=\mathbf{u}\right)\right]^\intercal =: \mathbf{C}_F(\mathbf{u}) - \mathbf{M}_F(\mathbf{u})\mathbf{M}_F(\mathbf{u})^\intercal.$$

We can apply MAMAR to both $\mathbf{C}_F(\mathbf{u})$ and $\mathbf{M}_F(\mathbf{u})$ and obtain their approximations, denoted as $\mathbf{C}^o_F(\mathbf{u})$ and $\mathbf{M}^o_F(\mathbf{u})$, respectively, and then obtain the optimal MAMAR approximation to $\boldsymbol{\Sigma}_F(\mathbf{u})$:

$$\boldsymbol{\Sigma}^o_F(\mathbf{u}) := \mathbf{C}^o_F(\mathbf{u}) - \mathbf{M}^o_F(\mathbf{u})\mathbf{M}^o_F(\mathbf{u})^\intercal. \tag{3.2.6}$$

Letting $\boldsymbol{\Lambda}^o_F(\mathbf{u}) = \boldsymbol{\Lambda}\left[\boldsymbol{\Sigma}^o_F(\mathbf{u})\right]^{1/2}$, by virtue of (3.1.3), we obtain

$$\boldsymbol{\Omega}^o_X(\mathbf{u}) := \boldsymbol{\Omega}^o_\varepsilon(\mathbf{u}) - \boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\boldsymbol{\Lambda}^o_F(\mathbf{u})\left[\mathbf{I}_K + \boldsymbol{\Lambda}^o_F(\mathbf{u})^\intercal\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\boldsymbol{\Lambda}^o_F(\mathbf{u})\right]^{-1}\boldsymbol{\Lambda}^o_F(\mathbf{u})^\intercal\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u}), \tag{3.2.7}$$

which is the theoretically optimal MAMAR approximation to $\boldsymbol{\Omega}_X(\mathbf{u})$.

Our main interest lies in estimating $\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})$ and $\boldsymbol{\Omega}^o_X(\mathbf{u})$, which can be seen as "proxies" for $\boldsymbol{\Omega}_\varepsilon(\mathbf{u})$ and $\boldsymbol{\Omega}_X(\mathbf{u})$, respectively. Chen *et al.* (2019) use a similar MAMAR approximation technique for large dynamic covariance matrix estimation under a sparsity assumption. To relax the sparsity

assumption, they also discuss, although very briefly, the estimation of the MAMAR approximation for $\mathbf{\Sigma}_X(\mathbf{u})$ under the factor structure (3.1.2), without providing rigorous theoretical derivation. In this chapter, we focus on large precision matrices rather than covariance matrices. In order to derive sensible estimation theory, we assume that $\mathbf{\Sigma}_F^o(\mathbf{u}) \succ 0$ and $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$ satisfies the uniform sparsity assumption, i.e., $\{\mathbf{\Omega}_\varepsilon^o(\mathbf{u}) : \mathbf{u} \in \mathscr{U}\} \subseteq \mathscr{S}(q, \varpi_N, M, \mathscr{U})$, where $\mathscr{S}(q, \varpi_N, M, \mathscr{U})$ is defined as

$$\left\{ \mathbf{\Omega}(\mathbf{u}) = [\omega_{ij}(\mathbf{u})]_{N\times N}, \mathbf{u} \in \mathscr{U} \middle| \ \mathbf{\Omega}(\mathbf{u}) \succ 0, \ \sup_{\mathbf{u}\in\mathscr{U}} \|\mathbf{\Omega}(\mathbf{u})\|_1 \leq M, \ \sup_{\mathbf{u}\in\mathscr{U}} \max_{1\leq i\leq N} \sum_{j=1}^{N} |\omega_{ij}(\mathbf{u})|^q \leq \varpi_N \right\}, \tag{3.2.8}$$

for some $0 \leq q < 1$, $0 < M < \infty$, and $\varpi_N > 0$ (which may depend on N). The notation "$\mathbf{\Omega} \succ 0$" denotes that $\mathbf{\Omega}$ is positive definite.

3.2.2 Factor model estimation

Before developing a feasible estimation procedure for $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$ and $\mathbf{\Omega}_X^o(\mathbf{u})$, we need to estimate the latent components in the approximate factor model (3.1.1), i.e., the factor loadings matrix $\mathbf{\Lambda}$, common factors $\mathbf{F}_t$, as well as the idiosyncratic errors $\boldsymbol{\varepsilon}_t$. We will use the principal component analysis (PCA) technique, which has been commonly used for factor model estimation (e.g., Bai and Ng, 2002; Stock and Watson, 2002; Fan *et al.*, 2013).

We first assume that the number of factors, K, is known, a priori, and will discuss its selection later. Letting $\mathbb{X}_{N,T} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)^\intercal$ be the $T \times N$ matrix of observations and conducting an eigenanalysis on the $T \times T$ matrix $\left(\mathbb{X}_{N,T}\mathbb{X}_{N,T}^\intercal\right)/(NT)$, we obtain a $T \times K$ matrix $\widehat{\mathbb{F}} = \left(\widehat{\mathbf{F}}_1, \ldots, \widehat{\mathbf{F}}_T\right)^\intercal$, whose columns are the K eigenvectors (multiplied by $T^{1/2}$) corresponding to the K largest eigenvalues. Replacing $\mathbf{F}_t$ with $\widehat{\mathbf{F}}_t$ in (3.1.1) and applying the least squares estimation, we obtain, using the normalisation restriction on $\widehat{\mathbb{F}}$, the following estimate of factor loadings matrix: $\widehat{\mathbf{\Lambda}} = \left(\widehat{\boldsymbol{\lambda}}_1, \ldots, \widehat{\boldsymbol{\lambda}}_N\right)^\intercal = \mathbb{X}_{N,T}^\intercal \widehat{\mathbb{F}}/T$. Finally, the idiosyncratic error $\boldsymbol{\varepsilon}_t$ can be approximated by $\widehat{\boldsymbol{\varepsilon}}_t = (\widehat{\varepsilon}_{t,1}, \ldots, \widehat{\varepsilon}_{t,N})^\intercal = \mathbf{X}_t - \widehat{\mathbf{\Lambda}}\widehat{F}_t$, $t = 1, \ldots, T$.

We next give some regularity conditions which are sufficient for deriving the uniform consistency results for $\widehat{\mathbf{F}}_t$, $\widehat{\boldsymbol{\lambda}}_i$, and $\widehat{\varepsilon}_{t,i}$.

Assumption 3.A. *(i) The process $\{(\mathbf{F}_t^\intercal, \boldsymbol{\varepsilon}_t^\intercal)^\intercal\}_{t=1}^\infty$ is stationary and α-mixing with the mixing coefficient α_t satisfying $\alpha_t = O\left(\theta^t\right)$, where θ is a constant that satisfies $0 < \theta < 1$.*

(ii) The $K \times K$ matrix $\frac{1}{N}\mathbf{\Lambda}^\intercal\mathbf{\Lambda}$ is positive definite with its smallest eigenvalue bounded away from zero, and $\|\boldsymbol{\lambda}_i\|$ is uniformly bounded over $1 \leq i \leq N$, where $\boldsymbol{\lambda}_i$ is the i-th column of $\mathbf{\Lambda}^\intercal$.

(iii) The covariance matrix $\mathsf{Var}(\mathbf{F}_t)$ is positive definite. In addition, there exists a constant $c_1 > 0$ such that $\mathsf{E}\left[\exp\left(c_1\|\mathbf{F}_t\|^2\right)\right] < \infty$.

(iv) The idiosyncratic errors satisfy

$$\mathsf{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}, \quad \mathsf{E}[\varepsilon_{t,i}\mathbf{F}_t] = \mathbf{0}, \quad \text{and} \quad \max_{1\leq i\leq N} \max_{1\leq t\leq T} \mathsf{E}\left[\exp\left(c_1\varepsilon_{t,i}^2\right)\right] < \infty,$$

where c_1 is defined in (iii).

(v) There exist $0 < c_2 < \infty$ *and* $\delta > 2$ *such that*

$$\max_{1\leq t\leq T} \mathsf{E}\left[\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i\varepsilon_{t,i}\right\|^{\delta}\right] \leq c_2 N^{\delta/2}, \quad \max_{1\leq s,t\leq T} \mathsf{E}\left[\left|\sum_{i=1}^{N}\left[\varepsilon_{s,i}\varepsilon_{t,i} - \mathsf{E}\left(\varepsilon_{s,i}\varepsilon_{t,i}\right)\right]\right|^{\delta}\right] \leq c_2 N^{\delta/2}.$$

Most of the above assumptions are standard in the approximate factor model estimation theory (e.g., Bai and Ng, 2002; Fan *et al.*, 2013; Chen *et al.*, 2018). The sub-Gaussian moment conditions in Assumption 3.A(iii) and (iv) are required to cover the ultra-high dimensional case where N diverges at an exponential rate of T, and can be weakened if N diverges at a polynomial rate of T.

Define a $K \times K$ rotation matrix

$$\mathbf{R} := \boldsymbol{\Lambda}_K^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{F}}_t\mathbf{F}_t^{\intercal}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{\intercal}\right), \tag{3.2.9}$$

where $\boldsymbol{\Lambda}_K$ is a $K \times K$ diagonal matrix with the first K largest eigenvalues of $\mathbb{X}_{N,T}\mathbb{X}_{NT}^{\intercal}/(NT)$ (arranged in descending order) being the diagonal elements. The following proposition gives the uniform consistency results for $\widehat{\mathbf{F}}_t$, $\widehat{\boldsymbol{\lambda}}_i$, and $\widehat{\varepsilon}_{t,i}$, which are comparable to those obtained in the existing literature (e.g., Bai and Ng, 2002; Fan *et al.*, 2013; Chen *et al.*, 2018; Li *et al.*, 2023).

Proposition 3.2.1. *Suppose that Assumption 3.A is satisfied,* $N \gg T^{4/\delta}$ *with* δ *defined in Assumption 3.A(v), and* $N = O\left(\exp\{T^{\nu}\}\right)$ *with* $0 < \nu < 1/5$*. Then, we have the following uniform consistency results:*

(i)

$$\max_{1\leq t\leq T}\left\|\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right\| = O_P\left(\frac{1}{T^{1/2}} + \frac{T^{2/\delta}}{N^{1/2}}\right); \tag{3.2.10}$$

(ii)

$$\max_{1\leq i\leq N}\left\|\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right\| = O_P\left(\left(\frac{\log N}{T}\right)^{1/2} + \frac{T^{2/\delta}}{N^{1/2}}\right); \tag{3.2.11}$$

(iii)

$$\max_{1\leq i\leq N}\max_{1\leq t\leq T}|\widehat{\varepsilon}_{t,i} - \varepsilon_{t,i}| = O_P\left((\log T)^{1/2}\left[\left(\frac{\log N}{T}\right)^{1/2} + \frac{T^{2/\delta}}{N^{1/2}}\right]\right), \tag{3.2.12}$$

where δ *is defined in Assumption 3.A(v).*

Proposition 3.2.1 shows that $\widehat{\mathbf{F}}_t$ and $\widehat{\boldsymbol{\Lambda}}$ are consistent estimators of the rotated latent factors $\mathbf{R}\mathbf{F}_t$ and rotated factor loadings matrix $\boldsymbol{\Lambda}\mathbf{R}^{-1}$, respectively, rather than the factors and factor loadings themselves (unless $\mathbf{R} = \mathbf{I}$). If, in addition, we assume that $N \gg T^{(2+\delta)/\delta}$, the rate $T^{2/\delta}/N^{1/2}$ would disappear in (3.2.10)–(3.2.12). In practice, the number of factors, K, is usually unknown but can be consistently estimated via an information criterion (Bai and Ng, 2002) or a simple ratio method (Lam and Yao, 2012).

3.2.3 Large precision matrix estimation

With the estimates of $\boldsymbol{\Lambda}$, $\mathbf{F}_t$, and $\boldsymbol{\varepsilon}_t$ from Section 3.2.2, the estimation procedure for $\boldsymbol{\Omega}_{\varepsilon}^{o}(\mathbf{u})$ and $\boldsymbol{\Omega}_{X}^{o}(\mathbf{u})$ includes the following steps: (i) use a semiparametric method to estimate $\boldsymbol{\Sigma}_{\varepsilon}^{o}(\mathbf{u})$ and $\boldsymbol{\Sigma}_{F}^{o}(\mathbf{u})$;

(ii) apply the method of constrained ℓ_1 minimisation for inverse matrix estimation (CLIME) to estimate $\mathbf{\Omega}^o_\varepsilon(\mathbf{u})$; and (iii) with the estimates of $\mathbf{\Lambda}$, $\mathbf{\Sigma}^o_F(\mathbf{u})$ and $\mathbf{\Omega}^o_\varepsilon(\mathbf{u})$, compute the estimate of $\mathbf{\Omega}^o_X(\mathbf{u})$ using (3.1.3).

Let $\sigma^o_{\varepsilon,ij}(\mathbf{u})$ be the (i,j)-entry of $\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})$. We next describe a semiparametric method to estimate $\sigma^o_{\varepsilon,ij}(\mathbf{u})$. First, with the estimated idiosyncratic errors, $\widehat{\varepsilon}_{t,i}$, constructed in Section 3.2.2, we estimate the univariate nonparametric function $\sigma_{\varepsilon,k,ij}(u_k)$ by kernel smoothing, i.e.,

$$\widehat{\sigma}_{\varepsilon,k,ij}(u_k) = \frac{\sum_{t=1}^{T-1} K\left(\frac{U_{t,k}-u_k}{h}\right)\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j}}{\sum_{t=1}^{T-1} K\left(\frac{U_{t,k}-u_k}{h}\right)}, \quad 1 \le k \le d,\ 1 \le i,j \le N, \tag{3.2.13}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth that tends to zero as N,T increase. To simplify the notation, we drop the dependence of h on N,T in its notation. Motivated by (3.2.1), we consider the approximate linear regression models:

$$\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j} \approx a_{0,ij} + \sum_{k=1}^{d} a_{k,ij}\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k}), \quad 1 \le i,j \le N. \tag{3.2.14}$$

Treating $\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j}$ and $\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})$ in (3.2.14) as the "response" and "regressors", respectively, and using the ordinary least squares, we can obtain the following estimate of the optimal weights defined in (3.2.3):

$$(\widehat{a}_{1,ij},\ldots,\widehat{a}_{d,ij})^\intercal = \widehat{\mathbf{\Delta}}^{-1}_{\varepsilon,ij}\widehat{\mathbf{W}}_{\varepsilon,ij} \tag{3.2.15}$$

and

$$\widehat{a}_{0,ij} = \frac{1}{T-1}\sum_{t=2}^{T}\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j} - \sum_{k=1}^{d}\widehat{a}_{k,ij}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})\right), \tag{3.2.16}$$

where $\widehat{\mathbf{\Delta}}_{\varepsilon,ij}$ is a $d\times d$ matrix with the (k,l)-entry being

$$\widehat{\delta}_{ij,kl} = \frac{1}{T-1}\sum_{t=1}^{T-1}\widehat{\sigma}^c_{\varepsilon,k,ij}(U_{t,k})\widehat{\sigma}^c_{\varepsilon,l,ij}(U_{t,l}), \quad \widehat{\sigma}^c_{\varepsilon,k,ij}(U_{t,k}) = \widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k}) - \frac{1}{T-1}\sum_{s=1}^{T-1}\widehat{\sigma}_{\varepsilon,k,ij}(U_{s,k}),$$

and $\widehat{\mathbf{W}}_{\varepsilon,ij}$ is a d-dimensional vector with the k-th element being

$$\widehat{w}_{ij,k} = \frac{1}{T-1}\sum_{t=1}^{T-1}\widehat{\sigma}^c_{\varepsilon,k,ij}(U_{t,k})\widehat{\varepsilon}^c_{t+1,(i,j)}, \quad \widehat{\varepsilon}^c_{t+1,(i,j)} = \widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j} - \frac{1}{T-1}\sum_{s=2}^{T}\widehat{\varepsilon}_{s,i}\widehat{\varepsilon}_{s,j}.$$

Combining (3.2.1), (3.2.13), (3.2.15) and (3.2.16), we obtain

$$\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u}) = [\widehat{\sigma}_{\varepsilon,ij}(\mathbf{u})]_{N\times N} \quad \text{with} \quad \widehat{\sigma}_{\varepsilon,ij}(\mathbf{u}) = \widehat{a}_{0,ij} + \sum_{k=1}^{p}\widehat{a}_{k,ij}\widehat{\sigma}_{\varepsilon,k,ij}(u_k). \tag{3.2.17}$$

Similarly, using the estimated factors, $\widehat{\mathbf{F}}_t$, and the same kernel function and bandwidth as in (3.2.13), we can construct $\widehat{\mathbf{C}}_F(\mathbf{u})$ and $\widehat{\mathbf{M}}_F(\mathbf{u})$, which are semiparametric estimates of $\mathbf{C}^o_F(\mathbf{u})$ and $\mathbf{M}^o_F(\mathbf{u})$

(subject to a rotation), and subsequently obtain

$$\widehat{\mathbf{\Sigma}}_F(\mathbf{u}) = \widehat{\mathbf{C}}_F(\mathbf{u}) - \widehat{\mathbf{M}}_F(\mathbf{u})\widehat{\mathbf{M}}_F(\mathbf{u})^\intercal. \tag{3.2.18}$$

In light of (3.2.7), to estimate $\mathbf{\Omega}_X^o(\mathbf{u})$, we also need to have an estimate of the inverse of $\mathbf{\Sigma}_\varepsilon^o(\mathbf{u})$, i.e., $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$. As the size of the matrix $\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})$ may be ultra large, it is often ill-conditioned and computation of its inverse becomes very challenging. To alleviate this problem, we use the uniform sparsity assumption (3.2.8) on $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$ and apply the CLIME method for estimating the inverse of a large matrix. To this end, we define the estimator

$$\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) = \underset{\mathbf{\Omega}}{\operatorname{argmin}} |\mathbf{\Omega}|_1 \quad \text{subject to} \quad \left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\mathbf{\Omega} - \mathbf{I}_N\right\|_{\max} \leq \rho, \tag{3.2.19}$$

where ρ is a tuning parameter that tends to zero as N, T increase. As $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$ is symmetric, we can modify the above estimator by symmetrising it. This leads to the final estimator of $\mathbf{\Omega}_\varepsilon^o(\mathbf{u})$, $\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) = [\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})]_{N\times N}$, where

$$\widehat{\omega}_{\varepsilon,ji}(\mathbf{u}) = \widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) = \widetilde{\omega}_{\varepsilon,ij}(\mathbf{u}) I\left(|\widetilde{\omega}_{\varepsilon,ij}(\mathbf{u})| \leq |\widetilde{\omega}_{\varepsilon,ji}(\mathbf{u})|\right) + \widetilde{\omega}_{\varepsilon,ji}(\mathbf{u}) I\left(|\widetilde{\omega}_{\varepsilon,ij}(\mathbf{u})| > |\widetilde{\omega}_{\varepsilon,ji}(\mathbf{u})|\right), \tag{3.2.20}$$

where $\widetilde{\omega}_{\varepsilon,ij}(\mathbf{u})$ is the (i,j)-entry of $\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})$. Lastly, by the Sherman-Morrison-Woodbury formula (3.2.7), we obtain the following estimate of $\mathbf{\Omega}_X^o(\mathbf{u})$:

$$\widehat{\mathbf{\Omega}}_X(\mathbf{u}) = \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}_F(\mathbf{u})\left[\mathbf{I}_K + \widehat{\mathbf{\Lambda}}_F(\mathbf{u})^\intercal\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}_F(\mathbf{u})\right]^{-1}\widehat{\mathbf{\Lambda}}_F(\mathbf{u})^\intercal\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}), \tag{3.2.21}$$

where $\widehat{\mathbf{\Lambda}}_F(\mathbf{u}) = \widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Sigma}}_F^{1/2}(\mathbf{u})$ with $\widehat{\mathbf{\Lambda}}$ defined in Section 3.2.2.

3.2.4 Dynamic minimum variance portfolio

One of the most common uses of precision matrices is in financial portfolio choice. We next consider an example of using the MAMAR approximation and the large dynamic precision matrix estimation introduced in Sections 3.2.1 and 3.2.3 to construct a dynamic version of the minimum-variance portfolio. To this end, let $\mathbf{X}_{t+1}$ be a vector of N asset returns at time $t+1$ and $\mathbf{U}_t$ be a vector of d conditioning variables, which can be chosen as style factors (such as returns on value stocks, returns on large, small or medium cap stocks) and (macro)economic and financial variables (such as interest rates, inflation rates, returns on market indices). Assume that $\mathbf{X}_{t+1}$ satisfies the approximate factor model structure (3.1.1). Recall that $\mathbf{\Sigma}_X(\mathbf{u}) = \mathsf{Var}(\mathbf{X}_{t+1}|\mathbf{U}_t = \mathbf{u})$ and $\mathbf{\Omega}_X(\mathbf{u}) = \mathbf{\Sigma}_X^{-1}(\mathbf{u})$. The dynamic minimum-variance portfolio for time $t+1$, given $\mathbf{U}_t = \mathbf{u}$, can be obtained by solving

$$\underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^\intercal\mathbf{\Sigma}_X(\mathbf{u})\mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\intercal\mathbf{1} = 1, \tag{3.2.22}$$

where $\mathbf{1}$ is an N-dimensional vector of ones. The analytical solution to (3.2.22) can be written as

$$\mathbf{w}^*(\mathbf{u}) = \frac{\mathbf{\Sigma}_X^{-1}(\mathbf{u})\mathbf{1}}{\mathbf{1}^\intercal\mathbf{\Sigma}_X^{-1}(\mathbf{u})\mathbf{1}} = \frac{\mathbf{\Omega}_X(\mathbf{u})\mathbf{1}}{\mathbf{1}^\intercal\mathbf{\Omega}_X(\mathbf{u})\mathbf{1}}. \tag{3.2.23}$$

Note that we allow short sellings in the construction of the optimal portfolio, i.e., we allow elements of $\mathbf{w}^{\star}(\mathbf{u})$ to be negative. We also assume that there are no transaction costs.

Using the semiparametric MAMAR approximation in Section 3.2.1, we can replace $\mathbf{\Sigma}_X(\mathbf{u})$ and $\mathbf{\Omega}_X(\mathbf{u})$ in (3.2.22) with $\mathbf{\Sigma}_X^o(\mathbf{u}) := \mathbf{\Lambda}\mathbf{\Sigma}_F^o(\mathbf{u})\mathbf{\Lambda}^{\intercal} + \mathbf{\Sigma}_\varepsilon^o(\mathbf{u})$ and $\mathbf{\Omega}_X^o(\mathbf{u}) = \left[\mathbf{\Sigma}_X^o(\mathbf{u})\right]^{-1}$. Consequently, $\mathbf{w}^*(\mathbf{u})$ can be approximated by

$$\mathbf{w}^o(\mathbf{u}) = \frac{\mathbf{\Omega}_X^o(\mathbf{u})\mathbf{1}}{\mathbf{1}^{\intercal}\mathbf{\Omega}_X^o(\mathbf{u})\mathbf{1}}. \tag{3.2.24}$$

Using the methods in Sections 3.2.2 and 3.2.3, we can estimate $\mathbf{\Omega}_X^o(\mathbf{u})$ by $\widehat{\mathbf{\Omega}}_X(\mathbf{u})$, which is defined in (3.2.21). Hence, in practice we estimate $\mathbf{w}^o(\mathbf{u})$ by

$$\widehat{\mathbf{w}}(\mathbf{u}) = \frac{\widehat{\mathbf{\Omega}}_X(\mathbf{u})\mathbf{1}}{\mathbf{1}^{\intercal}\widehat{\mathbf{\Omega}}_X(\mathbf{u})\mathbf{1}}. \tag{3.2.25}$$

3.3 Main theoretical results

Define $c_{F,k,ij}(u_k) = \mathsf{E}\left(F_{t+1,i}F_{t+1,j}|U_{t,k} = u_k\right)$ and $m_{F,k,i}(u_k) = \mathsf{E}\left(F_{t+1,i}|U_{t,k} = u_k\right)$, $1 \le i,j \le K$, $1 \le k \le d$. Let $\mathbf{\Delta}_{F,ij}^*$ and $\mathbf{\Delta}_{F,i}^{\diamond}$ be $d \times d$ matrices whose (k,l)-entries are $\delta_{F,ij,kl}^* = \mathsf{Cov}[c_{F,k,ij}(U_{t,k}), c_{F,l,ij}(U_{t,l})]$ and $\delta_{F,i,kl}^{\diamond} = \mathsf{Cov}\left[m_{F,k,i}(U_{t,k}), m_{F,l,i}(U_{t,l})\right]$, respectively. We introduce the following assumptions, which will be used for establishing uniform consistency results for $\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})$ and $\widehat{\mathbf{\Omega}}_X(\mathbf{u})$.

Assumption 3.B. *(i) The α-mixing dependence condition in Assumption 3.A(i) is satisfied for the joint process $\{(\mathbf{U}_t^{\intercal}, \mathbf{F}_t^{\intercal}, \boldsymbol{\varepsilon}_t^{\intercal})^{\intercal}\}_{t=1}^{\infty}$.*

(ii) The d-dimensional random vector, $\mathbf{U}_t$, has a compact support, $\mathcal{U} = \prod_{k=1}^{p}\mathcal{U}_k$, where $\mathcal{U}_k = [a_k, b_k]$ is the support of the k-th conditioning variable, $U_{t,k}$. The marginal density functions of $U_{t,k}$, $f_k(\cdot)$, $1 \le k \le d$, have continuous second-order derivatives and satisfy

$$\min_{1\le k\le d}\inf_{a_k\le u_k\le b_k} f_k(u_k) \ge c_3 > 0.$$

for some positive constant c_3.

Assumption 3.C. *(i) For each k, $1 \le k \le d$, the univariate nonparametric functions $\sigma_{\varepsilon,k,ij}(\cdot)$ $(1 \le i,j \le N)$, $c_{F,k,ij}(\cdot)$ $(1 \le i,j \le K)$, and $m_{F,k,i}(\cdot)$ $(1 \le i \le K)$ have continuous and uniformly bounded derivatives up to the second order.*

(ii) For all $1 \le i,j \le N$, the $d \times d$ matrix $\mathbf{\Delta}_{\varepsilon,ij}^*$, defined in (3.2.3), is positive definite and satisfies

$$0 < c_4 \le \min_{1\le i,j\le N}\lambda_{\min}(\mathbf{\Delta}_{\varepsilon,ij}^*) \le \max_{1\le i,j\le N}\lambda_{\max}(\mathbf{\Delta}_{\varepsilon,ij}^*) \le c_5 < \infty, \tag{3.3.1}$$

where c_4 and c_5 are some positive constants. The same also holds for the matrices $\mathbf{\Delta}_{F,ij}^*$ and $\mathbf{\Delta}_{F,i}^{\diamond}$.

Assumption 3.D. *(i) The kernel function $K(\cdot)$ is symmetric and Lipschitz continuous and has a compact support $[-1,1]$.*

(ii) The bandwidth h and the dimension N satisfy

$$h \to 0, \quad \frac{T^{1-2\iota}h}{\log^3(N\vee T)} \to \infty, \quad N \gg T^{4/\delta}, \quad (NT)\exp\{-c_1T^{\iota}\} = o(1), \tag{3.3.2}$$

where $0 < \iota < 1/2$, c_1 *is defined in Assumption 3.A(iii) and* δ *is defined in Assumption 3.A(v).*

(iii) The tuning parameter ρ *in (3.2.19) satisfies* $\rho = c_6(\zeta_{NT,1} + \zeta_{NT,2})$, *where* c_6 *is a sufficiently large positive constant,*

$$\zeta_{NT,1} = (\log T)^{1/2}\left[\left(\frac{\log N}{T}\right)^{1/2} + \frac{T^{2/\delta}}{N^{1/2}}\right], \quad \zeta_{NT,2} = \sqrt{\log(N \vee T)/(Th)} + h^2.$$

The compact support restriction on the random vector $\mathbf{U}_t$ in Assumption 3.B(ii) is imposed mainly to facilitate the proofs of our uniform consistency results and can be removed by using an appropriate truncation technique (e.g., Remark 1 in Chen *et al.*, 2018). A recent paper by Wang *et al.* (2021b) derives the uniform consistency properties for the nonparametric large covariance matrix estimation without the compact support assumption on $\mathbf{U}_t$. We conjecture that a similar extension can be achieved for the large precision matrix estimation in this chapter. The smoothness condition on the univariate nonparametric functions in Assumption 3.C(i) is common when the kernel smoothing method is applied. Assumption 3.C(ii) ensures that the optimal weights in the MAMAR approximation are well defined (see, for example, (3.2.3)). The conditions in Assumption 3.D(ii) indicate that N can diverge exponentially fast with respect to T, thus covering the ultra-high dimensional time series setting. There is also a trade-off between the bandwidth condition and the divergence rate of N. The convergence rate for ρ in Assumption 3.D(iii) is partly from the uniform convergence result in Proposition 3.3.1 below and is crucial for the validity of the CLIME method.

Proposition 3.3.1. *Suppose that Assumptions 3.A–3.C and 3.D(i)-(ii) are satisfied. Then we have*

$$\max_{1\le i,j\le N} \sup_{\mathbf{u}\in\mathcal{U}_h} \left|\widehat{\sigma}_{\varepsilon,ij}(\mathbf{u}) - \sigma^o_{\varepsilon,ij}(\mathbf{u})\right| = O_P\left(\zeta_{NT,1} + \zeta_{NT,2}\right), \tag{3.3.3}$$

where $\mathcal{U}_h = \prod_{k=1}^d \mathcal{U}_{k,h}$ *with* $\mathcal{U}_{k,h} = [a_k + h, b_k - h]$, $\zeta_{NT,1} = (\log T)^{1/2}\left[(\log N/T)^{1/2} + T^{2/\delta}/N^{1/2}\right]$, *and* $\zeta_{NT,2} = \sqrt{\log(N \vee T)/(Th)} + h^2$.

Proposition 3.3.2. *Suppose that Assumptions 3.A–3.C and 3.D(i)-(ii) are satisfied. Then we have*

$$\sup_{\mathbf{u}\in\mathcal{U}_h} \left\|\widehat{\mathbf{\Sigma}}_F(\mathbf{u}) - \mathbf{R}\mathbf{\Sigma}^o_F(\mathbf{u})\mathbf{R}^\intercal\right\| = O_P\left(\zeta_{NT,1} + \zeta_{NT,2}\right), \tag{3.3.4}$$

where $\mathbf{R}$ *is the rotation matrix defined in (3.2.9),* $\mathbf{\Sigma}^o_F(\mathbf{u})$ *and* $\widehat{\mathbf{\Sigma}}_F(\mathbf{u})$ *are defined in (3.2.6) and (3.2.18), respectively.*

Theorem 3.3.1. *Suppose that Assumptions 3.A–3.D are satisfied and* $\{\mathbf{\Omega}^o_\varepsilon(\mathbf{u}) : \mathbf{u} \in \mathcal{U}\} \subseteq \mathcal{S}(q, \varpi_N, M, \mathcal{U})$, *where* $\mathcal{S}(q, \varpi_N, M, \mathcal{U})$ *is defined in (3.2.8). Then, we have the following uniform consistency results:*

(i)

$$\sup_{\mathbf{u}\in\mathcal{U}_h} \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max} = O_P\left(\zeta_{NT,1} + \zeta_{NT,2}\right); \tag{3.3.5}$$

(ii)

$$\sup_{\mathbf{u}\in\mathcal{U}_h} \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_O = O_P\left(\varpi_N\left(\zeta_{NT,1} + \zeta_{NT,2}\right)^{1-q}\right); \tag{3.3.6}$$

(iii)

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_F^2=O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{2-q}\right). \tag{3.3.7}$$

If we assume that $\mathbf{\Sigma}^o_\varepsilon(\cdot)$ is sufficiently close to $\mathbf{\Sigma}_\varepsilon(\cdot)$ in the sense that

$$\sup_{\mathbf{u}\in\mathscr{U}}\|\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})-\mathbf{\Sigma}_\varepsilon(\mathbf{u})\|_{\max}=O(b_{NT,1})$$

for some $b_{NT,1}\to 0$, then by Proposition 3.3.1, we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})-\mathbf{\Sigma}_\varepsilon(\mathbf{u})\right\|_{\max}=O_P\left(\zeta_{NT,1}+\zeta_{NT,2}+b_{NT,1}\right).$$

If we further assume that the true dynamic precision matrix $\mathbf{\Omega}_\varepsilon(\mathbf{u})$, $\mathbf{u}\in\mathscr{U}$, also belongs to $\mathscr{S}(q,\varpi_N,M,\mathscr{U})$, then by following the proof of Theorem 3.3.1 and setting $\rho=c_5(\zeta_{NT,1}+\zeta_{NT,2}+b_{NT,1})$, we can show that

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_{\max}=O_P\left(\zeta_{NT,1}+\zeta_{NT,2}+b_{NT,1}\right), \tag{3.3.8}$$

$$\begin{aligned}\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_O &\le \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_1\\ &= O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}+b_{NT,1}\right)^{1-q}\right),\end{aligned} \tag{3.3.9}$$

and

$$\begin{aligned}\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_F^2 &\le \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_{\max}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}_\varepsilon(\mathbf{u})\right\|_1\\ &= O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}+b_{NT,1}\right)^{2-q}\right).\end{aligned} \tag{3.3.10}$$

If the MAMAR approximation rate $b_{NT,1}$ satisfies $b_{NT,1}\ll\zeta_{NT,1}+\zeta_{NT,2}$, the uniform convergence rates in (3.3.8)–(3.3.10) would be the same as those in (3.3.5)–(3.3.7).

Theorem 3.3.2. *Suppose that the conditions of Theorem 3.3.1 are satisfied. Then, we have the following uniform consistency results:*

(i)

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}^o_X(\mathbf{u})\right\|_O=O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{1-q}\right); \tag{3.3.11}$$

(ii)

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}^o_X(\mathbf{u})\right\|_F^2=O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{2-q}+\frac{1}{N}\varpi_N^2\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{2-2q}\right). \tag{3.3.12}$$

3.4 Monte-Carlo simulation

In this section, we conduct some Monte-Carlo experiments to examine the finite-sample performance of the proposed methods for estimating large dynamic precision matrices. In order to provide a comprehensive performance study, we examine factor models under four distinct settings. The precision matrices of the idiosyncratic errors in the four factor models exhibit different structures: a dynamic block-diagonal structure, a varying-sparsity structure, a dynamic banded structure, and a dynamic non-sparse structure.

We compare the proposed method for precision matrix estimation with two alternatives, both of which use MAMAR and CLIME in some way but disregard the factor structure to some extent. Specifically, Method 1 uses formula (3.1.2) and the procedure in (3.2.13)–(3.2.18) to compute the covariance matrice estimate $\widehat{\mathbf{\Sigma}}_X(\mathbf{u})$, but applies CLIME directly to $\widehat{\mathbf{\Sigma}}_X(\mathbf{u})$ to compute the precision matrix $\widehat{\mathbf{\Omega}}_X(\mathbf{u})$ instead of utilising the factor structure and the Sherman-Morrison-Woodbury formula (3.2.21). Method 2 completely ignores the factor structure and applies MAMAR directly to $\mathbf{X}_t$ to obtain $\widehat{\mathbf{\Sigma}}_X(\mathbf{u})$ and then CLIME directly to $\widehat{\mathbf{\Sigma}}_X(\mathbf{u})$ to obtain $\widehat{\mathbf{\Omega}}_X(\mathbf{u})$. In all three methods, we use the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$ with the rule-of-thumb bandwidths as the smoothing parameters.

To determine the number of factors, K, we use a commonly-used information criterion proposed by Bai and Ng (2002). For any $1 \leq k \leq \overline{K}$, where $\overline{K}$ is a predetermined positive number, we let $\widehat{\mathbb{F}}(k) = \left[\widehat{\mathbf{F}}_1(k), \cdots \widehat{\mathbf{F}}_T(k)\right]^\intercal$ be the estimated factors given $K = k$. Define

$$\mathsf{V}_n(k) \quad = \quad \min_{\mathbf{\Lambda}(k)} \frac{1}{NT} \sum_{t=1}^{T} \left[\mathbf{X}_t - \mathbf{\Lambda}(k)\widehat{\mathbf{F}}_t(k)\right]^\intercal \left[\mathbf{X}_t - \mathbf{\Lambda}(k)\widehat{\mathbf{F}}_t(k)\right]$$

where $\mathbf{\Lambda}(k) = [\boldsymbol{\lambda}_1(k), \cdots, \boldsymbol{\lambda}_N(k)]^\intercal$ is a $N \times k$ factor loading matrix. Consequently, we can choose the following objective function:

$$\mathsf{IC}(k) = \log\left[\mathsf{V}_n(k)\right] + k \cdot \left(\frac{N+T}{NT}\right) \log(N \wedge T), \tag{3.4.1}$$

and obtain the estimate $\widehat{K}$ via

$$\widehat{K} = \underset{0 \leq k \leq \overline{K}}{\operatorname{argmin}}\, \mathsf{IC}(k). \tag{3.4.2}$$

When $\widehat{K} = 0$, the common components disappear and our method degenerates to Method 2.

3.4.1 KSIS + PMAMAR method

In this subsection, we introduce the KSIS + PMAMAR method, which combines the approach of kernel sure independence screening (KSIS) and the Penalised Model Averaging MArginal Regression (PMAMAR). This method is proposed by Chen *et al.* (2018), aiming to use KSIS to screen out the unimportant marginal regression functions, and use PMAMAR to further select the most relevant regression functions.

The feasibility of the MAMAR procedure in Section 3.2.3 depends on the positive definiteness of $\mathbf{\Delta}^*_{F,ij}$, $\mathbf{\Delta}^\diamond_{F,i}$, and $\mathbf{\Delta}^*_{\varepsilon,ij}$ (see, for example, (3.2.3) and Assumption 3.C). However, when there are

irrelevant conditioning variables and one or more of these matrices are near singular, the MAMAR optimal weights become ill-defined. For instance, if $U_{t,k}$ is irrelevant to $\varepsilon_{t+1,i}\varepsilon_{t+1,j}$ in the sense that $\sigma_{\varepsilon,k,ij}(U_{t,k}) = \mathsf{E}\big(\varepsilon_{t+1,i}\varepsilon_{t+1,j}|U_{t,k}\big) = \mathsf{E}\big(\varepsilon_{t+1,i}\varepsilon_{t+1,j}\big)$, which is a constant that does not depend on $U_{t,k}$, then $\delta^*_{\varepsilon,ij,kl} = \mathsf{Cov}\left[\sigma_{\varepsilon,k,ij}(U_{t,k}), \sigma_{\varepsilon,l,ij}(U_{t,l})\right] = 0$, for all $1 \le l \le d$. Consequently, the kth column and kth row of $\mathbf{\Delta}^*_{\varepsilon,ij}$ are all 0's and $\mathbf{\Delta}^*_{\varepsilon,ij}$ is not positive definite. This issue will arise if any of the matrices $\mathbf{\Sigma}_{\varepsilon,k}(u_k)$, $1 \le k \le d$, is sparse (so that many of the elements $\sigma_{\varepsilon,k,ij}(u_k)$ are zero) and the sparsity does not depend on the conditioning variable.

The above highlights that in implementation, we need to properly deal with possible irrelevant variables in each MAMAR regression such as (3.2.14). We use the KSIS+PMAMAR approach, which implements a preliminary KSIS step before using a penalised MAMAR to eliminate irrelevant variables and obtain estimates of optimal weights. Taking the regression in (3.2.14) as an example: in the KSIS step, we calculate, for each (i, j) pair and $1 \le k \le K$, the variances of the response variable and the regressors as follows,

$$\widehat{\mathsf{Var}}(\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j}) = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j})^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j}\right)^2 \tag{3.4.3}$$

and

$$\widehat{\mathsf{Var}}(\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})) = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k}))^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})\right)^2, \tag{3.4.4}$$

and screen out those $\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})$'s that satisfy $\widehat{\mathsf{Var}}(\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})) < \kappa \cdot \widehat{\mathsf{Var}}(\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j})$, where κ is a constant within the range of $(0, 1)$. Subsequently, we perform a ridge regression of $\widehat{\varepsilon}_{t,i}\widehat{\varepsilon}_{t,j}$ on the remaining $\widehat{\sigma}_{\varepsilon,k,ij}(U_{t,k})$'s to obtain estimates of the optimal weights. Similarly, we can apply the KSIS+PMAMAR method to each entry of $\mathbf{C}^o_F(\mathbf{u})$ and $\mathbf{M}^o_F(\mathbf{u})$. In the simulation and real data application, we set $\kappa = 0.2$ and use cross-validation to determine the tuning parameter for each ridge regression.

3.4.2 Data generating processes

Throughout this section, the dimension N takes one of the values of 100, 300, and 500. The sample size T is fixed at 300. The conditioning variables $\mathbf{U}_t$ is defined as

$$\mathbf{U}_t = (U_{t1}, U_{t2}, U_{t3})^\intercal = \left(\Phi(\widetilde{U}_{t1}/\sigma_{\widetilde{U}}), \Phi(\widetilde{U}_{t2}/\sigma_{\widetilde{U}}), \Phi(\widetilde{U}_{t3}/\sigma_{\widetilde{U}})\right)^\intercal,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, $\sigma_{\widetilde{U}} = \sqrt{4/3}$, $\widetilde{\mathbf{U}}_t = \left(\widetilde{U}_{t1}, \widetilde{U}_{t2}, \widetilde{U}_{t3}\right)^\intercal$ are drawn from a VAR(1) process:

$$\widetilde{\mathbf{U}}_t = 0.5\widetilde{\mathbf{U}}_{t-1} + \mathbf{v}_t, \quad t = 1, \ldots, T$$

with $\widetilde{\mathbf{U}}_0 = \mathbf{0}$, $\mathbf{v}_t$ are i.i.d. three-dimensional random vectors following the $\mathrm{N}(\mathbf{0}, \Sigma_v)$ distribution with $\Sigma_v = \{\sigma^v_{ij}\}_{3\times3}$ and $\sigma_{ij} = I(i = j) + 0.2I(|i - j| = 1) + 0.1I(|i - j| = 2)$, for i, $j = 1$, 2, and 3. The dynamic precision matrix $\mathbf{\Omega}_X(\mathbf{u})$ is estimated at $\mathbf{U} = \mathbf{u} \in \{\Phi(-0.5/\sigma_{\widetilde{U}}), \Phi(0/\sigma_{\widetilde{U}}), \Phi(0.5/\sigma_{\widetilde{U}})\}^3$ or equivalently $\widetilde{\mathbf{U}} = \widetilde{\mathbf{u}} \in \{-0.5, 0, 0.5\}^3$, which are 27 grid points in total.

Example 3.1 (Dynamic Block Diagonal Precision Matrix). For each $t = 1, \ldots, T$, the factor $\mathbf{F}_t$ is generated independently from a 3-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mu_F(\mathbf{U}_t), \Sigma_F\left(\mathbf{U}_t\right)\right)$, where

$$\mu_F(\mathbf{U}_t) = (\sin(2\pi\widetilde{U}_1), \sin(2\pi\widetilde{U}_2), \sin(2\pi\widetilde{U}_3))^{\intercal},$$

$$\mathbf{\Sigma}_F\left(\mathbf{U}_t\right) = \left\{\sigma_{ij}^F\left(\mathbf{U}_t\right)\right\}_{3\times 3} \text{ with } \sigma_{ij}^F\left(\mathbf{U}_t\right) = 0.4\varsigma_{ij}^F\left(\widetilde{U}_{t1}\right) + 0.3\varsigma_{ij}^F\left(\widetilde{U}_{t2}\right) + 0.3\varsigma_{ij}^F\left(\widetilde{U}_{t3}\right),$$

and

$$\begin{aligned}\varsigma_{ij}^F(v) = \quad & (2 + \arctan(v/2))\left\{I(i=j) + (2.5v + 0.75)I(-0.3 \leqslant v \leqslant 0.1)I(|i-j| = 1)\right. \\ & \left. +(2v - 0.4)I(0.2 \leqslant v \leqslant 0.4)I(|i-j| = 2)\right\}.\end{aligned}$$

The idiosyncratic error $\boldsymbol{\varepsilon}_t$ is independently generated from an N-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mathbf{0}, \Sigma_\varepsilon\left(\mathbf{U}_t\right)\right)$, where

$$\mathbf{\Sigma}_\varepsilon\left(\mathbf{U}_t\right) = \mathbf{I}_{N/5\times N/5} \otimes \widetilde{\mathbf{\Sigma}}_\varepsilon\left(\mathbf{U}_t\right) \text{ with } \widetilde{\mathbf{\Sigma}}_\varepsilon\left(\mathbf{U}_t\right) = \left\{\widetilde{\sigma}_{ij}^\varepsilon\left(\mathbf{U}_t\right)\right\}_{5\times 5},$$

$$\widetilde{\sigma}_{ii}^\varepsilon\left(\mathbf{U}_t\right) = \phi(4\widetilde{U}_1 + 3 - i) + \phi(4\widetilde{U}_2) + \phi(4\widetilde{U}_3 + 3 - i), \text{ for } i = 1, ..., 5,$$

$$\widetilde{\sigma}_{ij}^\varepsilon\left(\mathbf{U}_t\right) = 0.1\left(\widetilde{\sigma}_{ii}^\varepsilon\left(\widetilde{\mathbf{U}}_t\right)\widetilde{\sigma}_{jj}^\varepsilon\left(\widetilde{\mathbf{U}}_t\right)\right)^{1/2}, \text{ for } i \neq j,$$

and $\phi(\cdot)$ is the probability density function of the standard normal distribution. In this example, the dynamic covariance matrix of $\mathbf{X}$ is additive with respect to the elements of $\mathbf{U}_t$.

Example 3.2 (Dynamic Precision Matrix with Varying Sparsity). For each $t = 1, \ldots, T$, the factor $\mathbf{F}_t$ is generated independently from a 3-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mu_F\left(\mathbf{U}_t\right), \mathbf{\Omega}_F^{-1}\left(\mathbf{U}_t\right)\right)$, where

$$\mu_F\left(\mathbf{U}_t\right) = (\sin(\widetilde{U}_1), \sin(\widetilde{U}_2), \sin(\widetilde{U}_3))^{\intercal},$$

$$\mathbf{\Omega}_F\left(\mathbf{U}_t\right) = \left\{\omega_{ij}^F\left(\mathbf{U}_t\right)\right\}_{3\times 3} \text{ with } \omega_{ij}^F\left(\mathbf{U}_t\right) = 0.3\varsigma_{ij}\left(\widetilde{U}_{t1}\right) + 0.3\varsigma_{ij}\left(\widetilde{U}_{t2}\right) + 0.4\varsigma_{ij}\left(\widetilde{U}_{t3}\right),$$

and

$$\begin{aligned}\varsigma_{ij}(v) = \exp(v/2) \quad & \left\{I(i=j) + 0.5\exp\left[-\tfrac{(v-0.25)^2}{0.75^2-(v-0.25)^2}\right]I(-0.49 \leqslant v \leqslant 0.99)I(|i-j| = 1)\right. \\ & \left. +0.4\exp\left[-\tfrac{(v-0.65)^2}{0.35^2-(v-0.65)^2}I(0.31 \leqslant v \leqslant 0.99)I(|i-j| = 2)\right]\right\}. \qquad (3.4.5)\end{aligned}$$

The idiosyncratic error $\boldsymbol{\varepsilon}_t$ is independently generated from an N-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mathbf{0}, \mathbf{\Omega}_\varepsilon^{-1}\left(\widetilde{\mathbf{U}}_t\right)\right)$, where

$$\mathbf{\Omega}_\varepsilon\left(\mathbf{U}_t\right) = \left\{\omega_{ij}^\varepsilon\left(\mathbf{U}_t\right)\right\}_{N\times N} \text{ with } \omega_{ij}^\varepsilon\left(\mathbf{U}_t\right) = 0.4\varsigma_{ij}\left(\widetilde{U}_{t1}\right) + 0.3\varsigma_{ij}\left(\widetilde{U}_{t2}\right) + 0.3\varsigma_{ij}\left(\widetilde{U}_{t3}\right).$$

Note that even if the conditional precision matrices have an additive structure, the conditional covariance matrices do not maintain the additive structure. This example enables us to evaluate the performance of the MAMAR method for approximating entries of a conditional covariance matrix

that are non-additive.

Example 3.3 (Dynamic Banded Precision Matrix). For each $t = 1, \ldots, T$, the factor $\mathbf{F}_t$ is generated from an 3-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mathbf{0}, \Omega_F^{-1}(\mathbf{U}_t)\right)$, and the idiosyncratic error $\boldsymbol{\varepsilon}_t$ is generated from an N-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mathbf{0}, \Omega_\varepsilon^{-1}(\mathbf{U}_t)\right)$, where

$$\boldsymbol{\Omega}_F(\mathbf{U}_t) = \left\{\omega_{ij}^F(\mathbf{U}_t)\right\}_{3\times 3} \text{ with } \omega_{ij}^F(\mathbf{U}_t) = (2 + \arctan(|\mathbf{U}_t|_1/9)) \times \varsigma_{ij}(|\mathbf{U}_t|_1),$$

$$\boldsymbol{\Omega}_\varepsilon(\mathbf{U}_t) = \left\{\omega_{ij}^\varepsilon(\mathbf{U}_t)\right\}_{N\times N} \text{ with } \omega_{ij}^\varepsilon(\mathbf{U}_t) = (2 + \arctan(|\mathbf{U}_t|_1/9)) \times \varsigma_{ij}(|\mathbf{U}_t|_1),$$

and

$$\varsigma_{ij}(v) = I(i=j) + [\phi(v) + 0.1]I(|i-j| = 1) + \phi(v)I(|i-j| = 2), \tag{3.4.6}$$

in which $\phi(v)$ is the probability density function of the standard normal distribution.

Example 3.4 (Dynamic Non-Sparse Precision Matrix). For each $t = 1, \ldots, T$, the factor $\mathbf{F}_t$ is generated independently from a 3-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mu_F(\mathbf{U}_t), \boldsymbol{\Omega}_F^{-1}(\mathbf{U}_t)\right)$, where

$$\mu_F(\mathbf{U}_t) = (\sin(\widetilde{U}_1/2), \sin(\widetilde{U}_2/2), \sin(\widetilde{U}_3/2))^\intercal,$$

$$\boldsymbol{\Omega}_F(\mathbf{U}_t) = \left\{\omega_{ij}^F(\mathbf{U}_t)\right\}_{3\times 3} \text{ with } \omega_{ij}^F(\mathbf{U}_t) = \varsigma_{ij}^F\left(\widetilde{U}_{t1} + \widetilde{U}_{t2} + \widetilde{U}_{t3}\right),$$

and

$$\begin{aligned}\varsigma_{ij}^F(v) = \quad & (\exp(v/4))\left\{I(i=j) + (0.1 + \phi(v))I(|i-j| = 1)\right. \\ & \left. + \phi(v)I(|i-j| = 2)\right\}.\end{aligned}$$

The idiosyncratic error $\boldsymbol{\varepsilon}_t$ is independently generated from an N-dimensional multivariate Gaussian distribution $\mathrm{N}\left(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon^{-1}(\mathbf{U}_t)\right)$, where

$$\boldsymbol{\Omega}_\varepsilon(\mathbf{U}_t) = \left\{\omega_{ij}^\varepsilon(\mathbf{U}_t)\right\}_{N\times N} \text{ with } \omega_{ij}^\varepsilon(\mathbf{U}_t) = \varsigma_{ij}^\varepsilon\left(\widetilde{U}_{t1} + \widetilde{U}_{t2} + \widetilde{U}_{t3}\right)$$

and

$$\varsigma_{ij}^\varepsilon(v) = \quad \exp(v/4)\phi(v)^{|i-j|}.$$

This Toeplitz structure enables us to evaluate the performance of the MAMAR method for approximating the precision matrix with non-sparse dynamic inverse.

3.4.3 Simulation results

To measure estimation accuracy, we consider the average value (averaged over the 27 grid points) of the scaled estimation errors $N^{-1/2}\left\|\widehat{\boldsymbol{\Sigma}}_X(\mathbf{u}) - \boldsymbol{\Sigma}_X(\mathbf{u})\right\|_F$ and $N^{-1/2}\left\|\widehat{\boldsymbol{\Omega}}_X(\mathbf{u}) - \boldsymbol{\Omega}_X(\mathbf{u})\right\|_F$. In addition, we report the relative estimation error of the portfolio weights $\|\widehat{\mathbf{w}}(\mathbf{u}) - \mathbf{w}(\mathbf{u})\| / \|\mathbf{w}(\mathbf{u})\|$ and the volatility of the portfolio with weight $\widehat{\mathbf{w}}(\mathbf{u})$, that is $\sqrt{\widehat{\mathbf{w}}^\intercal(\mathbf{u})\boldsymbol{\Sigma}_X(\mathbf{u})\widehat{\mathbf{w}}(\mathbf{u})}$.

Table 3.1 reports estimation results for Example 3.1. Our method outperforms the other two

methods in terms of the covariance matrix and precision matrix estimation, as evidenced by all four performance measurements. Importantly, our method exhibits significantly lower relative estimation error in portfolio weights and smaller portfolio volatility compared to the other two methods. Method 1, while having the same estimated covariance matrix as our method, neglects the factor structure during precision matrix estimation. Instead of utilising the Sherman-Morrison-Woodbury formula, Method 1 directly calculates $\widehat{\mathbf{\Omega}}_X(\mathbf{u})$ from $\widehat{\mathbf{\Sigma}}_X(\mathbf{u})$ using the CLIME method, resulting in a sparse estimate of $\mathbf{\Omega}_X(\mathbf{u})$. However, since the true $\mathbf{\Omega}_X(\mathbf{u})$ is not sparse due to the factor structure, this approach leads to poor performance of Method 1 in the estimation of the precision matrix. Method 2 neglects the factor structure also in covariance estimation, leading to a less accurate approximation of the covariance matrix. Similar to Method 1, the direct use of the CLIME method for precision matrix estimation exacerbates the estimation performance.

Tables 3.2–3.4 report estimation results for Examples 3.2–3.4. The same pattern as in Table 3.1 is observed, where our method consistently outperforms the other two regardless of the form of the precision matrix. Although Method 2 exhibits inferior performance compared to Method 1 in terms of estimation errors in the Frobenius norm for covariance and precision matrix estimation in all four examples, this does not necessarily result in worse performance in the estimation of the optimal portfolio weights. In other words, the approximation error and the estimation error in large matrix estimation may accumulate non-linearly in the construction of estimators related to optimal portfolio weights. As we can see from Example 3.1, Method 2 even yields a slightly superior estimate of the portfolio weights than Method 1.

In summary, taking into account and utilising the factor structure in the estimation of the covariance and precision matrices can lead to more accurate estimates, which can further lead to better performance in portfolio choice.

3.5 An empirical application

We now apply the proposed method to daily returns of S&P 500 Index constituents in the construction of global minimum variance portfolios. The data are collected from the Thomson Reuters Eikon database and cover a period from 1 Jan 2021 to 31 Dec 2022. As for the conditioning variables, we use the one-day-before returns on the Fama–French three factors, which are downloaded from Keneth French's data library website [1].

We use a rolling window structure to test the performance of our model. Specifically, at the beginning of each month, we re-estimate the model parameters using data from the most recent 12 months. At the beginning of each trading day, we calculate the weights of the assets using the returns of the three factors in the previous trading day. Thus, the out-of-sample period is from 1 Jan 2022 to 31 Dec 2022.

After obtaining all the out-of-sample global minimum variance portfolio returns, we compute their annualised average return (AVR), annualised standard deviation (STD) and the max drawdown (MDD). These measures are used as measures of the performance of portfolios constructed using the proposed method and Method 1 defined in Section 3.4. In addition, we construct a

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data-library.html

Table 3.1: Average losses (standard error) for Example 3.1.

	Our Method	Method 1	Method 2
	$N^{-1/2}\left\|\widehat{\mathbf{\Sigma}}_X(\mathbf{u})-\mathbf{\Sigma}_X(\mathbf{u})\right\|_F$		
$N=100$	47.193	47.193	56.750
	(13.087)	(13.087)	(22.062)
$N=300$	77.530	77.530	105.375
	(20.932)	(20.932)	(62.907)
$N=500$	107.032	107.032	145.094
	(33.855)	(33.855)	(95.681)
	$N^{-1/2}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}_X(\mathbf{u})\right\|_F$		
$N=100$	3.119	4.006	4.217
	(0.022)	(0.052)	(0.153)
$N=300$	3.259	4.247	4.390
	(0.026)	(0.033)	(0.110)
$N=500$	3.325	4.348	4.469
	(0.028)	(0.031)	(0.073)
	$\left\|\widehat{\mathbf{w}}(\mathbf{u})-\mathbf{w}(\mathbf{u})\right\| / \left\|\mathbf{w}(\mathbf{u})\right\|$		
$N=100$	0.481	1.056	1.036
	(0.016)	(0.012)	(0.047)
$N=300$	0.484	1.096	1.067
	(0.011)	(0.010)	(0.063)
$N=500$	0.480	1.094	1.064
	(0.011)	(0.007)	(0.060)
	$\sqrt{\widehat{\mathbf{w}}^\intercal(\mathbf{u})\mathbf{\Sigma}_X(\mathbf{u})\widehat{\mathbf{w}}(\mathbf{u})}$		
$N=100$	0.632	4.864	4.777
	(0.043)	(0.086)	(0.241)
$N=300$	0.361	5.086	4.874
	(0.012)	(0.052)	(0.271)
$N=500$	0.278	5.059	4.840
	(0.008)	(0.036)	(0.240)

Table 3.2: Average losses (standard error) for Example 3.2.

	Our Method	Method 1	Method 2
	$N^{-1/2}\left\|\widehat{\mathbf{\Sigma}}_X(\mathbf{u})-\mathbf{\Sigma}_X(\mathbf{u})\right\|_F$		
$N=100$	8.261	8.261	12.654
	(1.368)	(1.368)	(0.00)
$N=300$	15.062	15.062	31.328
	(2.048)	(2.048)	(8.338)
$N=500$	19.233	19.233	47.491
	(2.381)	(2.381)	(14.934)
	$N^{-1/2}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}_X(\mathbf{u})\right\|_F$		
$N=100$	0.583	0.633	1.030
	(0.014)	(0.014)	(0.045)
$N=300$	0.688	0.714	1.113
	(0.013)	(0.013)	(0.025)
$N=500$	0.745	0.755	1.131
	(0.013)	(0.014)	(0.013)
	$\left\|\widehat{\mathbf{w}}(\mathbf{u})-\mathbf{w}(\mathbf{u})\right\| / \left\|\mathbf{w}(\mathbf{u})\right\|$		
$N=100$	0.173	0.349	0.224
	(0.024)	(0.038)	(0.068)
$N=300$	0.127	0.387	0.137
	(0.013)	(0.032)	(0.042)
$N=500$	0.105	0.429	0.108
	(0.010)	(0.025)	(0.030)
	$\sqrt{\widehat{\mathbf{w}}^\intercal(\mathbf{u})\mathbf{\Sigma}_X(\mathbf{u})\widehat{\mathbf{w}}(\mathbf{u})}$		
$N=100$	0.080	0.151	0.151
	(0.003)	(0.050)	(0.050)
$N=300$	0.044	0.098	0.100
	(0.001)	(0.028)	(0.033)
$N=500$	0.034	0.083	0.079
	(0.0004)	(0.027)	(0.029)

Table 3.3: Average losses (standard error) for Example 3.3.

	Our Method	Method 1	Method 2
	$N^{-1/2}\left\|\widehat{\mathbf{\Sigma}}_X(\mathbf{u})-\mathbf{\Sigma}_X(\mathbf{u})\right\|_F$		
$N=100$	3.365	3.365	3.416
	(0.462)	(0.462)	(0.447)
$N=300$	5.523	5.523	6.057
	(0.717)	(0.717)	(1.178)
$N=500$	7.182	7.182	8.337
	(1.015)	(1.015)	(1.711)
	$N^{-1/2}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}_X(\mathbf{u})\right\|_F$		
$N=100$	1.614	1.723	1.931
	(0.025)	(0.023)	(0.157)
$N=300$	1.832	1.881	2.243
	(0.022)	(0.021)	(0.097)
$N=500$	1.929	1.963	2.322
	(0.020)	(0.019)	(0.076)
	$\left\|\widehat{\mathbf{w}}(\mathbf{u})-\mathbf{w}(\mathbf{u})\right\| / \left\|\mathbf{w}(\mathbf{u})\right\|$		
$N=100$	0.167	0.270	0.286
	(0.051)	(0.094)	(0.102)
$N=300$	0.124	0.291	0.206
	(0.019)	(0.021)	(0.044)
$N=500$	0.102	0.287	0.182
	(0.015)	(0.018)	(0.042)
	$\sqrt{\widehat{\mathbf{w}}^\intercal(\mathbf{u})\mathbf{\Sigma}_X(\mathbf{u})\widehat{\mathbf{w}}(\mathbf{u})}$		
$N=100$	0.051	0.094	0.102
	(0.003)	(0.032)	(0.036)
$N=300$	0.028	0.068	0.072
	(0.0006)	(0.023)	(0.026)
$N=500$	0.021	0.056	0.056
	(0.0004)	(0.019)	(0.021)

Table 3.4: Average losses (standard error) for Example 3.4.

	Our Method	Method 1	Method 2
	$N^{-1/2}\left\|\widehat{\mathbf{\Sigma}}_X(\mathbf{u})-\mathbf{\Sigma}_X(\mathbf{u})\right\|_F$		
$N=100$	7.979	7.979	15.215
	(1.165)	(1.165)	(3.630)
$N=300$	13.935	13.935	39.040
	(1.979)	(1.979)	(9.760)
$N=500$	17.946	17.946	61.546
	(2.550)	(2.550)	(21.051)
	$N^{-1/2}\left\|\widehat{\mathbf{\Omega}}_X(\mathbf{u})-\mathbf{\Omega}_X(\mathbf{u})\right\|_F$		
$N=100$	0.640	0.682	1.044
	(0.019)	(0.019)	(0.018)
$N=300$	0.752	0.772	1.102
	(0.020)	(0.020)	(0.008)
$N=500$	0.805	0.814	1.114
	(0.020)	(0.022)	(0.006)
	$\left\|\widehat{\mathbf{w}}(\mathbf{u})-\mathbf{w}(\mathbf{u})\right\| / \left\|\mathbf{w}(\mathbf{u})\right\|$		
$N=100$	0.211	0.367	0.201
	(0.020)	(0.037)	(0.020)
$N=300$	0.146	0.393	0.125
	(0.011)	(0.037)	(0.035)
$N=500$	0.121	0.415	0.099
	(0.010)	(0.035)	(0.027)
	$\sqrt{\widehat{\mathbf{w}}^\intercal(\mathbf{u})\mathbf{\Sigma}_X(\mathbf{u})\widehat{\mathbf{w}}(\mathbf{u})}$		
$N=100$	0.079	0.158	0.158
	(0.003)	(0.055)	(0.057)
$N=300$	0.044	0.104	0.106
	(0.001)	(0.032)	(0.038)
$N=500$	0.033	0.086	0.084
	(0.001)	(0.029)	(0.032)

Table 3.5: Out-of-sample performance of the constructed minimum variance portfolios

	Our Method	Equal-weighted	Static	Method 1
AVR(%)	-3.55	-16.0	-11.9	-14.1
STD(%)	14.8	23.2	20.7	22.0
MDD(%)	25.3	26.9	23.5	25.3

portfolio using (Static) sample covariance matrix and CLIME precision matrix estimate with the same rolling window structure. As a benchmark, we also consider the equally-weighted portfolio (Equally-weighted).

The results are summarised in Table 3.5. In 2022, all four portfolios experienced negative returns, with the S&P 500 index recording a return of -18.11%. Among the four portfolios, our method achieved the highest return of -3.55% and exhibited the lowest standard deviation. The equal-weighted portfolio performed the worst, with a return of -16.0%. The Static model, which utilises the information from the covariance matrix, performed slightly better. In contrast, our method incorporated the information from the factor returns, which may explain its superior performance. Regarding the maximum drawdown, the four portfolios exhibited behaved similarly, indicating that they experienced comparable declines in value over the period.

3.6 Conclusion

In this chapter, we estimate large dynamic precision matrices for high-dimensional time series data where the conditioning random variables are multivariate. To overcome the challenges posed by the curse of dimensionality, we introduce the approximate factor structure and employ the semi-parametric MAMAR approximation to estimate the underlying dynamic covariance matrix of the factors and the idiosyncratic components. By using the Sherman-Morrison-Woodbury formula, we obtain the dynamic precision matrix for the time series. Under some mild conditions such as the approximate sparsity assumption, the proposed precision matrix estimation is proved to be uniformly consistent. The simulation highlights the importance of correctly specifying the low-rank plus sparse structure. In the empirical analysis, we apply the proposed method to the returns of S&P 500 constituents. The results indicate that our method performs well in the portfolio selection problem.

Conclusions

This dissertation has made significant methodological contributions to the existing literature by studying factor-model-based models and methods to analyse different types of data and data features, such as high-frequency data analysis, network analysis, and precision matrix estimation. The findings and contributions of each chapter are summarised below.

Chapter 1 contributes to the market microstructure literature by being the first study, to our best knowledge, to examine and estimate common factors for microstructure noise. We develop the Double Principle Component Analysis, which provides a robust method for estimating separate factor structures for efficient prices and microstructure noise in high-frequency data. By avoiding strong parametric assumptions, DPCA overcomes limitations of existing approaches such as the PCA-VECM method. The consistent estimators obtained through DPCA enable the identification of co-movements in both efficient prices and microstructure noise, offer tools for portfolio management, and facilitate the construction of factor-mimicking portfolios to hedge risks associated with microstructure noise.

Chapter 2 contributes to the high-dimensional VAR literature. Our contributions lie in the development of a three-stage estimation procedure for modelling time-varying networks in high-dimensional locally stationary time series. The proposed methodology, incorporating time-varying LASSO, weighted group LASSO, and time-varying CLIME techniques, provides reliable estimators of transition and error precision matrices. These estimators are used to construct directed Granger causality networks and undirected partial correlation networks, revealing the dependencies among a large panel of time series. The established uniform consistency and oracle properties under sparsity assumptions validate the efficacy of the proposed estimates. Additionally, by extending the methodology to factor-adjusted time-varying VAR, we account for high correlation among large-scale time series, enhancing the applicability of the approach.

Chapter 3 contributes to the high-dimensional precision matrix estimation literature. By introducing the approximate factor structure and employing the semiparametric Model Averaging Marginal Regression approximation, we address the challenges posed by the curse of dimensionality. By utilising the Sherman-Morrison-Woodbury formula and the CLIME method, the estimate of the dynamic precision matrix for the original time series is then obtained. The resulting estimators demonstrate uniform consistency under mild conditions. The simulation results highlight the advantage of utilising the factor structure when estimating large dynamic precision matrices.

In addition to the methodological advancements, empirical contributions are made in each chapter, further enhancing the practical relevance of this dissertation. In Chapter 1, empirical analysis using intraday returns of S&P 500 constituents provides evidence of co-movement in both mi-

crostructure noise and prices caused by common systematic risk factors. In Chapter 2, the developed methodology for modelling time-varying networks is applied to a large macro dataset. The empirical analysis demonstrates the presence of time-varying Granger causal relations and dynamic contemporaneous partial correlations. In Chapter 3, empirical analysis using returns of S&P 500 constituents showcases the performance of the proposed method for estimating large dynamic precision matrices. The results demonstrate the effectiveness of the approach in the portfolio selection problem.

While this dissertation has made significant contributions to the literature on high-dimensional methods, there are some directions for future research. First, the exploration of sparse factor models (e.g., Uematsu and Yamagata, 2023b,a; Freyaldenhoven, 2022) and quantile factor models (e.g., Ando and Bai, 2020; Chen *et al.*, 2021b) presents two interesting directions for future research. Moreover, high-dimensional inference techniques, such as debiased LASSO (e.g., Van de Geer *et al.*, 2014; Zhang and Zhang, 2014) could be incorporated into the factor analysis, and tuning-insensitive approaches, such as scaled LASSO (e.g., Sun and Zhang, 2013; Liu and Wang, 2017), could be more appealing in the precision matrix estimation. Finally, expanding the application domains beyond finance, to areas such as health economics and environmental economics, would broaden the empirical scope of this research.

Appendix A

Appendix to Chapter 1

In the subsequent proofs, we often make use of the following Weyl's inequality, for two $n \times n$ symmetric matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, with eigenvalues $\mu_j(\mathbf{M}_1)$ and $\mu_j(\mathbf{M}_2)$:

$$|\mu_j(\mathbf{M}_1) - \mu_j(\mathbf{M}_2)| \leq \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}, \tag{A.0.1}$$

for $j = 1, ..., n$. If $\mathbf{M}_1$ and $\mathbf{M}_2$ are invertible and $\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}} < 1$, we have

$$\begin{aligned}\|\mathbf{M}_1^{-1} - \mathbf{M}_2^{-1}\|_{\mathsf{O}} &\leq \|\mathbf{M}_1^{-1}\|_{\mathsf{O}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}} \\ &\leq \|\mathbf{M}_1^{-1} - \mathbf{M}_2^{-1}\|_{\mathsf{O}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}} + \|\mathbf{M}_2^{-1}\|_{\mathsf{O}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}} \\ &\leq \frac{\|\mathbf{M}_2^{-1}\|_{\mathsf{O}}\|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}}}{1 - \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathsf{O}}\|\mathbf{M}_2^{-1}\|_{\mathsf{O}}}. \end{aligned} \tag{A.0.2}$$

Note that the max norm is not sub-multiplicative, but we can use $\|\mathbf{M}_1\mathbf{M}_2\|_{\max} \leq \|\mathbf{M}_1\|_\infty\|\mathbf{M}_2\|_{\max}$ or $\|\mathbf{M}_1\mathbf{M}_2\|_{\max} \leq \|\mathbf{M}_1\|_{\max}\|\mathbf{M}_2\|_1$.

A.1 Proofs of main results

We first provide some lemmas that will be useful in the proofs of the main results.

Lemma A.1.1. *Under Assumptions 1.A–1.D,* ***1.E****, *1.F and 1.G(i), we have* $\mu_{K_H}(\boldsymbol{x}^\intercal\boldsymbol{x}) \geq Cdn^{2\tau_{\bar{G}}}$.

Proof. Recall that

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Sigma}_h\mathbf{D}_H\boldsymbol{\Lambda}_H^\intercal + \boldsymbol{\Sigma}_w.$$

By Weyl's inequality,

$$|\mu_{K_H}(\boldsymbol{x}^\intercal\boldsymbol{x}) - \mu_{K_H}(\boldsymbol{\Sigma}_x)| \leq \|\boldsymbol{x}^\intercal\boldsymbol{x} - \boldsymbol{\Sigma}_x\|_{\mathsf{O}}.$$

Thus, we have

$$\begin{aligned}\mu_{K_H}(x^\intercal x) &\geq \mu_{K_H}(\mathbf{\Sigma}_x) - \|x^\intercal x - \mathbf{\Sigma}_x\|_{\mathsf{O}} \\ &\geq \mu_{K_H}(\mathbf{\Sigma}_x) - \|x^\intercal x - \mathbf{\Lambda}_H\mathbf{D}_H\hbar^\intercal\hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}} - \|\mathbf{\Sigma}_w\|_{\mathsf{O}} \\ &\quad -\|\mathbf{\Lambda}_H\mathbf{D}_H\hbar^\intercal\hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}},\end{aligned}$$

where $\mathbf{\Sigma}_w = \mathbf{\Sigma}_U + n\mathbf{D}_V\mathbf{\Sigma}_v\mathbf{D}_V$. To prove the lemma, we only need to show that

$$\mu_{K_H}(\mathbf{\Sigma}_x) \geq Cdn^{2\tau_G^-}, \tag{A.1.3}$$

$$\|x^\intercal x - \mathbf{\Lambda}_H\mathbf{D}_H\hbar^\intercal\hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}} = o_P(dn^{2\tau_G^-}), \tag{A.1.4}$$

and

$$\|\mathbf{\Lambda}_H\mathbf{D}_H\hbar^\intercal\hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}} = o_P(dn^{2\tau_G^-}). \tag{A.1.5}$$

As for (A.1.3), let $\mathbf{B} = \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\mathbf{Q} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{K_H})$ with $\|\boldsymbol{b}_j\|_2$'s sorted in a descending order, where $\mathbf{Q}$ is an orthogonal matrix such that $\mathbf{Q}^\intercal\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\mathbf{Q}$ is a diagonal matrix. Then $\|\boldsymbol{b}_j\|_2^2$, $1 \leq j \leq K_H$, are the non-zero eigenvalues of $\mathbf{B}\mathbf{B}^\intercal = \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^\intercal$ and also the eigenvalues of $\mathbf{B}^\intercal\mathbf{B} = \mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}$. Therefore,

$$\|\boldsymbol{b}_j\|_2^2 \leq \|\boldsymbol{b}_1\|_2^2 = \|\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}\|_{\mathsf{O}} \leq \|\mathbf{D}_H\|_{\mathsf{O}}^2\|\mathbf{\Sigma}_h\|_{\mathsf{O}} \cdot \|\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\|_{\mathsf{O}} = O(dn^{2\bar{\tau}_G^+}),$$

where the last equality holds by Assumptions 1.B and 1.D. On the other hand,

$$\begin{aligned}\|\boldsymbol{b}_{K_H}\|_2^2 &= \mu_{K_H}(\mathbf{\Sigma}_h^{1/2}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h^{1/2}) \\ &\geq \mu_{K_H}(\mathbf{\Sigma}_h)\mu_{K_H}^2(\mathbf{D}_H)\mu_{K_H}(\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H) \geq Cdn^{2\tau_G^-}.\end{aligned} \tag{A.1.6}$$

By Weyl's inequality and triangle inequality, we have

$$|\mu_j(\mathbf{\Sigma}_x) - \|\boldsymbol{b}_j\|_2^2| \leq \|\mathbf{\Sigma}_w\|_{\mathsf{O}} = O(m_{w,nd}), \tag{A.1.7}$$

for $1 \leq j \leq K_H$, where $m_{w,nd} = m_{U,d} + n^{2\bar{\tau}_V}m_{v,d}$. Therefore by Assumption **1.E***, (A.1.6) and (A.1.7), we have

$$\mu_{K_H}(\mathbf{\Sigma}_x) \geq \|\boldsymbol{b}_{K_H}\|_2^2 - \big|\|\boldsymbol{b}_{K_H}\|_2^2 - \mu_{K_H}(\mathbf{\Sigma}_x)\big| \geq Cdn^{2\tau_G^-}.$$

As for (A.1.4), using Lemma A.2.4, we have

$$\begin{aligned}\|x^\intercal x - \mathbf{\Lambda}_H\mathbf{D}_H\hbar^\intercal\hbar\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}} &\leq 2\|\mathbf{\Lambda}_H\|_{\mathsf{O}}\|\mathbf{D}_H\hbar^\intercal w\|_{\mathsf{O}} + \|w^\intercal w - \mathbf{\Sigma}_w\|_{\mathsf{O}} + \|\mathbf{\Sigma}_w\|_{\mathsf{O}} \\ &= O_P(d(\log d/n)^{1/2} \cdot n^{\bar{\tau}_V^+ + \bar{\tau}_G^+})\end{aligned}$$

$$+O_P(d(\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^+})+O(m_{w,nd}), \tag{A.1.8}$$

since $\|\mathbf{\Lambda}_H\|_{\mathsf{O}}=O(d^{1/2})$, $\|\mathbf{D}_H\hbar^{\intercal}w\|_{\mathsf{O}}\le d^{1/2}\|\mathbf{D}_H\hbar^{\intercal}w\|_1$ and $\|w^{\intercal}w-\mathbf{\Sigma}_w\|_{\mathsf{O}}\le d\|w^{\intercal}w-\mathbf{\Sigma}_w\|_{\max}$. Then, under Assumptions **1.E*** and 1.G(i),

$$\|x^{\intercal}x-\mathbf{\Lambda}_H\mathbf{D}_H\hbar^{\intercal}\hbar\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}\|_{\mathsf{O}}=o_P(dn^{2\underline{\tau}_G^-}),$$

if $n^{1+4\underline{\tau}_G^--2\bar{\tau}_V^+-2(\bar{\tau}_G^+\vee\bar{\tau}_V^+)}/(\log d)\to\infty$ and $m_{w,nd}=o(dn^{2\underline{\tau}_G^-})$. As for (A.1.5), by Assumptions 1.D and 1.G(i), we have

$$\begin{aligned}
&\|\mathbf{\Lambda}_H\mathbf{D}_H\hbar^{\intercal}\hbar\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}-\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}\|_{\mathsf{O}}\\
\le\;& C_0d\|\mathbf{D}_H\hbar^{\intercal}\hbar\mathbf{D}_H-\mathbf{D}_H\mathbf{\Sigma}_h\mathbf{D}_H\|_{\mathsf{O}}\\
\le\;& C_0K_Hd\cdot\max\left\{\|\ell^{\intercal}\ell-\mathbf{\Sigma}_F\|_{\max},\|\mathbf{D}_G\mathcal{G}^{\intercal}\mathcal{G}\mathbf{D}_G-\mathbf{D}_G\mathbf{\Sigma}_g\mathbf{D}_G\|_{\max},\|\ell^{\intercal}\mathcal{G}\mathbf{D}_G\|_{\max}\right\}\\
=\;& O_P(dn^{2\bar{\tau}_G^+}(\log d/n)^{1/2})=o_P(dn^{2\underline{\tau}_G^-}),
\end{aligned}$$

where C_0 is a constant larger than 1, and the last line follows from Lemmas A.2.1(ii), A.2.2(ii) and A.2.3(iv). Hence we complete the proof. □

Although we use PCA* in the first step of our estimation procedure, the $n\times n$ matrix $xx^{\intercal}$, on which PCA* is based, is conceptually more difficult to analyse (as the spot covariance matrices are time-varying). It is easier to establish the asymptotic theory of $\widehat{\hbar}$ than that of $\widehat{\hbar}^\star$. Therefore, we first prove the consistency of $\widehat{\hbar}$ (Lemma A.1.2 below) and then use the relations in (1.3.4) to prove the consistency of $\widehat{\hbar}^\star$ in Theorem 1.3.1.

Lemma A.1.2. *Suppose that Assumptions 1.A–1.G are satisfied. Define the rotation matrix*

$$\mathbf{R}=\hbar^{\intercal}\hbar\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}\widehat{\mathbf{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{-1}, \tag{A.1.9}$$

where $\widehat{\mathbf{D}}_{x,K_H}$ is a $K_H\times K_H$ diagonal matrix with the diagonal elements being the first K_H largest eigenvalues of $x^{\intercal}x$ arranged in descending order. Then, we have
(i)

$$\left\|\widehat{\mathbf{\Lambda}}_H-\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right\|_{\max}=O_P\left(n^{-2\underline{\tau}_G^-}\cdot a_{nd}\right)=o_P(1), \tag{A.1.10}$$

where

$$a_{nd}=(\log d)^{1/2}\frac{n^{\bar{\tau}_V^++\bar{\tau}_G^+\vee\bar{\tau}_V^+}}{n^{1/2}}+\frac{m_{w,nd}}{d},$$

in which $\bar{\tau}_G^+=(1/2+\bar{\tau}_G^\diamond)_+$, $\underline{\tau}_G^-=(1/2+\underline{\tau}_G^\diamond)_-$, $\bar{\tau}_V^+=(1/2+\bar{\tau}_V^\diamond)_+$, $m_{w,nd}=m_{U,d}+n^{2\bar{\tau}_V}m_{v,d}$, and $\bar{\tau}_V=1/2+\bar{\tau}_V^\diamond$.

(ii)

$$\|(\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H)^{-1}-\mathbf{I}_{K_H}\|_{\mathsf{O}}=o_P(1),\qquad \|\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H-\mathbf{I}_{K_H}\|_{\mathsf{O}}=o_P(1), \tag{A.1.11}$$

and

$$d^{-1/2}\|\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\|_{\mathsf{O}}=O_P(1),\qquad d^{1/2}\|(\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2})^{-1}\|_{\mathsf{O}}=O_P(1). \tag{A.1.12}$$

(iii)

$$\left\|\widehat{\mathscr{h}}^\intercal-(\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\mathscr{h}^\intercal\right\|_{\mathsf{O}}=O_P\left(n^{-\underline{\tau}_G^-}\cdot\widetilde{a}_{nd}\right), \tag{A.1.13}$$

and

$$\left\|\widehat{\mathscr{h}}^\intercal-(\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\mathscr{h}^\intercal\right\|_{\max}=O_P\left(n^{-1/2+\bar{\tau}_V^+}\cdot b_{nd}\right), \tag{A.1.14}$$

where

$$\widetilde{a}_{nd}=(\log d)^{1/2}\frac{n^{\bar{\tau}_V^++\bar{\tau}_G^+\vee\bar{\tau}_V^+}}{n^{1/2}}+\frac{m_{w,nd}}{d^{1/2}}$$

and

$$b_{nd}=(\log(nd))^{1/(\gamma_2\wedge1)}\cdot n^{-2\underline{\tau}_G^-}\cdot a_{nd}+(\log n)^{1/(\gamma_3\wedge1)}\cdot\frac{1}{d^{1/2}}.$$

Proof of Lemma A.1.2. By the definition of PCA estimation, we may show that

$$\begin{aligned}\left(\widehat{\mathbf{\Lambda}}_H-\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right)\widehat{\mathbf{D}}_{x,K_H}&=(x^\intercal x-\mathbf{\Lambda}_H\mathbf{D}_H\mathscr{h}^\intercal\mathscr{h}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal)\widehat{\mathbf{\Lambda}}_H\\&=\mathbf{\Lambda}_H\mathbf{D}_H\mathscr{h}^\intercal w\widehat{\mathbf{\Lambda}}_H+w^\intercal\mathscr{h}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H\\&\quad+(w^\intercal w-\mathbf{\Sigma}_w)\widehat{\mathbf{\Lambda}}_H+\mathbf{\Sigma}_w\widehat{\mathbf{\Lambda}}_H.\end{aligned}$$

We can control the four terms with respect to max norm as follows,

$$\|\mathbf{\Lambda}_H\mathbf{D}_H\mathscr{h}^\intercal w\widehat{\mathbf{\Lambda}}_H\|_{\max}\le\|\mathbf{\Lambda}_H\|_{\max}\|\mathbf{D}_H\mathscr{h}^\intercal w\|_1\|\widehat{\mathbf{\Lambda}}_H\|_1=O_P(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^++\bar{\tau}_G^+}), \tag{A.1.15}$$

$$\|w^\intercal\mathscr{h}\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H\|_{\max}\le\|w^\intercal\mathscr{h}\mathbf{D}_H\|_{\max}\|\mathbf{\Lambda}_H^\intercal\|_1\|\widehat{\mathbf{\Lambda}}_H\|_1=O_P(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^++\bar{\tau}_G^+}), \tag{A.1.16}$$

$$\|(w^\intercal w-\mathbf{\Sigma}_w)\widehat{\mathbf{\Lambda}}_H\|_{\max}\le\|w^\intercal w-\mathbf{\Sigma}_w\|_{\max}\|\widehat{\mathbf{\Lambda}}_H\|_1=O_P(d(\log d/n)^{1/2}\cdot n^{2\bar{\tau}_V^+}), \tag{A.1.17}$$

and

$$\|\mathbf{\Sigma}_w\widehat{\mathbf{\Lambda}}_H\|_{\max}\le\|\mathbf{\Sigma}_w\|_\infty\|\widehat{\mathbf{\Lambda}}_H\|_{\max}=O_P(d^{1/2}m_{w,nd}), \tag{A.1.18}$$

since $\|\mathbf{\Lambda}_H\|_{\max}=O(1)$, $\left\|\mathscr{h}^\intercal w\right\|_1\le K_H\left\|\mathscr{h}^\intercal w\right\|_{\max}=O_P\left((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^++\bar{\tau}_G^+}\right)$ by Lemma A.2.4, $\|\widehat{\mathbf{\Lambda}}_H\|_1\le d^{1/2}\|\widehat{\mathbf{\Lambda}}_H\|_F=dK_H$, and $\|\widehat{\mathbf{\Lambda}}_H\|_{\max}\le\|\widehat{\mathbf{\Lambda}}_H\|_F=d^{1/2}K_H$. Therefore

$$\left\|\left(\widehat{\mathbf{\Lambda}}_H-\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right)\widehat{\mathbf{D}}_{x,K_H}\right\|_{\max}=O_P\left(d(\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^++\bar{\tau}_G^+\vee\bar{\tau}_V^+}+d^{1/2}m_{w,nd}\right)=O_P(d\cdot\widetilde{a}_{nd}). \tag{A.1.19}$$

Since $\|\widehat{\mathbf{D}}_{x,K_H}^{-1}\|_{\mathsf{O}} = O_P(d^{-1}n^{-2\underline{\tau}_G^-})$ by Lemma A.1.1, we can prove that

$$\left\|\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right\|_{\max} = O_P\left(n^{-2\underline{\tau}_G^-}\cdot\widetilde{a}_{nd}\right) = o_P(1), \tag{A.1.20}$$

by noting that

$$n^{-2\underline{\tau}_G^-}\cdot\widetilde{a}_{nd} = o_P(1)$$

when $n^{1+4\underline{\tau}_G^- - 2\bar{\tau}_V^+ - 2(\bar{\tau}_G^+\vee\bar{\tau}_V^+)}/(\log d)\to\infty$ and $m_{w,nd}/(d^{1/2}n^{2\underline{\tau}_G^-})\to 0$. Note that (A.1.20) already shows the consistency of the factor loading estimator, and we can following the argument of part (ii) to prove $\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\max} = O_P(1)$. Since

$$\|\widehat{\mathbf{\Lambda}}_H\|_{\max} \le \|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\max} + \left\|\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right\|_{\max},$$

we can improve the bound for $\|\mathbf{\Sigma}_w\widehat{\mathbf{\Lambda}}_H\|_{\max}$ in (A.1.18), that is

$$\|\mathbf{\Sigma}_w\widehat{\mathbf{\Lambda}}_H\|_{\max} \le \|\mathbf{\Sigma}_w\|_\infty\|\widehat{\mathbf{\Lambda}}_H\|_{\max} = O_P(m_{w,nd}). \tag{A.1.21}$$

Combining (A.1.15)–(A.1.17) and (A.1.21), we prove the result.

For part (ii), noting that

$$d^{-1}\widehat{\mathbf{\Lambda}}_H^\intercal\widehat{\mathbf{\Lambda}}_H = \mathbf{I}_{K_H} \quad\text{and}\quad \left\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right\|_{\mathsf{O}} \le (dK_H)^{1/2}\left\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right\|_{\max},$$

we have

$$\begin{aligned}
\|d^{-1}\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H - (\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H)^{-1}\|_{\mathsf{O}} &= \|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{O}}\\
&= d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H^\intercal\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{O}}\\
&\le 2d^{-1}\|\widehat{\mathbf{\Lambda}}_H\|_{\mathsf{O}}\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\|_{\mathsf{O}} + d^{-1}\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\|_{\mathsf{O}}^2\\
&= O_P(n^{-2\underline{\tau}_G^-}\cdot a_{nd}) = o_P(1).
\end{aligned} \tag{A.1.22}$$

Then, by triangle inequality and Assumption 1.D, we have $\|(\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H)^{-1} - \mathbf{I}_{K_H}\|_{\mathsf{O}} = o_P(1)$.

One the other hand, by (A.0.2) and (A.1.22),

$$\|(\mathbf{D}_H\mathbf{R})^{-1}(\frac{\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H}{d})^{-1}(\mathbf{R}^\intercal\mathbf{D}_H)^{-1} - \mathbf{I}_{K_H}\|_{\mathsf{O}} \le \frac{\|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{O}}}{1 - \|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{O}}} = o_P(1).$$

Then following the same argument as in (A.1.22), we can prove $\|\mathbf{D}_H\mathbf{R}\mathbf{R}^\intercal\mathbf{D}_H - \mathbf{I}_{K_H}\|_{\mathsf{O}} = o_P(1)$.

As for (A.1.12), since $\mathbf{R}^\intercal(\mathscr{H}^\intercal\mathscr{H})^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H} = \mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\widehat{\mathbf{\Lambda}}_H$, by (A.1.10) and (A.1.22), we have

$$\|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^\intercal(\mathscr{H}^\intercal\mathscr{H})^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2} - \mathbf{I}_{K_H}\|_{\mathsf{O}}$$

$$
\begin{aligned}
&= \|d^{-1}\mathbf{R}^\intercal(\hbar^\intercal\hbar)^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H} - \mathbf{I}_{K_H}\|_{\mathsf{O}} \\
&\le \|d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{O}} + d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal\|_{\mathsf{O}}\|\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\mathsf{O}} \\
&= O_P(n^{-2\underline{\tau}_G^-}\cdot a_{nd}) = o_P(1).
\end{aligned}
$$

Then by Lemmas A.2.1 and A.2.2, we can prove $\|(\hbar^\intercal\hbar)^{-1}\|_{\mathsf{O}} = O_P(1)$, and therefore we have

$$
d^{-1/2}\|\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\|_{\mathsf{O}} = O_P(1).
$$

Similarly, we can prove the second half of (A.1.12).

For part (iii), we use the following decomposition

$$
\widehat{\hbar}^\intercal - \mathbf{R}^{-1}\hbar^\intercal = d^{-1}\widehat{\mathbf{\Lambda}}_H^\intercal\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)\mathbf{R}^{-1}\hbar^\intercal - d^{-1}\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)^\intercal w^\intercal + d^{-1}\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal w^\intercal. \tag{A.1.23}
$$

For the first term on the right hand side (RHS) of (A.1.23), we have

$$
\begin{aligned}
&\|\widehat{\mathbf{\Lambda}}_H^\intercal\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)\mathbf{R}^{-1}\hbar^\intercal\|_{\mathsf{O}} \\
&\le \|\widehat{\mathbf{\Lambda}}_H^\intercal\|_{\mathsf{O}}\left\|\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)\widehat{\mathbf{D}}_{x,K_H}\right\|_{\mathsf{O}}\left\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\right\|_{\mathsf{O}}\left\|(\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2})^{-1}\right\|_{\mathsf{O}}\|\hbar^\intercal\|_{\mathsf{O}} \\
&= O_P(d^{1/2})\cdot O_P(d^{3/2}\cdot a_{nd})\cdot O_P(d^{-1/2}n^{-\underline{\tau}_G^-})\cdot O_P(d^{-1/2})\cdot O_P(1) \\
&= O_P(dn^{-\underline{\tau}_G^-}a_{nd}).
\end{aligned} \tag{A.1.24}
$$

For the second term on the RHS of (A.1.23), when $n^{1+4\underline{\tau}_G^- - 4\bar{\tau}_V^+}/\log d \to \infty$, we have

$$
\begin{aligned}
&\left\|\left(\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)^\intercal w^\intercal\right\|_{\mathsf{O}} \le \left\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right\|_{\mathsf{O}}\|w\|_{\mathsf{O}} \\
&= O_P(d^{1/2}n^{-2\underline{\tau}_G^-}a_{nd})\cdot O_P(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^+} + m_{w,nd}^{1/2}) \\
&= O_P(dn^{-\underline{\tau}_G^-}a_{nd})\cdot O_P(n^{-\underline{\tau}_G^-}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^+} + m_{w,nd}^{1/2}d^{-1/2}) \\
&= o_P(dn^{-\underline{\tau}_G^-}a_{nd}).
\end{aligned} \tag{A.1.25}
$$

For the last term on the RHS of (A.1.23), by (A.1.11), Lemma A.2.4(v), and Assumption 1.G(ii), we have

$$
\begin{aligned}
d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\mathbf{\Lambda}_H^\intercal w^\intercal\|_{\mathsf{O}} &\le d^{-1}\|\mathbf{R}^\intercal\mathbf{D}_H\|_{\mathsf{O}}\|\mathbf{\Lambda}_H^\intercal w^\intercal\|_{\mathsf{O}} \\
&= d^{-1}\cdot O_P(1)\cdot O_P\left(d^{1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^+}m_{w,nd}^{1/2} + d^{1/2}m_{w,nd}^{1/2}\right) \\
&= o_P(n^{-\underline{\tau}_G^-}a_{nd}),
\end{aligned} \tag{A.1.26}
$$

when $d^{-1/2}(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^+}m_{w,nd}^{1/2} = o((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^+ + \bar{\tau}_G^+\vee\bar{\tau}_V^+ - \underline{\tau}_G^-})$, or equivalently, $n^{1-4(\bar{\tau}_G^+\vee\bar{\tau}_V^+)+4\underline{\tau}_G^-} = o(d^2\log d/m_{w,nd}^2)$. Combing (A.1.24)–(A.1.26), we have $\|\widehat{\hbar}^\intercal - \mathbf{R}^{-1}\hbar^\intercal\|_{\mathsf{O}} =$

$O_P(n^{-\underline{\tau}_G^-} a_{nd})$, which completes the proof of (A.1.13).

Now we consider (A.1.14) and use the decomposition (A.1.23) again. For the first term on the RHS of (A.1.23), we have

$$\begin{aligned}
&\|\widehat{\mathbf{\Lambda}}_H^\intercal \left(\mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right) \mathbf{R}^{-1} \mathscr{h}^\intercal\|_{\max} \\
&\le \|\widehat{\mathbf{\Lambda}}_H^\intercal \left(\mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right) \mathbf{R}^{-1}\|_\infty \|\mathscr{h}^\intercal\|_{\max} \\
&\le K_H^{1/2} \cdot \|\widehat{\mathbf{\Lambda}}_H^\intercal\|_{\mathsf{O}} \left\|\left(\mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right) \widehat{\mathbf{D}}_{x,K_H}\right\|_{\mathsf{O}} \left\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\right\|_{\mathsf{O}} \left\|(\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2})^{-1}\right\|_{\mathsf{O}} \|\mathscr{h}^\intercal\|_{\max} \\
&= O_P(d^{1/2}) \cdot O_P(d^{3/2} \cdot a_{nd}) \cdot O_P(d^{-1/2} n^{-\underline{\tau}_G^-}) \cdot O_P(d^{-1/2}) \cdot O_P(n^{-1/2} (\log n)^{1/(\gamma_2 \wedge 1)}) \\
&= O_P(d (\log n)^{1/(\gamma_2 \wedge 1)} n^{-1/2 - \underline{\tau}_G^-} a_{nd}),
\end{aligned} \tag{A.1.27}$$

since

$$\|\mathscr{h}^\intercal\|_{\max} \le \|\mathscr{f}\|_{\max} + \|n^{-1/2}\mathscr{g}\|_{\max} = O_P(n^{-1/2}(\log n)) + O_P(n^{-1/2}(\log n)^{1/\gamma_2})$$

by Lemma A.2.7.

For the second term on the RHS of (A.1.23), we have

$$\begin{aligned}
&\left\|\left(\mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right)^\intercal \mathscr{w}^\intercal\right\|_{\max} \le d \cdot \left\|\mathbf{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\mathbf{\Lambda}}_H\right\|_{\max} \|\mathscr{w}\|_{\max} \\
&= d \cdot O_P(n^{-2\underline{\tau}_G^-} a_{nd}) \cdot O_P(n^{-1/2 + \bar{\tau}_V^+} (\log(nd))^{1/(\gamma_2 \wedge 1)}) \\
&= O_P(d (\log(nd))^{1/(\gamma_2 \wedge 1)} n^{-1/2 + \bar{\tau}_V^+ - 2\underline{\tau}_G^-} a_{nd}),
\end{aligned} \tag{A.1.28}$$

since

$$\|\mathscr{w}\|_{\max} \le \|\mathscr{u}\|_{\max} + \|\mathbf{D}_V \mathscr{v}\|_{\max} = O_P(n^{-1/2}(\log(nd))) + O_P(n^{\bar{\tau}_V^\diamond}(\log(nd))^{1/\gamma_2})$$

by Lemma A.2.7.

For the last term on the RHS of (A.1.23), using Lemma A.2.7(v) and (vi), we have

$$\|\mathbf{R}^\intercal \mathbf{D}_H \mathbf{\Lambda}_H^\intercal \mathscr{w}^\intercal\|_{\max} \le K_H \|\mathbf{R}^\intercal \mathbf{D}_H\|_{\max} \|\mathbf{\Lambda}_H^\intercal \mathscr{w}^\intercal\|_{\max} = O_P\left(d^{1/2} n^{-1/2 + \bar{\tau}_V^+} \log(n)^{1/(\gamma_3 \wedge 1)}\right). \tag{A.1.29}$$

Combing (A.1.27)–(A.1.29), we have

$$\|\widehat{\mathscr{h}}^\intercal - \mathbf{R}^{-1}\mathscr{h}^\intercal\|_{\max} = O_P((\log(nd))^{1/(\gamma_2 \wedge 1)} n^{-1/2 + \bar{\tau}_V^+ - 2\underline{\tau}_G^-} a_{nd}) + O_P\left(d^{-1/2} n^{-1/2 + \bar{\tau}_V^+} \log(n)^{1/(\gamma_3 \wedge 1)}\right),$$

which completes the proof of Lemma A.1.2. ■

Proof of Theorem 1.3.1. (i) Following (A.1.19) and noting that

$$\widehat{\mathbf{\Lambda}}_H^\star = \widehat{\mathbf{\Lambda}}_H (\widehat{\mathscr{h}}^\intercal \widehat{\mathscr{h}})^{1/2} = d^{-1/2} \widehat{\mathbf{\Lambda}}_H \widehat{\mathbf{D}}_{x,K_H}^{1/2},$$

we have

$$\left\|\widehat{\mathbf{\Lambda}}_H^{\star} - d^{-1/2}\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G^-}\cdot a_{nd}\right).$$

Using the notation of $\mathbf{R}^{\star}$, it can be equivalently written as

$$\left\|\widehat{\mathbf{\Lambda}}_H^{\star} - \mathbf{\Lambda}_H\mathbf{D}_H(\hbar^{\intercal}\hbar)^{1/2}\mathbf{R}^{\star}\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G^-}\cdot a_{nd}\right).$$

(ii) Note that $\widehat{\hbar}^{\star} = \widehat{\hbar}(\widehat{\hbar}^{\intercal}\widehat{\hbar})^{-1/2} = d^{1/2}\widehat{\hbar}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}$ and

$$(\mathbf{R}^{\star})^{-1} = d^{1/2}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\mathbf{R}^{-1}(\hbar^{\intercal}\hbar)^{1/2}. \tag{A.1.30}$$

By Lemma A.1.2(iii), we have

$$\begin{aligned}\left\|\widehat{\hbar}^{\star\intercal} - (\mathbf{R}^{\star})^{-1}(\hbar^{\intercal}\hbar)^{-1/2}\hbar^{\intercal}\right\|_{\mathsf{O}} &= d^{1/2}\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}(\widehat{\hbar}^{\intercal} - \mathbf{R}^{-1}\hbar^{\intercal})\|_{\mathsf{O}} \\ &= O_P\left(n^{-2\underline{\tau}_G^-}\cdot\widetilde{a}_{nd}\right)\end{aligned}$$

and

$$\begin{aligned}\left\|\widehat{\hbar}^{\star\intercal} - (\mathbf{R}^{\star})^{-1}(\hbar^{\intercal}\hbar)^{-1/2}\hbar^{\intercal}\right\|_{\max} &= d^{1/2}\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}(\widehat{\hbar}^{\intercal} - \mathbf{R}^{-1}\hbar^{\intercal})\|_{\max} \\ &= O_P\left(n^{-1/2+\bar{\tau}_V^+-\underline{\tau}_G^-}\cdot b_{nd}\right).\end{aligned}$$

(iii) For the first part, it is obvious that $\widehat{\hbar}^{\star}\widehat{\mathbf{\Lambda}}_H^{\star\intercal} = \widehat{\hbar}\widehat{\mathbf{\Lambda}}_H^{\intercal}$. For the second part, we have

$$\begin{aligned}&\left\|\widehat{\hbar}^{\star}\widehat{\mathbf{\Lambda}}_H^{\star\intercal} - \hbar\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}\right\|_{\max} \\ &\le K_H\left\|\widehat{\hbar}^{\intercal} - (\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\hbar^{\intercal}\right\|_{\max}\|\mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\max} \\ &+ K_H\left\|\widehat{\mathbf{D}}_{x,K_H}^{-1}(\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\hbar^{\intercal}\right\|_{\max}\left\|(\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R})\widehat{\mathbf{D}}_{x,K_H}\right\|_{\max} \\ &+ K_H\left\|\widehat{\hbar}^{\intercal} - (\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\hbar^{\intercal}\right\|_{\max}\left\|\widehat{\mathbf{\Lambda}}_H - \mathbf{\Lambda}_H\mathbf{D}_H\mathbf{R}\right\|_{\max} \\ &= O_P\left(n^{-1/2+\bar{\tau}_V^+}\cdot b_{nd}\right)\cdot O_P(1) + O_P(d^{-1}(\log d)^{1/(\gamma_2\wedge 1)}n^{-1/2-\underline{\tau}_G^-})\cdot O_P(d\cdot a_{nd}) \\ &= O_P\left(n^{-1/2+\bar{\tau}_V^+}\cdot b_{nd}\right),\end{aligned}$$

as

$$\left\|\widehat{\mathbf{D}}_{x,K_H}^{-1}(\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\right\|_{\mathsf{O}} = \left\|(\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R})^{-1}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}(\mathbf{D}_H\mathbf{R})^{-1}\mathbf{D}_H\mathbf{R}\right\|_{\mathsf{O}} = O_P\left(d^{-1}n^{-\underline{\tau}_G^-}\right)$$

by (A.1.11) and (A.1.12). Thus, we obtain the uniformly convergence rate for the common components.

(iv) We next prove that $\mathbf{R}^{\star}$ is an asymptotically orthogonal matrix. By (A.1.30), we have

$$\mathbf{R}^{\star\intercal}\mathbf{R}^{\star} = d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^{\intercal}(\hbar^{\intercal}\hbar)^{-1}\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2} = d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^{\intercal}\mathbf{D}_H\mathbf{\Lambda}_H^{\intercal}\widehat{\mathbf{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{-1/2}.$$

Thus by (A.1.10) and (A.1.22), we have

$$\begin{aligned}
& \|\mathbf{R}^{\star\intercal}\mathbf{R}^{\star} - \mathbf{I}_{K_H}\|_{\mathsf{O}} \\
\leq\ & \|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\widehat{\boldsymbol{\Lambda}}_H\widehat{\mathbf{D}}_{x,K_H}^{-1/2} - d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\|_{\mathsf{O}} \\
& +\|d^{-1}\widehat{\mathbf{D}}_{x,K_H}^{1/2}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{-1/2} - \mathbf{I}_{K_H}\|_{\mathsf{O}} \\
=\ & \|d^{-1}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\widehat{\boldsymbol{\Lambda}}_H - d^{-1}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R}\|_{\mathsf{O}} \\
& +\|d^{-1}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \mathbf{I}_{K_H}\|_{\mathsf{O}} \\
=\ & O_P(n^{-2\underline{\tau}_G^-}a_{nd}).
\end{aligned}$$

We thus complete the proof of Theorem 1.3.1. ■

Next, we prove 1.3.2. For this, we need some intermediate estimators or infeasible estimators related to $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_\perp$. Recall that $\widehat{\boldsymbol{\beta}}_\perp$ is the matrix of eigenvectors associated with the largest K_F eigenvalues of $\widehat{\mathbf{S}}_{HH} = n^{-1}\widehat{\mathscr{H}^{\star c\intercal}}\widehat{\mathscr{H}^{\star c}}$, and $\widehat{\boldsymbol{\beta}}$ is the matrix of eigenvectors associated with the rest of the K_G eigenvalues. For a $K_H \times K_H$ matrix, $\boldsymbol{\Xi}$, we define $\mathbf{S}_{HH}^{\boldsymbol{\Xi}} := n^{-1}\boldsymbol{\Xi}\mathscr{H}^{c\intercal}\mathscr{H}^{c}\boldsymbol{\Xi}^{\intercal}$, where $\mathscr{H}^c = \mathscr{H} - \overline{\mathscr{H}}$ and $\overline{\mathscr{H}} = n^{-1}\mathbf{1}_n\sum_{s=1}^n \boldsymbol{H}_{s\Delta}^{\intercal}$. Replacing $\widehat{\mathbf{S}}_{HH}$ with $\mathbf{S}_{HH}^{\boldsymbol{\Xi}}$ in the second-step PCA, we can obtain the infeasible estimators, $\boldsymbol{\beta}_\perp^{\boldsymbol{\Xi}}$ and $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$. For simplicity, we use $\mathbf{S}_{HH}$ to denote $\mathbf{S}_{HH}^{\mathbf{I}_{K_H}}$. Later on, we will determine a proper choice of $\boldsymbol{\Xi}$.

Lemma A.1.3. *Suppose that Assumptions 1.A, 1.C and 1.F are satisfied. If the eigenvalues of $\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}$ are bounded away from zero and infinity uniformly with probability approaching one, then $\boldsymbol{\beta}_\perp^{\boldsymbol{\Xi}}$ and $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$ are super-consistent in the sense that*

$$\boldsymbol{\beta}^{\boldsymbol{\Xi}} - \boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}[\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}} = O_P(n^{-1}), \tag{A.1.31}$$

and

$$\boldsymbol{\beta}_\perp^{\boldsymbol{\Xi}} - \boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\boldsymbol{\beta}_\perp^{\boldsymbol{\Xi}} = O_P(n^{-1}). \tag{A.1.32}$$

Proof. We decompose $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$ in the directions of $\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}$ and $\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp$ (which are orthogonal) as

$$\begin{aligned}
\boldsymbol{\beta}^{\boldsymbol{\Xi}} =\ & \boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}[\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
& +\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}}.
\end{aligned} \tag{A.1.33}$$

Note that $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$ satisfies $\boldsymbol{\Xi}^{-1}\mathbf{S}_{HH}(\boldsymbol{\Xi}^{\intercal})^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} = \boldsymbol{\beta}^{\boldsymbol{\Xi}}\mathbf{D}_S^{\boldsymbol{\Xi}}$, where $\mathbf{D}_S^{\boldsymbol{\Xi}}$ is a $K_G \times K_G$ diagonal matrix with the diagonal elements being the K_G smallest eigenvalues of $\boldsymbol{\Xi}^{-1}\mathbf{S}_{HH}(\boldsymbol{\Xi}^{\intercal})^{-1}$ arranged in a descending

order. Using (A.1.33) and the equality $\mathbf{I}_{K_H} = \boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal$, we have

$$\begin{aligned}
\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}}\mathbf{D}_S^{\boldsymbol{\Xi}} =& \boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^{-1}\mathbf{S}_{HH}(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
=& \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
&+ \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\mathbf{I}_{K_H}(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
=& \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
&+ \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
&+ \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}}, \\
=& \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
&+ \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} \\
&+ \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}}.
\end{aligned}$$

Vectorising this expression, we have

$$\begin{aligned}
\mathsf{vec}(\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}}) \;=\; & \{\mathbf{D}_S^{\boldsymbol{\Xi}}\otimes\mathbf{I}_{K_F} - \mathbf{I}_{K_G}\otimes\boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp \\
& -\mathbf{I}_{K_G}\otimes\boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\}^{-1} \\
& \cdot\mathsf{vec}\left(\boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\beta}^{\boldsymbol{\Xi}}\right). \qquad \text{(A.1.34)}
\end{aligned}$$

Recall that $\boldsymbol{\beta} = (\mathbf{O}_{K_G\times K_F} \quad \mathbf{I}_{K_G})^\intercal$ and $\boldsymbol{\beta}_\perp = (\mathbf{I}_{K_F} \quad \mathbf{O}_{K_F\times K_G})^\intercal$. By Lemma A.2.5, we have

$$\begin{aligned}
&\boldsymbol{\beta}^\intercal\mathbf{S}_{HH}\boldsymbol{\beta} = n^{-2}\sum_{s=1}^n \boldsymbol{G}_{s\Delta}^c\boldsymbol{G}_{s\Delta}^{c\intercal} = O_P(n^{-1}), \\
&\boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp = n^{-1}\sum_{s=1}^n \boldsymbol{F}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal} \text{ is bounded away from zero}, \\
&\boldsymbol{\beta}^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp = n^{-3/2}\sum_{s=1}^n \boldsymbol{G}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal} = O_P(n^{-1}),
\end{aligned}$$

where $\boldsymbol{G}_{s\Delta}^c = \boldsymbol{G}_{s\Delta} - n^{-1}\sum_{s=1}^n \boldsymbol{G}_{s\Delta}$ and $\boldsymbol{F}_{s\Delta}^c = \boldsymbol{F}_{s\Delta} - n^{-1}\sum_{s=1}^n \boldsymbol{F}_{s\Delta}$. Thus, only the first block, $\boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp$, of the matrix $\mathbf{S}_{HH} = \begin{bmatrix} \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp & \boldsymbol{\beta}_\perp^\intercal\mathbf{S}_{HH}\boldsymbol{\beta} \\ \boldsymbol{\beta}^\intercal\mathbf{S}_{HH}\boldsymbol{\beta}_\perp & \boldsymbol{\beta}^\intercal\mathbf{S}_{HH}\boldsymbol{\beta} \end{bmatrix}$ does not converge to zero. Therefore, $\mathbf{D}_S^{\boldsymbol{\Xi}} = o_P(1)$, and we have $\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}^{\boldsymbol{\Xi}} = O_P(n^{-1})$. Then using (A.1.33) again, we can prove the consistency of $\boldsymbol{\beta}^{\boldsymbol{\Xi}}$. Using the same argument, we can prove the consistency of $\boldsymbol{\beta}_\perp^{\boldsymbol{\Xi}}$. □

When $\boldsymbol{\Xi} = \mathbf{I}_{K_H}$, the results in Lemma A.1.3 degenerate to Lemma 1 of Harris (1997). When $\boldsymbol{\Xi} = (\boldsymbol{\hbar}^\intercal\boldsymbol{\hbar})^{1/2}(\mathbf{R}^\star)$, and replacing $\mathbf{S}_{HH}$ with $\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal$, we can prove 1.3.2.

Proof of 1.3.2. Note that $\widehat{\boldsymbol{\beta}}$ satisfies $\widehat{\mathbf{S}}_{HH}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}\widehat{\mathbf{D}}_S$, where $\widehat{\mathbf{D}}_S$ is a $K_G\times K_G$ diagonal matrix with the diagonal elements being the K_G smallest eigenvalues of $\widehat{\mathbf{S}}_{HH}$ arranged in a descending

order. Let $\boldsymbol{\Xi} = (\hbar^\intercal \hbar)^{1/2}(\mathbf{R}^\star)^\intercal$. Following similar arguments in the proof of Lemma A.1.3, we have

$$\begin{aligned}
\mathsf{vec}(\boldsymbol{\beta}_\perp^\intercal \boldsymbol{\Xi}^{-1}\widehat{\boldsymbol{\beta}}) = &\Big\{\widehat{\mathbf{D}}_S^{\boldsymbol{\Xi}} \otimes \mathbf{I}_{K_F} - \mathbf{I}_{K_G} \otimes \boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}_\perp \\
&-\mathbf{I}_{K_G} \otimes \boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\Big\}^{-1} \\
&\cdot \mathsf{vec}\left(\boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}[\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\boldsymbol{\Xi}^\intercal\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^\intercal\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}\right). \qquad \text{(A.1.35)}
\end{aligned}$$

Using the convergence results in Lemma A.2.6 and $\|\widehat{\mathbf{D}}_S\|_{\mathsf{O}} \le \|\mathbf{D}_S^{\boldsymbol{\Xi}}\|_{\mathsf{O}} + \|\widehat{\mathbf{S}}_{HH} - \boldsymbol{\Xi}^{-1}\mathbf{S}_{HH}(\boldsymbol{\Xi}^\intercal)^{-1}\|_{\mathsf{O}} = o_P(1)$, we can prove $\boldsymbol{\beta}_\perp^\intercal\boldsymbol{\Xi}^{-1}\widehat{\boldsymbol{\beta}} = O_P\left(\boldsymbol{\beta}_\perp^\intercal[\boldsymbol{\Xi}\widehat{\mathbf{S}}_{HH}\boldsymbol{\Xi}^\intercal]\boldsymbol{\beta}\right) = O_P((\log n)n^{-\underline{\tau}_G}\cdot b_{nd})$.

Then following the same arguments as in the proof of Lemma A.1.3, we can prove the results. ■

Proof of Theorem 1.3.5. (ii) and (iii) follow directly from Theorem 1.3.1 and 1.3.2.

As for (i), by Theorem 1.3.1 and 1.3.2, we have

$$\begin{aligned}
&\left\|\widehat{\hbar}^\star\widehat{\boldsymbol{\beta}}_\perp - \hbar(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp}\right\|_{\mathsf{O}} \\
= \;& O_P\left(\left\|\widehat{\hbar}^\star - \hbar(\boldsymbol{\Xi}^\intercal)^{-1}\right\|_{\mathsf{O}}\|\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp}\|_{\mathsf{O}} + \|\hbar(\boldsymbol{\Xi}^\intercal)^{-1}\|_{\mathsf{O}}\left\|\widehat{\boldsymbol{\beta}}_\perp - \boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp}\right\|_{\mathsf{O}}\right) \\
= \;& O_P\left(n^{-2\underline{\tau}_G}\cdot\widetilde{a}_{nd}\right)\cdot O_P(1) + O_P(1)\cdot O_P\left((\log n)n^{-\underline{\tau}_G}\cdot b_{nd}\right) \\
= \;& O_P\left((\log n)n^{-\underline{\tau}_G}\cdot\widetilde{b}_{nd}\right). \qquad \text{(A.1.36)}
\end{aligned}$$

Using $(\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal) = \mathbf{I}_{K_H}$, we have

$$\begin{aligned}
\hbar(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp} &= \hbar(\boldsymbol{\beta}_\perp\boldsymbol{\beta}_\perp^\intercal + \boldsymbol{\beta}\boldsymbol{\beta}^\intercal)(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp \\
&= \not{\ell}\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp + n^{-1/2}\not{g}\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp[\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp]^{-1}\boldsymbol{\beta}_\perp^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\widehat{\boldsymbol{\beta}}_\perp. \qquad \text{(A.1.37)}
\end{aligned}$$

Therefore, we only need to prove that the second term on the RHS of the second equality of (A.1.37) is $O_P(n^{-2\underline{\tau}_G}a_{nd})$. Indeed, $\|n^{-1/2}\not{g}\|_{\mathsf{O}} = O_P(1)$ and

$$\begin{aligned}
\boldsymbol{\beta}^\intercal(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp &= \boldsymbol{\beta}^\intercal(\hbar^\intercal\hbar)^{-1/2}(\mathbf{R}^{\star\intercal}\mathbf{R}^\star)^{-1}(\hbar^\intercal\hbar)^{-1/2}\boldsymbol{\beta}_\perp \\
&= \boldsymbol{\beta}^\intercal(\hbar^\intercal\hbar)^{-1}\boldsymbol{\beta}_\perp\cdot(1 + O_P(n^{-2\underline{\tau}_G}a_{nd})) = O_P(n^{-2\underline{\tau}_G}a_{nd}),
\end{aligned}$$

where the last two equalities follow from (1.3.13) and the result that $\hbar^\intercal\hbar$ converges to a block diagonal matrix at rate $(\log d/n)^{1/2}$ using Lemmas A.2.1–A.2.3. Thus we complete the proof of (i) under the spectral norm.

As for the max norm convergence, by Theorem 1.3.1 and 1.3.2, we have

$$\left\|\widehat{\hbar}^\star\widehat{\boldsymbol{\beta}}_\perp - \hbar(\boldsymbol{\Xi}^\intercal)^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_\perp\mathbf{Q}_{\boldsymbol{\beta}_\perp}\right\|_{\max}$$

$$
\begin{aligned}
&= O_P\left(\left\|\widehat{\not{h}}^{\star} - \not{h}(\boldsymbol{\Xi}^{\intercal})^{-1}\right\|_{\max} \left\|\boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_{\perp}\mathbf{Q}_{\boldsymbol{\beta}_{\perp}}\right\|_{\max} + \left\|\not{h}(\boldsymbol{\Xi}^{\intercal})^{-1}\right\|_{\max} \left\|\widehat{\boldsymbol{\beta}}_{\perp} - \boldsymbol{\Xi}^{-1}\boldsymbol{\beta}_{\perp}\mathbf{Q}_{\boldsymbol{\beta}_{\perp}}\right\|_{\max}\right) \\
&= O_P\left(n^{-1/2+\bar{\tau}_V^{+}-\underline{\tau}_G^{-}} \cdot b_{nd}\right) \cdot O_P(1) + O_P(n^{-1/2}(\log n)^{1/(\gamma_2\wedge 1)}) \cdot O_P\left((\log n) n^{-\underline{\tau}_G^{-}} \cdot b_{nd}\right) \\
&= O_P\left((\log n)^{2/(\gamma_2\wedge 1)} n^{-1/2+\bar{\tau}_V^{+}-\underline{\tau}_G^{-}} \cdot b_{nd}\right). \qquad \text{(A.1.38)}
\end{aligned}
$$

Using (A.1.37) again, and noting that $\|n^{-1/2}\not{g}\|_{\max} = O_P(n^{-1/2}(\log n)^{1/\gamma_2})$ by Lemma A.2.7(ii), we can prove the result.

As for (iv), by Theorem 1.3.1 and 1.3.2, we have

$$
\left\|\widehat{\boldsymbol{\Lambda}}_H^{\star}\widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}\mathbf{Q}_{\boldsymbol{\beta}}\right\|_{\max} = O_P\left(n^{-2\underline{\tau}_G^{-}} \cdot a_{nd}\right). \qquad \text{(A.1.39)}
$$

Following similar arguments to the proof of part (i), we can show that $\boldsymbol{\beta}_{\perp}^{\intercal}\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta} = O_P(n^{-2\underline{\tau}_G^{-}} a_{nd})$ and that

$$
\left\|\boldsymbol{\Lambda}_H\mathbf{D}_H\boldsymbol{\Xi}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\beta}\mathbf{Q}_{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_G\mathbf{D}_G\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}}\right\|_{\max} = O_P\left(n^{-2\underline{\tau}_G^{-}} \cdot a_{nd}\right). \qquad \text{(A.1.40)}
$$

Combining (A.1.39) and (A.1.40), we complete the proof. ■

Proof of Corollary 1.3.4

The result is a direct consequence of Theorem 1.3.5 by noting that

$$
\boldsymbol{\beta}_{\perp}^{\intercal}(\boldsymbol{\Xi}^{\intercal})^{-1}\widehat{\boldsymbol{\beta}}_{\perp}\mathbf{Q}_{\boldsymbol{\beta}_{\perp}}^{\intercal} = \mathbf{I}_{K_F} \quad \text{and} \quad \mathbf{Q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\intercal}\boldsymbol{\Xi}\widehat{\boldsymbol{\beta}})^{\intercal} = \mathbf{I}_{K_G}.
$$

■

Proof of Theorem 1.3.5. We first prove the uniform consistency of $\widehat{\mathscr{H}}$, and then, following a similar argument as in the proof of Theorem 1.3.1(ii), we can prove the uniform consistency of $\widehat{\mathscr{H}^{\star}}$. We use the following decomposition

$$
\begin{aligned}
\widehat{\mathscr{H}}^{\intercal} - \mathbf{R}^{-1}(\mathscr{H}^{\intercal} - \boldsymbol{H}_0\mathbf{1}_n^{\intercal}) &= d^{-1}\widehat{\boldsymbol{\Lambda}}_H^{\intercal}\left(\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right)\mathbf{R}^{-1}(\mathscr{H}^{\intercal} - \boldsymbol{H}_0\mathbf{1}_n^{\intercal}) \\
&\quad -d^{-1}\left(\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right)^{\intercal}(\mathscr{W}^{\intercal} - \boldsymbol{W}_0\mathbf{1}_n^{\intercal}) \\
&\quad +d^{-1}\mathbf{R}^{\intercal}\mathbf{D}_H\boldsymbol{\Lambda}_H^{\intercal}(\mathscr{W}^{\intercal} - \boldsymbol{W}_0\mathbf{1}_n^{\intercal}). \qquad \text{(A.1.41)}
\end{aligned}
$$

and follow the similar arguments as in (A.1.27)–(A.1.29). The first term on the right hand side (RHS) of (A.1.41) can be bounded as follows,

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Lambda}}_H^{\intercal}\left(\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right)\mathbf{R}^{-1}(\mathscr{H}^{\intercal} - \boldsymbol{H}_0\mathbf{1}_n^{\intercal})\|_{\max} \\
&\le \|\widehat{\boldsymbol{\Lambda}}_H^{\intercal}\left(\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right)\mathbf{R}^{-1}\|_{\infty}\|\mathscr{H}^{\intercal} - \boldsymbol{H}_0\mathbf{1}_n^{\intercal}\|_{\max} \\
&\le K_H^{1/2} \cdot \|\widehat{\boldsymbol{\Lambda}}_H^{\intercal}\|_{\mathsf{O}} \left\|\left(\boldsymbol{\Lambda}_H\mathbf{D}_H\mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H\right)\widehat{\mathbf{D}}_{x,K_H}\right\|_{\mathsf{O}} \left\|\widehat{\mathbf{D}}_{x,K_H}^{-1/2}\right\|_{\mathsf{O}} \left\|(\mathbf{R}\widehat{\mathbf{D}}_{x,K_H}^{1/2})^{-1}\right\|_{\mathsf{O}} \|\mathscr{H}^{\intercal} - \boldsymbol{H}_0\mathbf{1}_n^{\intercal}\|_{\max}
\end{aligned}
$$

$$= O_P(d^{1/2}) \cdot O_P(d \cdot a_{nd}) \cdot O_P(d^{-1/2} n^{-\underline{\tau}_G^-}) \cdot O_P(1) \cdot O_P(\log n)$$
$$= O_P(d(\log n) n^{-\underline{\tau}_G^-} a_{nd}), \tag{A.1.42}$$

since

$$\|\mathscr{H}^\intercal - \boldsymbol{H}_0 \mathbf{1}_n^\intercal\|_{\max} \le \|\mathscr{F} - \boldsymbol{F}_0 \mathbf{1}_n^\intercal\|_{\max} + \|n^{-1/2}(\mathscr{G} - \boldsymbol{G}_0 \mathbf{1}_n^\intercal)\|_{\max} = O_P((\log n)) + O_P(n^{-1/2}(\log n)^{1/\gamma_2})$$

by Lemma A.2.8.

For the second term on the RHS of (A.1.41), we have

$$\left\| \left(\boldsymbol{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H \right)^\intercal (\mathscr{W}^\intercal - \boldsymbol{W}_0 \mathbf{1}_n^\intercal) \right\|_{\max} \le d \cdot \left\| \boldsymbol{\Lambda}_H \mathbf{D}_H \mathbf{R} - \widehat{\boldsymbol{\Lambda}}_H \right\|_{\max} \|\mathscr{W}^\intercal - \boldsymbol{W}_0 \mathbf{1}_n^\intercal\|_{\max}$$
$$= d \cdot O_P(n^{-2\underline{\tau}_G^-} a_{nd}) \cdot O_P(\log(nd))$$
$$= O_P(d(\log(nd)) n^{-2\underline{\tau}_G^-} a_{nd}), \tag{A.1.43}$$

since

$$\|\mathscr{W}^\intercal - \boldsymbol{W}_0 \mathbf{1}_n^\intercal\|_{\max} \le \|\mathscr{U} - \boldsymbol{U}_0 \mathbf{1}_n^\intercal\|_{\max} + \|\mathbf{D}_V(\mathscr{V} - \boldsymbol{V}_0 \mathbf{1}_n^\intercal)\|_{\max} = O_P((\log(nd))) + O_P(n^{\tau_V^\diamond}(\log(nd))^{1/\gamma_2})$$

by Lemma A.2.8 and $\tau_V^\diamond < -1/4$ by Assumption 1.G(i).

For the last term on the RHS of (A.1.41), by (A.1.11) and Lemma A.2.8(v) and (vi), we have

$$\begin{aligned} \|\mathbf{R}^\intercal \mathbf{D}_H \boldsymbol{\Lambda}_H^\intercal (\mathscr{W}^\intercal - \boldsymbol{W}_0 \mathbf{1}_n^\intercal)\|_{\max} &\le K_H \|\mathbf{R}^\intercal \mathbf{D}_H\|_{\max} \|\boldsymbol{\Lambda}_H^\intercal (\mathscr{W}^\intercal - \boldsymbol{W}_0 \mathbf{1}_n^\intercal)\|_{\max} \\ &= O_P\left(d^{1/2}(\log n)\right). \end{aligned} \tag{A.1.44}$$

Combing (A.1.42)–(A.1.44), we have

$$\|\widehat{\mathscr{H}}^\intercal - \mathbf{R}^{-1}(\mathscr{H}^\intercal - \boldsymbol{H}_0 \mathbf{1}_n^\intercal)\|_{\max} = O_P((\log(nd)) n^{-2\underline{\tau}_G^-} a_{nd}) + O_P\left(d^{-1/2}(\log n)\right),$$

which completes the proof of the uniform consistency of $\widehat{\mathscr{H}}$. Note that $\widehat{\mathscr{H}^\star} = d^{1/2} \widehat{\mathscr{H}} \widehat{\mathbf{D}}_{x,K_H}^{-1/2}$ and $\|\widehat{\mathbf{D}}_{x,K_H}^{-1}\|_{\mathsf{O}} = O_P(d^{-1} n^{-2\underline{\tau}_G^-})$ by Lemma A.1.1, we can prove the uniform consistency of $\widehat{\mathscr{H}^\star}$.

Then following a similar argument as in the proof of Theorem 1.3.5, we can prove the uniform consistency of $\widehat{\mathscr{G}}^\star$ and $\widehat{\mathscr{F}}^\star$.

As for part (ii), it is a direct consequence of part (i). ■

A.2 Auxiliary lemmas

This appendix provides some auxiliary lemmas, which are used in the proofs in Appendix A.

Recall that $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are increments of continuous-time processes, between t and $t-\Delta$, while $\boldsymbol{g}_t$ and $\boldsymbol{v}_t$ are the first-order differences of stationary time series, $\boldsymbol{G}_t$ and $\boldsymbol{V}_t$, for $t = 0, \Delta, \ldots, n\Delta$. Lemmas A.2.1–A.2.3 give the large deviation theory for them. Specially, Lemma A.2.1 is for $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ only, Lemma A.2.2 for $\boldsymbol{g}_t$ and $\boldsymbol{v}_t$ only, and Lemma A.2.3 for mixtures of the continuous-time processes and the discrete-time processes. In addition, Lemma A.2.4 provides bounds for quantities related to $\boldsymbol{w}$. Lemmas A.2.5 and A.2.6 provide bounds for quantities related to the demeaned cumulated factors. Lemmas A.2.7 and A.2.8 provide uniform bounds for the common factors and the idiosyncratic errors.

Lemma A.2.1. *Under Assumption 1.A, we have*
(i) $\big\|\sum_{s=1}^{n} \boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_U\big\|_{\max} = O_P((\log d/n)^{1/2})$;
(ii) $\big\|\sum_{s=1}^{n} \boldsymbol{f}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_F\big\|_{\max} = O_P((\log d/n)^{1/2})$;
(iii) $\big\|\sum_{s=1}^{n} \boldsymbol{u}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal}\big\|_{\max} = O_P((\log d/n)^{1/2})$.
(iv) In addition, if Assumptions 1.D and 1.E hold, we have
$\big\|d^{-1}\sum_{s=1}^{n} \boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^{\intercal}\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\big\|_{\max} = O_P(m_{U,d}(\log d/n)^{1/2})$.

Proof. Parts (i)–(iii) are the same as Lemma 1 in Aït-Sahalia and Xiu (2017). We only prove part (iv), as parts (i)–(iii) can be proved similarly. By Bonferroni inequality and Lemma 10 of Tao *et al.* (2013b), we have

$$\mathsf{P}\left(\left\|\sum_{s=1}^{n} d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{u}_{s\Delta}\boldsymbol{u}_{s\Delta}^{\intercal}\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\right\|_{\max} > c\right) \le K_h^2 \cdot 4\exp(-nc^2/(64C_1))$$

for all $0 \le c \le \mu_d^2(d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H)\cdot n^{1/2}$, where $C_1 = 8\|d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\|_{\max}^2$ is obtained from Lemma 3 of Fan *et al.* (2012). By Assumptions 1.A and 1.D, we have that $\mu_d^2(d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H)$ is bounded away from zero, and $C_1 \le 8\|d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_U\boldsymbol{\Lambda}_H\|_{\mathsf{O}}^2 = O(m_{U,d}^2)$. Then using the exponential inequality and taking $c = m_{U,d}(\log d/n)^{1/2}$, we can prove the result. □

Lemma A.2.2. *Under Assumption 1.F, we have*
(i) $\big\|n^{-1}\sum_{s=1}^{n} \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_v\big\|_{\max} = O_P((\log d/n)^{1/2})$;
(ii) $\big\|n^{-1}\sum_{s=1}^{n} \boldsymbol{g}_{s\Delta}\boldsymbol{g}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_g\big\|_{\max} = O_P((\log d/n)^{1/2})$;
(iii) $\big\|n^{-1}\sum_{s=1}^{n} \boldsymbol{v}_{s\Delta}\boldsymbol{g}_{s\Delta}^{\intercal}\big\|_{\max} = O_P((\log d/n)^{1/2})$.
(iv) In addition, if Assumptions 1.D and 1.E hold, we have
$\big\|(nd)^{-1}\sum_{s=1}^{n} \boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\boldsymbol{\Lambda}_H - d^{-1}\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_v\boldsymbol{\Lambda}_H\big\|_{\max} = O_P((\log d/n)^{1/2})$.

Proof. Parts (i)–(iii) are the same as Lemma C.3 in Fan *et al.* (2013). Note that, under Assumption 1.D(i), the mixing coefficient of $\{\boldsymbol{v}_{s\Delta}\}$ is bounded by $C'_\alpha \exp(-s^{\gamma_2})$, for some positive constant C'_α. Also note that $v_{i,s\Delta}$ still satisfies the exponential-type tail condition, since

$$\begin{aligned}\max_{1\le i\le d} \mathsf{P}\left(|v_{i,s\Delta}| > c\right) &\le \max_{1\le i\le d} 2P\left(|V_{i,s\Delta}| > c/2\right) \\ &\le 2\exp(1-(c/(2b_1))^{\gamma_2}) \le \exp(1-(c/b_3)^{\gamma_4}), \end{aligned} \tag{A.2.1}$$

for $1 \le i \le d$, $s = 1, \cdots, n$, and $c > 0$, where $\gamma_4 \in (0, \gamma_2)$ and $b_3 > 2b_1 \max\{(\gamma_4/\gamma_2)^{1/\gamma_2}, (1+\log 2)^{1/\gamma_2}\}$, and the last inequality is shown in the proof of Lemma C.2 of Fan *et al.* (2011). Again by Lemma C.2 of Fan *et al.* (2011), $|v_{i_1,s\Delta}v_{i_2,s\Delta}|$ still satisfies the exponential-type tail condition,

$$\max_{1\le i\le d} \mathsf{P}\left(|v_{i_1,s\Delta}v_{i_2,s\Delta} - \mathsf{E}[v_{i_1,s\Delta}v_{i_2,s\Delta}]| > c\right) \le \exp(1-(c/b_4)^{\gamma_5}), \tag{A.2.2}$$

for $1 \le i_1, i_2 \le d$, $s = 1, \cdots, n$, $c > 0$, some b_4, and $\gamma_5 \in (0, \gamma_4/2)$. Therefore, using the arguments in the proof of Lemma A.3 in Fan *et al.* (2011), we can show that

$$\mathsf{P}\left(\left\| n^{-1}\sum_{s=1}^n \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal - \boldsymbol{\Sigma}_v \right\|_{\max} > C_2\sqrt{\frac{\log d}{n}}\right) = O\left(\frac{1}{d^2}\right)$$

for some positive constant C_2, which proves part (i). Parts (ii) and (iii) are similar to part (i) and can be obtained from the inequalities derived in Lemma B.1 of Fan *et al.* (2011). As for part (iv), we have

$$\begin{aligned}&\mathsf{P}\left(\left\| n^{-1}\sum_{s=1}^n \boldsymbol{\Lambda}_H^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\Lambda}_H - \boldsymbol{\Lambda}_H^\intercal \boldsymbol{\Sigma}_v \boldsymbol{\Lambda}_H \right\|_{\max} > d\cdot x\right) \\ \le\ & K_H^2 \max_{1\le j_1,j_2\le K_H} \mathsf{P}\left(\left| n^{-1}\sum_{s=1}^n \boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\lambda}_{H,j_2} - \boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{\Sigma}_v \boldsymbol{\lambda}_{H,j_2}\right| > d\cdot x\right). \end{aligned} \tag{A.2.3}$$

Applying similar arguments in (A.2.1) and (A.2.2) and using Lemma C.2 of Fan *et al.* (2011) under Assumption 1.E(iv), we have

$$\max_{1\le j_1,j_2\le K_H} \mathsf{P}\left(d^{-1}|\boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\lambda}_{H,j_2} - \mathsf{E}[\boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\lambda}_{H,j_2}]| > c\right) \le \exp(1-(c/b_5)^{\gamma_6}), \tag{A.2.4}$$

for $\gamma_6 \in (0, \gamma_2\gamma_3/(\gamma_2+\gamma_3))$, $c > 0$, and some $b_5 > 0$ which does not depend on n and d. Since $|\boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\lambda}_{H,j_2}|$ satisfies the strong mixing condition, we can follow the same arguments as the proof of Lemma B.1 in Fan *et al.* (2011) by applying the Bernstein's inequality in Theorem 1 of Merlevède *et al.* (2011) to obtain

$$\mathsf{P}\left(d^{-1}\left| n^{-1}\sum_{s=1}^n \boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{v}_{s\Delta}\boldsymbol{v}_{s\Delta}^\intercal \boldsymbol{\lambda}_{H,j_2} - \boldsymbol{\lambda}_{H,j_1}^\intercal \boldsymbol{\Sigma}_v \boldsymbol{\lambda}_{H,j_2}\right| > C_3\sqrt{\frac{\log d}{n}}\right) = O\left(\frac{1}{d^2}\right), \tag{A.2.5}$$

for some positive constant C_3, which only depends on γ_1, γ_6 and b_5. Then by (A.2.3) and (A.2.5), we can complete the proof of part (iv). □

Lemma A.2.3. *Under Assumptions 1.A, 1.C and 1.F, we have*
(i) $\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}\|_{\max}=O_P((\log d/n)^{1/2})$;
(ii) $\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{g}_{s\Delta}^{\intercal}\|_{\max}=O_P((\log d/n)^{1/2})$;
(iii) $\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{v}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal}\|_{\max}=O_P((\log d/n)^{1/2})$;
(iv) $\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{g}_{s\Delta}\boldsymbol{f}_{s\Delta}^{\intercal}\|_{\max}=O_P((1/n)^{1/2})$;
(v) $\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{G}_{s\Delta}(n^{-1/2}\boldsymbol{F}_{s\Delta}^{\intercal})\|_{\max}=O_P((1/n)^{1/2})$.

Proof. (i) The proof is similar to that of Lemma 11 in Tao *et al.* (2013b). Since $\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{v}_{s\Delta}^{\intercal}=\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}-\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{(s-1)\Delta}^{\intercal}$, we only need to prove

$$n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}=O_P((\log d/n)^{1/2}) \tag{A.2.6}$$

and

$$n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{(s-1)\Delta}^{\intercal}=O_P((\log d/n)^{1/2}). \tag{A.2.7}$$

The proofs of (A.2.6) and (A.2.7) are similar, so we only provide the former. Denote

$$\Omega_0=\{\max_{1\le i\le d}\max_{1\le s\le n}|u_{i,s\Delta}|\le 1\}.$$

Using the Bonferroni and Markov inequalities, we have

$$\mathsf{P}(\Omega_0^c) \;=\; \mathsf{P}\left(\max_{1\le i\le d}\max_{1\le s\le n}|u_{i,s\Delta}|>1\right)\le nde^{C_\sigma/2-n}. \tag{A.2.8}$$

Note that $\boldsymbol{U}_t$ and $\boldsymbol{V}_t$ are independent. Conditional on the whole path of $\boldsymbol{U}_t$, we have

$$\begin{aligned}
&\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\max}>c(\log d/n)^{1/2}\right)\\
\le\;&\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\max}>c(\log d/n)^{1/2},\Omega_0\right)+\mathsf{P}\left(\Omega_0^c\right)\\
\le\;&\mathsf{E}\left[\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\max}>c(\log d/n)^{1/2},\Omega_0\middle|\boldsymbol{U}_t,\,t\in[0,1]\right)\right]+O(nde^{-n})\\
=\;&\mathsf{E}\left[\mathsf{P}\left(\left\|n^{-1/2}\sum_{s=1}^{n}\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}\right\|_{\max}>c(\log d/n)^{1/2}\middle|\Omega_0,\boldsymbol{U}_t,\,t\in[0,1]\right)\right]+O(nde^{-n}).
\end{aligned} \tag{A.2.9}$$

Note that conditional on the path of $\boldsymbol{U}_t$ and Ω_0, $\boldsymbol{u}_{s\Delta}\boldsymbol{V}_{s\Delta}^{\intercal}$ satisfies the same mixing condition and

exponential-tail condition for $\boldsymbol{V}_{s\Delta}$, and the coefficients in these conditions only depend on γ_1, γ_2, and b_1. Thus we can apply the Bernstein's inequality in Theorem 1 of Merlevède *et al.* (2011) to obtain (letting $\bar{c} = c(\log d/n)^{1/2}$)

$$\begin{aligned} & \mathsf{P}\left(\left\| n^{-1/2} \sum_{s=1}^{n} \boldsymbol{u}_{s\Delta} \boldsymbol{V}_{s\Delta}^{\intercal} \right\|_{\max} > \bar{c} \middle| \Omega_0, \boldsymbol{U}_t,\, t \in [0,1] \right) \\ \leq\ & nd^2 \exp\left(-\frac{\bar{c}^{\gamma}}{C_4} \right) + d^2 \exp\left(-\frac{\bar{c}^2}{C_5(1+C_6 n)} \right) + d^2 \exp\left(-\frac{\bar{c}^2}{C_7 n} \exp\left(\frac{\bar{c}^{\gamma(1-\gamma)}}{C_8((\log \bar{c})^{\gamma})} \right) \right) \\ =\ & O(1/d^2), \end{aligned} \tag{A.2.10}$$

when $(\log d)^{2/\gamma - 1} = o(n)$ and c is large enough, where $\gamma = 1/\gamma_1 + 1/\gamma_2$, and C_4–C_8 only depends on γ_1, γ_2, and b_1. Therefore (A.2.10) holds true uniformly for all path of $\boldsymbol{U}_t$ satisfying Ω_0. Combining (A.2.9) and (A.2.10), we can prove (A.2.6). The proofs of (ii)–(v) are similar to that of (i) by choosing proper $\bar{c}$. So we omit them to save space. □

Lemma A.2.4. *Under Assumptions 1.A–1.F, we have*
(i) $\left\| w^{\intercal} w - \boldsymbol{\Sigma}_w \right\|_{\max} = O_P((\log d/n)^{1/2} \cdot n^{2\bar{\tau}_V^+})$;
(ii) $\left\| w^{\intercal} \hbar \mathbf{D}_H \right\|_{\max} = O_P\left((\log d/n)^{1/2} \cdot n^{\bar{\tau}_V^+ + \bar{\tau}_G^+} \right)$;
(iii) $\left\| \mathbf{D}_H \hbar^{\intercal} w \right\|_1 = O_P\left((\log d/n)^{1/2} \cdot n^{\bar{\tau}_V^+ + \bar{\tau}_G^+} \right)$;
(iv) $\|w\|_{\mathsf{O}} = O_P(d^{1/2}(\log d/n)^{1/4} \cdot n^{\bar{\tau}_V^+} + m_{U,d}^{1/2} + n^{\bar{\tau}_V} m_{v,d}^{1/2})$;
(v) $\|\boldsymbol{\Lambda}_H^{\intercal} w^{\intercal}\|_{\mathsf{O}} = O_P\left(d^{1/2}(m_{U,d}^{1/2} + n^{\bar{\tau}_V} m_{v,d}^{1/2})(1 + n^{\bar{\tau}_V^+}(\log d/n)^{1/4}) \right)$;
where $\bar{\tau}_G^+ = (1/2 + \bar{\tau}_G^{\diamond})_+$, $\bar{\tau}_V = 1/2 + \bar{\tau}_V^{\diamond}$, *and* $\bar{\tau}_V^+ = (1/2 + \bar{\tau}_V^{\diamond})_+$.

Proof. For part (i), recall that $w w^{\intercal} = \sum_{s=1}^{n} (\boldsymbol{u}_{s\Delta} + \mathbf{D}_V \boldsymbol{v}_{s\Delta})^{\intercal} (\boldsymbol{u}_{s\Delta} + \mathbf{D}_V \boldsymbol{v}_{s\Delta})$ and $\boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_U + n\mathbf{D}_V \boldsymbol{\Sigma}_v \mathbf{D}_V$. By Lemmas A.2.1(i), A.2.2(i) and A.2.3(i), we have

$$\begin{aligned} \left\| w^{\intercal} w - \boldsymbol{\Sigma}_w \right\|_{\max} \leq\ & \left\| \sum_{s=1}^{n} \boldsymbol{u}_{s\Delta} \boldsymbol{u}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_U \right\|_{\max} + 2n^{1/2} \|\mathbf{D}_V\| \left\| n^{-1/2} \sum_{s=1}^{n} \boldsymbol{u}_{s\Delta} \boldsymbol{v}_{s\Delta}^{\intercal} \right\|_{\max} \\ & + n\|\mathbf{D}_V\|_{\mathsf{O}}^2 \left\| n^{-1} \sum_{s=1}^{n} \boldsymbol{v}_{s\Delta} \boldsymbol{v}_{s\Delta}^{\intercal} - \boldsymbol{\Sigma}_v \right\|_{\max} \\ =\ & O_P\left((\log d/n)^{1/2} \cdot n^{2\bar{\tau}_V^+} \right). \end{aligned}$$

For part (ii), by Lemmas A.2.1(iii), A.2.2(iii), A.2.3(ii) and A.2.3(iii), we have

$$\begin{aligned} \left\| w^{\intercal} \hbar \mathbf{D}_H \right\|_{\max} \leq\ & \left\| \sum_{s=1}^{n} \boldsymbol{u}_{s\Delta} \boldsymbol{f}_{s\Delta}^{\intercal} \right\|_{\max} + n\|\mathbf{D}_V\|_{\mathsf{O}} \left\| n^{-1} \sum_{s=1}^{n} \boldsymbol{v}_{s\Delta} \boldsymbol{g}_{s\Delta}^{\intercal} \right\|_{\max} \|\mathbf{D}_G\|_{\mathsf{O}} \\ & + n^{1/2} \left\| n^{-1/2} \sum_{s=1}^{n} \boldsymbol{u}_{s\Delta} \boldsymbol{g}_{s\Delta}^{\intercal} \right\|_{\max} \|\mathbf{D}_G\|_{\mathsf{O}} + \|\mathbf{D}_V\|_{\mathsf{O}} \left\| \sum_{s=1}^{n} \boldsymbol{v}_{s\Delta} \boldsymbol{f}_{s\Delta}^{\intercal} \right\|_{\max} \\ =\ & O_P\left((\log d/n)^{1/2} \cdot (1 + n^{1+\bar{\tau}_V^{\diamond} + \bar{\tau}_G^{\diamond}} + n^{1/2 + \bar{\tau}_G^{\diamond}} + n^{1/2 + \bar{\tau}_V^{\diamond}}) \right) \end{aligned}$$

$$
\begin{aligned}
&= O_P\left((\log d/n)^{1/2}\cdot(1+n^{1/2+\bar{\tau}_V^{\diamond}})(1+n^{1/2+\bar{\tau}_G^{\diamond}})\right)\\
&= O_P\left((\log d/n)^{1/2}\cdot n^{\bar{\tau}_V^{+}+\bar{\tau}_G^{+}}\right).
\end{aligned}
$$

Part (iii) follows from part (ii) as $\left\|\hbar^{\intercal}w\right\|_1 \leq K_H\left\|\hbar^{\intercal}w\right\|_{\max}$. Part (iv) follows from $\|w\|_{\mathsf{O}} = \left\|w^{\intercal}w\right\|_{\mathsf{O}}^{1/2} \leq (d\left\|w^{\intercal}w-\boldsymbol{\Sigma}_w\right\|_{\max}+\|\boldsymbol{\Sigma}_w\|_{\mathsf{O}})^{1/2}$ and $\|\boldsymbol{\Sigma}_w\|_{\mathsf{O}} \leq (\|\boldsymbol{\Sigma}_w\|_1\|\boldsymbol{\Sigma}_w\|_\infty)^{1/2} = \|\boldsymbol{\Sigma}_w\|_1 = O(m_{U,d}+n^{2\tau_V}m_{v,d})$. Lastly, we consider part (v). By Lemmas A.2.1(iv) and A.2.2(iv), we have

$$
\begin{aligned}
\|\boldsymbol{\Lambda}_H^{\intercal}w^{\intercal}\|_{\mathsf{O}} &= \|\boldsymbol{\Lambda}_H^{\intercal}w^{\intercal}w\boldsymbol{\Lambda}_H\|_{\mathsf{O}}^{1/2}\\
&\leq (\|\boldsymbol{\Lambda}_H^{\intercal}(w^{\intercal}w-\boldsymbol{\Sigma}_w)\boldsymbol{\Lambda}_H\|_{\mathsf{O}}+\|\boldsymbol{\Lambda}_H^{\intercal}\boldsymbol{\Sigma}_w\boldsymbol{\Lambda}_H\|_1)^{1/2}\\
&\leq (K_H\|\boldsymbol{\Lambda}_H^{\intercal}(w^{\intercal}w-\boldsymbol{\Sigma}_w)\boldsymbol{\Lambda}_H\|_{\max}+\|\boldsymbol{\Lambda}_H^{\intercal}\|_1\|\boldsymbol{\Sigma}_w\|_1\|\boldsymbol{\Lambda}_H\|_1)^{1/2}\\
&= O_P\left(d^{1/2}(m_{U,d}^{1/2}+n^{\bar{\tau}_V})(\log d/n)^{1/4}\cdot n^{\bar{\tau}_V^{+}}+d^{1/2}(m_{U,d}^{1/2}+n^{\bar{\tau}_V}m_{v,d}^{1/2})\right)\\
&= O_P\left(d^{1/2}(m_{U,d}^{1/2}+n^{\bar{\tau}_V}m_{v,d}^{1/2})(1+n^{\bar{\tau}_V^{+}}(\log d/n)^{1/4})\right).
\end{aligned}
$$

□

Lemma A.2.5. *Under Assumptions 1.A, 1.C and 1.F, we have*
(i) $\left\|n^{-1}\sum_{s=1}^n(n^{-1/2}\boldsymbol{G}_{s\Delta}^c)(n^{-1/2}\boldsymbol{G}_{s\Delta}^{c\intercal})\right\|_{\max} = O_P(n^{-1})$;
(ii) $n^{-1}\sum_{s=1}^n \boldsymbol{F}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal} \xrightarrow{d} \int_0^1(\int_0^t\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F-\int_0^1\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F)(\int_0^t\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F-\int_0^1\boldsymbol{\sigma}_{fu}\mathsf{d}\boldsymbol{B}_u^F)^{\intercal}\mathsf{d}t$;
(iii) $\left\|n^{-1}\sum_{s=1}^n(n^{-1/2}\boldsymbol{G}_{s\Delta}^c)\boldsymbol{F}_{s\Delta}^{c\intercal}\right\|_{\max} = O_P(n^{-1})$;
where $\boldsymbol{G}_{s\Delta}^c = \boldsymbol{G}_{s\Delta}-\overline{\boldsymbol{G}}$, $\boldsymbol{F}_{s\Delta}^c = \boldsymbol{F}_{s\Delta}-\overline{\boldsymbol{F}}$, $\overline{\boldsymbol{G}} = n^{-1}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}$, *and* $\overline{\boldsymbol{F}} = n^{-1}\sum_{s=1}^n\boldsymbol{F}_{s\Delta}$.

Proof. Parts (i) and (ii) are trivial. For part (iii), by Lemma A.2.3(v), we have

$$
\left\|n^{-3/2}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal}\right\|_{\max} = n^{-1/2}\left\|n^{-1}\sum_{s=1}^n\boldsymbol{G}_{s\Delta}\boldsymbol{F}_{s\Delta}^{\intercal}-\overline{\boldsymbol{G}}\,\overline{\boldsymbol{F}}^{\intercal}\right\|_{\max} = O_P(n^{-1}).
$$

□

Lemma A.2.6. *Under Assumptions 1.A–1.G, we have*
(i) $\|n^{-1}\mathscr{H}^{\star c\intercal}\mathscr{H}^{\star c}\|_{\mathsf{O}} = O_P(1)$;
(ii) $\left\|\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal}-\mathscr{H}^c\right\|_{\max} = O_P\left(n^{-1/2-\underline{\tau}_G}\cdot b_{nd}\right)$;
(iii) $n^{-1}\|\boldsymbol{\Xi}\widehat{\mathscr{H}}^{\star c\intercal}\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal}-\mathscr{H}^{\star c\intercal}\mathscr{H}^{\star c}\|_{\mathsf{O}} = O_P\left((\log n)n^{-\underline{\tau}_G}\cdot b_{nd}\right)$;
(iv) $n^{-1}\|\boldsymbol{\beta}_\perp^{\intercal}(\boldsymbol{\Xi}\widehat{\mathscr{H}}^{\star c\intercal}\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal})\boldsymbol{\beta}\|_{\mathsf{O}} = O_P((\log n)n^{-\underline{\tau}_G}\cdot b_{nd})$,
where $\boldsymbol{\Xi}$ *is defined in 1.3.2.*

Proof. (i) By Lemma A.2.4, the dominate term is $n^{-1}\sum_{s=1}^n\boldsymbol{F}_{s\Delta}^c\boldsymbol{F}_{s\Delta}^{c\intercal}$, which is of order $O_P(1)$.

(ii) By Theorem 1.3.5, we have

$$
\left\|\widehat{\mathscr{H}^{\star}}-(\mathscr{H}-\mathbf{1}_n\boldsymbol{H}_0^{\intercal})(\boldsymbol{\Xi}^{\intercal})^{-1}\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G}\cdot b_{nd}\right).
$$

Since $\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal} = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\intercal}/n)\widehat{\mathscr{H}}\boldsymbol{\Xi}^{\intercal}$, $\mathscr{H}^c\boldsymbol{\Xi}^{\intercal} = (\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\intercal}/n)(\mathscr{H} - \mathbf{1}_n\boldsymbol{h}_0^{\intercal})$ and $\|\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\intercal}/n\|_{\mathsf{O}} = 1$, we then have $\left\|\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal} - \mathscr{H}^c\right\|_{\max} = O_P\left(n^{-\underline{\tau}_G^-}\cdot b_{nd}\right)$.

(iii) The result follows by noticing that

$$\begin{aligned} & n^{-1}\|\boldsymbol{\Xi}\widehat{\mathscr{H}}^{\star c\intercal}\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal} - \mathscr{H}^{c\intercal}\mathscr{H}^{\star c}\|_{\max} \\ \leq\ & 2\left\|\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal} - \mathscr{H}^c\right\|_{\max}\|\mathscr{H}^c\|_{\max} + \left\|\widehat{\mathscr{H}}^{\star c}\boldsymbol{\Xi}^{\intercal} - \mathscr{H}^c\right\|_{\max}^2 \\ =\ & O_P\left(n^{-\underline{\tau}_G^-}\cdot b_{nd}\right)\cdot O_P(\log n). \end{aligned}$$

(iv) is an immediate consequence of (iii).

□

Lemma A.2.7. *(i) Under Assumption 1.A, we have $\|\mathscr{f}\|_{\max} = O_P(n^{-1/2}(\log n))$;*
(ii) Under Assumption 1.F, we have $\|\mathscr{g}\|_{\max} = O_P((\log n)^{1/\gamma_2})$;
(iii) Under Assumptions 1.A and 1.E, we have $\|\mathscr{u}\|_{\max} = O_P(n^{-1/2}\log(nd))$;
(iv) Under Assumptions 1.E and 1.F, we have $\|\mathscr{v}\|_{\max} = O_P(\log(nd)^{1/\gamma_2})$;
(v) Under Assumptions 1.A, 1.D and 1.E, we have $d^{-1/2}\|\mathscr{u}\boldsymbol{\Lambda}_H\|_{\max} = O_P(n^{-1/2}(\log(nd)))$;
(vi) Under Assumptions 1.D–1.F, we have $d^{-1/2}\|\mathscr{v}\boldsymbol{\Lambda}_H\|_{\max} = O_P((\log n)^{1/\gamma_3})$.

Proof. (i) Denote the (i,j)-th entry of $\boldsymbol{\sigma}_{Ft}$ as $\sigma_{Ft,ij}$. We can prove that

$$\begin{aligned} \mathsf{P}(|f_{it}| > C_9 n^{-1/2}\log(n)) &= 2\mathsf{P}\left(\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}\sigma_{Fs,ij}\mathsf{d}B_{js}^F > C_9 n^{-1/2}\log(n)\right) \\ &\leq 2\mathsf{P}\left(\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}\sigma_{F,s\wedge\tau,ij}\mathsf{d}B_{js}^F > C_9 n^{-1/2}\log(n), \tau \geq 1\right) + 2\mathsf{P}(\tau < 1) \end{aligned}$$

where C_9 is a positive constant and the stopping time τ reduces Q_t to a local bounded process defined in Assumption 1.A. By the local (super)martingale property of stochastic exponential of $\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}\sigma_{Ft,ij}\mathsf{d}B_{js}^F$, we have

$$\mathsf{E}\left[\exp\left(\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}\sqrt{n}\sigma_{F,t\wedge\tau,ij}\mathsf{d}B_{js}^F - \frac{1}{2}\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}(\sqrt{n}\sigma_{F,s\wedge\tau,ij})^2\mathsf{d}s\right)\right] \leq 1,$$

and thus

$$\begin{aligned} \mathsf{E}\left[\exp\left(\sum_{j=1}^{K_F}\int_{t/n}^{(t+1)/n}\sqrt{n}\sigma_{F,s\wedge\tau,ij}\mathsf{d}B_{js}^F\right)\right] &\leq \mathsf{E}\left[\exp\left(\frac{n}{2}\int_{t/n}^{(t+1)/n}\sum_{j=1}^{K_F}\sigma_{F,s\wedge\tau,ij}^2\mathsf{d}s\right)\right] \\ &\leq \mathsf{E}\left[\exp\left(\frac{n}{2}\int_{t/n}^{(t+1)/n}Q_{s\wedge\tau}\mathsf{d}s\right)\right] < \infty. \end{aligned}$$

By Bonferroni and Markov inequalities and note that $\tau \to \infty$, we have

$$\mathsf{P}(\|\not{f}\|_{\max} > C_9 n^{-1/2}\log(n)) \leq \sum_{t=1}^{n}\sum_{j=1}^{K_F}\mathsf{P}(|f_{it}| > C_9 n^{-1/2}\log(n)) = O(n^{1-C_9}),$$

which converge to 0 when C_9 is large enough.

The proofs of (ii)–(iv) are similar to that of (i) so we omit them to save space.

(v) Denote the (i,j)-th entry of $\boldsymbol{\sigma}_{Ut}$ as $\sigma_{Ut,ij}$. Following the same argument as in the proof of part (i), we have

$$\begin{aligned}
& \mathsf{E}\left[\exp\left(\sum_{i=1}^{d}\sum_{j=1}^{d}\lambda_{H,ik}\int_{t/n}^{(t+1)/n}(n/d)^{1/2}\sigma_{U,s\wedge\tau,ij}\mathsf{d}B_{js}^{U}\right)\right] \\
\leq\ & \mathsf{E}\left[\exp\left(\frac{n}{2d}\int_{t/n}^{(t+1)/n}\sum_{i=1}^{d}\sum_{j=1}^{d}\lambda_{H,ik}\sigma_{U,s\wedge\tau,ij}^{2}\mathsf{d}s\right)\right] \\
\leq\ & \mathsf{E}\left[\exp\left(\frac{n\|\boldsymbol{\Lambda}_H\|_{\max}}{2}\int_{t/n}^{(t+1)/n}Q_{s\wedge\tau}\mathsf{d}s\right)\right] < \infty,
\end{aligned}$$

for $k = 1, ..., K_H$. Then using Bonferroni and Markov inequalities again, we can prove the result.

(vi) It can be proved by using Markov inequality and (1.2.10).

□

Lemma A.2.8. *(i) Under Assumption 1.A, we have* $\|\mathscr{F} - \mathbf{1}_n\boldsymbol{F}_0^{\intercal}\|_{\max} = O_P(\log n)$;
(ii) Under Assumption 1.F, we have and $\|\mathscr{G} - \mathbf{1}_n\boldsymbol{G}_0^{\intercal}\|_{\max} = O_P((\log n)^{1/\gamma_2})$;
(iii) Under Assumptions 1.A and 1.E, we have $\|\mathscr{U} - \mathbf{1}_n\boldsymbol{U}_0^{\intercal}\|_{\max} = O_P(\log(nd))$;
(iv) Under Assumptions 1.E and 1.F, we have $\|\mathscr{V} - \mathbf{1}_n\boldsymbol{V}_0^{\intercal}\|_{\max} = O_P(\log(nd)^{1/\gamma_2})$;
(v) Under Assumptions 1.A, 1.D and 1.E, we have $d^{-1/2}\|(\mathscr{U} - \mathbf{1}_n\boldsymbol{U}_0^{\intercal})\boldsymbol{\Lambda}_H\|_{\max} = O_P((\log(nd))$;
(vi) Under Assumptions 1.D–1.F, we have $d^{-1/2}\|(\mathscr{V} - \mathbf{1}_n\boldsymbol{V}_0^{\intercal})\boldsymbol{\Lambda}_H\|_{\max} = O_P((\log n)^{1/\gamma_3})$.

Proof. The proof is similar to Lemma A.2.7 so we omit the proof. □

Appendix B

Appendix to Chapter 2

B.1 Proofs of Theorems 2.4.1–2.4.4

Proof of Theorem 2.4.1. The main idea to be used in this proof is similar to that in Bickel *et al.* (2009), Lian (2012) and Li *et al.* (2015a) 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.2.3 in Appendix B.2). In fact, we next prove a strengthened version of (2.4.4) which also includes a uniform consistency of the derivative function estimates:

$$\max_{1\leq i\leq d}\max_{1\leq t\leq n}\left(\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\|+h\left\|\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)-\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\right\|\right)=O_P\left(\sqrt{s}\lambda_1\right). \tag{B.1.1}$$

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

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

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

$$\sum_{j\notin\mathscr{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\notin\mathscr{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\leq 2\left(\sum_{j\in\mathscr{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\in\mathscr{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\right), \tag{B.1.2}$$

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

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

for any $i=1,\cdots,d$ and $t=1,\cdots,n$, where $\mathscr{L}_i^*(\boldsymbol{\alpha},\boldsymbol{\beta}\mid\tau_t)$ is defined in (2.3.4). Then, we readily

have that

$$\begin{aligned}
&\mathscr{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}'_{i\bullet}(\tau_t) \mid \tau_t\right) - \mathscr{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t), \widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t) \mid \tau_t\right) \\
&\geq \lambda_1 \left[\sum_{j=1}^{d} |\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j=1}^{d} |\widetilde{\alpha}'_{i,j}(\tau_t)| - \sum_{j=1}^{d} |\alpha_{i,j}(\tau_t)| - h\sum_{j=1}^{d} |\alpha'_{i,j}(\tau_t)|\right]. \qquad \text{(B.1.3)}
\end{aligned}$$

Let

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

Note that

$$\begin{aligned}
&\mathscr{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}'_{i\bullet}(\tau_t) \mid \tau_t\right) - \mathscr{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t), \widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t) \mid \tau_t\right) \\
= \;& 2\left[L_{i,0}^\intercal(\tau_t)\delta_i(\tau_t) + L_{i,1}^\intercal(\tau_t)\delta'_i(\tau_t)\right] - \frac{1}{n}\sum_{s=1}^{n}\left\{\left[\delta_i(\tau_t) + \delta'_i(\tau_t)\left(\frac{\tau_s - \tau_t}{h}\right)\right]^\intercal X_{s-1}\right\}^2 K_h(\tau_s - \tau_t) \\
\leq \;& 2\left[L_{i,0}^\intercal(\tau_t)\delta_i(\tau_t) + L_{i,1}^\intercal(\tau_t)\delta'_i(\tau_t)\right], \qquad \text{(B.1.4)}
\end{aligned}$$

where $L_{i,0}(\tau_t)$ and $L_{i,1}(\tau_t)$ are defined in Appendix B.2. By Lemma B.2.3, we may show that

$$\left|L_{i,0}^\intercal(\tau_t)\delta_i(\tau_t) + L_{i,1}^\intercal(\tau_t)\delta'_i(\tau_t)\right| \leq O_P\left(\zeta_{n,d}\right) \cdot \left(\sum_{j=1}^{d} |\delta_{i,j}(\tau_t)| + \sum_{j=1}^{d} |\delta'_{i,j}(\tau_t)|\right) \qquad \text{(B.1.5)}$$

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

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

$$\begin{aligned}
&\lambda_1 \left[\sum_{j=1}^{d} |\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j=1}^{d} |\widetilde{\alpha}'_{i,j}(\tau_t)| - \sum_{j=1}^{d} |\alpha_{i,j}(\tau_t)| - h\sum_{j=1}^{d} |\alpha'_{i,j}(\tau_t)|\right] \\
= \;& \lambda_1 \left[\sum_{j\in\mathscr{J}_i(\tau_t)} \left(|\widetilde{\alpha}_{i,j}(\tau_t)| - |\alpha_{i,j}(\tau_t)|\right) + h\sum_{j\in\mathscr{J}'_i(\tau_t)} \left(|\widetilde{\alpha}'_{i,j}(\tau_t)| - |\alpha'_{i,j}(\tau_t)|\right)\right] + \\
& \lambda_1 \left[\sum_{j\notin\mathscr{J}_i(\tau_t)} |\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j\notin\mathscr{J}'_i(\tau_t)} |\widetilde{\alpha}'_{i,j}(\tau_t)|\right] \\
\geq \;& -\lambda_1 \left(\sum_{j\in\mathscr{J}_i(\tau_t)} |\delta_{i,j}(\tau_t)| + \sum_{j\in\mathscr{J}'_i(\tau_t)} |\delta'_{i,j}(\tau_t)|\right) + \lambda_1 \left(\sum_{j\notin\mathscr{J}_i(\tau_t)} |\delta_{i,j}(\tau_t)| + \sum_{j\notin\mathscr{J}'_i(\tau_t)} |\delta'_{i,j}(\tau_t)|\right). \qquad \text{(B.1.6)}
\end{aligned}$$

By (B.1.3)–(B.1.6) and the condition $\zeta_{n,d} = o(\lambda_1)$ in Assumption 2.C(i), we complete the proof of (B.1.2).

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

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

where M is a positive constant which may be sufficiently large. Note that for any $i = 1, \cdots, d$, $t = 1, \cdots, n$, and $u \in \mathscr{B}_i(\tau_t; M)$,

$$\mathscr{L}_i^* \left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, \boldsymbol{\alpha}_{i\bullet}'(\tau_t) + \sqrt{s}\lambda_1 u_2/h \mid \tau_t \right) - \mathscr{L}_i^* \left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}_{i\bullet}'(\tau_t) \mid \tau_t \right) = \sum_{k=1}^{3} \Xi_{i,k}(\tau_t), \quad \text{(B.1.7)}$$

where

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

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

$$\Xi_{i,1}(\tau_t) = -2\sqrt{s}\lambda_1 u^\intercal L_i(\tau_t) + s\lambda_1^2 u^\intercal \boldsymbol{\Psi}(\tau_t) u, \quad \text{(B.1.8)}$$

where $L_i(\tau) = \left[L_{i,0}^\intercal(\tau), L_{i,1}^\intercal(\tau) \right]^\intercal$, and $\boldsymbol{\Psi}(\tau)$ is defined in (2.4.2). By the definition of $\mathscr{B}_i(\tau_t; M)$, Lemma B.2.3 and the Cauchy-Schwarz inequality, we have

$$\max_{1 \le i \le d} \left| \sqrt{s}\lambda_1 u^\intercal L_i(\tau_t) \right| = o_P\left(s\lambda_1^2\right) \cdot \|u\|. \quad \text{(B.1.9)}$$

By (B.1.8), (B.1.9) and the uniform restricted eigenvalue condition (2.4.3), when n is sufficiently large and M is chosen to be large enough, we have

$$\min_{1 \le i \le d} \min_{1 \le t \le n} \inf_{u \in \mathscr{B}_i(\tau_t; M)} u^\intercal \Xi_{i,1}(\tau_t) = s\lambda_1^2 u^\intercal \boldsymbol{\Psi}(\tau_t) u (1 + o_P(1)) > \frac{1}{2}\kappa_0 s\lambda_1^2 \|u\|^2, \quad w.p.a.1. \quad \text{(B.1.10)}$$

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

$$\begin{aligned} \Xi_{i,2}(\tau_t) &= \lambda_1 \left(\sum_{j=1}^{d} |\alpha_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{1,j}| - \sum_{j=1}^{d} |\alpha_{i,j}(\tau_t)| \right) \\ &= \lambda_1 \sum_{j \in \mathscr{J}_i(\tau_t)} \left[|\alpha_{i,j}(\tau_t) + \sqrt{s}\lambda_1 u_{1,j}| - |\alpha_{i,j}(\tau_t)| \right] + \lambda_1 \sum_{j \notin \mathscr{J}_i(\tau_t)} |\sqrt{s}\lambda_1 u_{1,j}| \end{aligned}$$

$$= O\left(s\lambda_1^2\right)\cdot\|u_1\| + \lambda_1 \sum_{j\notin\mathscr{J}_i(\tau_t)} |\sqrt{s}\lambda_1 u_{1,j}| = O\left(s\lambda_1^2\right)\cdot\|u_1\|, \tag{B.1.11}$$

and similarly,

$$\Xi_{i,3}(\tau_t) = O\left(s\lambda_1^2\right)\cdot\|u_2\| + \lambda_1 \sum_{j\notin\mathscr{J}_i'(\tau_t)} |\sqrt{s}\lambda_1 u_{2,j}| = O\left(s\lambda_1^2\right)\cdot\|u_2\|, \tag{B.1.12}$$

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

With (B.1.7) and (B.1.10)–(B.1.12), letting M be large enough, we can prove that the leading term of

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

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

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

which, together with (B.1.2), completes the proof of (B.1.1). ■

Proof of Theorem 2.4.2. Define

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

where

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

From the KKT condition (e.g., Fan and Lv, 2011; Fan *et al.*, 2014b; Li *et al.*, 2015a), the oracle estimate $\left(\widehat{\mathbf{A}}_i^o, \widehat{\mathbf{B}}_i^o\right)$ is the unique minimiser to the objective function $\mathscr{Q}_i(\mathbf{A},\mathbf{B})$ if

$$\mathbf{L}_{i,j}^{\alpha} - \mathbf{P}_{i,j}^{\alpha} = \mathbf{0}_n \ \text{ for } \ j \in \mathscr{J}_i, \quad \mathbf{L}_{i,j}^{\beta} - \mathbf{P}_{i,j}^{\beta} = \mathbf{0}_n \ \text{ for } \ j \in \mathscr{J}_i', \tag{B.1.13}$$

$$\max_{j\in\overline{\mathscr{J}}_i}\left\|\mathbf{L}_{i,j}^{\alpha}\right\| < \min_{j\in\overline{\mathscr{J}}_i} p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right), \quad \max_{j\in\overline{\mathscr{J}}_i}\left\|\mathbf{L}_{i,j}^{\beta}\right\| < \min_{j\in\overline{\mathscr{J}}_i} p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right), \tag{B.1.14}$$

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 (B.1.13) automatically hold by the definition of the oracle estimates $\widehat{\mathbf{A}}_i^o$ and $\widehat{\mathbf{B}}_i^o$. It remains to prove (B.1.14). We next only show the proof of the first assertion in (B.1.14) as the proof of the second one is analogous. By Theorem 2.4.1 and the condition of $(ns)^{1/2}\lambda_1 = o(\lambda_2)$ in Assumption 2.D(i), we may show that $\min_{j\in\overline{\mathscr{I}}_i} p'_{\lambda_2}(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|) = \lambda_2$ *w.p.a.1.* Meanwhile, by Lemmas B.2.3 and B.2.4 as well as Assumption 2.D(i), we may prove that

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

when $\mathbf{A} = \widehat{\mathbf{A}}_i^o$ and $\mathbf{B} = \widehat{\mathbf{B}}_i^o$, leading to the first assertion in (B.1.14). Then, the mean squared convergence result (2.4.8) follows from Lemma B.2.4. ■

Proof of Corollary 2.4.3. By Theorem 2.4.2 and Assumption 2.D(ii), we may show that

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

and

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

leading to (2.4.9). ■

Proof of Theorem 2.4.4. By Lemma B.2.5 in Appendix B.2, we have

$$\sup_{0\leq\tau\leq 1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{\max} = O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right). \tag{B.1.15}$$

By (B.1.15), the sparsity assumption (2.3.7) and the inequality: $\|\mathbf{W}_1\mathbf{W}_2\|_{\max} \leq \|\mathbf{W}_1\|_1\|\mathbf{W}_2\|_{\max}$ for any two square matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ with the same size,

$$\begin{aligned}
\sup_{0\leq\tau\leq 1}\left\|\mathbf{I}_d - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{\max} &= \sup_{0\leq\tau\leq 1}\left\|\boldsymbol{\Sigma}(\tau)\boldsymbol{\Omega}(\tau) - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{\max} \\
&\leq \sup_{0\leq\tau\leq 1}\|\boldsymbol{\Omega}(\tau)\|_1\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{\max} \\
&\leq C_2\sup_{0\leq\tau\leq 1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{\max} \\
&= O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right),
\end{aligned} \tag{B.1.16}$$

where C_2 is defined in (2.3.7). By (B.1.16), the triangle inequality, Assumption 2.E(ii) and the definition of the time-varying CLIME estimate, we readily have that

$$\sup_{0\leq\tau\leq 1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau)\right]\right\|_{\max}$$

$$\begin{aligned} &\leq \sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Sigma}}(\tau)\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{I}_d\right\|_{\max}+\sup_{0\leq\tau\leq 1}\left\|\mathbf{I}_d-\widehat{\mathbf{\Sigma}}(\tau)\mathbf{\Omega}(\tau)\right\|_{\max} \\ &\leq \lambda_3+O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)=O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right). \end{aligned} \tag{B.1.17}$$

By Lemma 1 in Cai *et al.* (2011), $\left\|\widetilde{\mathbf{\Omega}}(\tau)\right\|_1 \leq \|\mathbf{\Omega}(\tau)\|_1 \leq C_2$ uniformly over $0\leq\tau\leq 1$. Then, by (B.1.16) and (B.1.17), we readily have that

$$\begin{aligned} &\sup_{0\leq\tau\leq 1}\left\|\mathbf{\Sigma}(\tau)\left[\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right]\right\|_{\max} \\ \leq\ &\sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Sigma}}(\tau)\left[\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right]\right\|_{\max}+\sup_{0\leq\tau\leq 1}\left\|\left[\widehat{\mathbf{\Sigma}}(\tau)-\mathbf{\Sigma}(\tau)\right]\left[\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right]\right\|_{\max} \\ \leq\ &O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right)+2C_2\sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Sigma}}(\tau)-\mathbf{\Sigma}(\tau)\right\|_{\max}=O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right). \end{aligned} \tag{B.1.18}$$

Using the assumption $\|\mathbf{\Omega}(\tau)\|_1 \leq C_2$ again and (B.1.18), we have

$$\begin{aligned} \sup_{0\leq\tau\leq 1}\left\|\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\|_{\max} &\leq \sup_{0\leq\tau\leq 1}\|\mathbf{\Omega}(\tau)\|_1\left\|\mathbf{\Sigma}(\tau)\left[\widetilde{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right]\right\|_{\max} \\ &= O_P\left(\nu_{n,d}^{*}+\nu_{n,d}^{\diamond}\right). \end{aligned} \tag{B.1.19}$$

By (B.1.19) and the definition of $\widehat{\mathbf{\Omega}}(\tau)$ in (2.3.10), we prove (2.4.10).

We next give the proof of (2.4.11). By Lemma 1 in Cai *et al.* (2011), we have

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

Noting that

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

we have

$$\begin{aligned} \sup_{0\leq\tau\leq 1}\left\|\widehat{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\|_{O} &\leq \sup_{0\leq\tau\leq 1}\max_{1\leq i\leq d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)| \\ &\leq 2\sup_{0\leq\tau\leq 1}\max_{1\leq i\leq d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)+ \\ &\quad 2\sup_{0\leq\tau\leq 1}\max_{1\leq i\leq d}\sum_{j=1}^{d}|\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|\leq\lambda_3\right) \end{aligned}$$

$$=: \quad \Delta_1 + \Delta_2. \tag{B.1.20}$$

Define an event

$$\mathscr{E}_\epsilon = \left\{ \sup_{0\le\tau\le1} \left\| \widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau) \right\|_{\max} \le c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \right\},$$

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

$$\Delta_1 \le c_\epsilon(\nu^\diamond_{n,d} + \nu^*_{n,d}) \sup_{0\le\tau\le1} \left[\max_{1\le i\le d} \sum_{j=1}^{d} I\left(|\widehat{\omega}_{ij}(\tau)| > \lambda_3 \right) \right]. \tag{B.1.21}$$

Note that on $\mathscr{E}_\epsilon$,

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

Choosing $C_3 = 2c_\epsilon$ in Assumption 2.E(ii), the event $\{|\widehat{\omega}_{ij}(\tau)| > \lambda_3\}$ implies that $\left\{ |\omega_{ij}(\tau)| > c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \right\}$ holds. Then, by (2.3.7) and (B.1.21), we may show that on $\mathscr{E}_\epsilon$,

$$\begin{aligned}
\Delta_1 &\le c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \left[\sup_{0\le\tau\le1} \max_{1\le i\le d} \sum_{j=1}^{d} I\left(|\omega_{ij}(\tau)| > c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \right) \right] \\
&\le c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \left[\sup_{0\le\tau\le1} \max_{1\le i\le d} \sum_{j=1}^{d} \frac{|\omega_{ij}(\tau)|^q}{c_\epsilon^q \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right)^q} \right] \\
&= O_P\left(\xi_d \cdot \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right)^{1-q} \right).
\end{aligned} \tag{B.1.22}$$

On the other hand, by the triangle inequality,

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

on $\mathscr{E}_\epsilon$. Hence, we readily show that $\{|\widehat{\omega}_{ij}(\tau)| \le \lambda_3\}$ indicates $\left\{ |\omega_{ij}(\tau)| \le 3c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \right\}$. Then, by (2.3.7) again, we have

$$\begin{aligned}
\Delta_2 &\le \sup_{0\le\tau\le1} \max_{1\le i\le d} \sum_{j=1}^{d} |\omega_{ij}(\tau)| \, I\left(|\omega_{ij}(\tau)| \le 3c_\epsilon \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right) \right) \\
&\le (3c_\epsilon)^{1-q} \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right)^{1-q} \sup_{0\le\tau\le1} \max_{1\le i\le d} \sum_{j=1}^{d} |\omega_{ij}(\tau)|^q \\
&= O_P\left(\xi_d \left(\nu^\diamond_{n,d} + \nu^*_{n,d} \right)^{1-q} \right).
\end{aligned} \tag{B.1.23}$$

The proof of (2.4.11) can be completed by (B.1.20), (B.1.22) and (B.1.23).

Following the proof of (2.4.11), we also have

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

which, together with the following inequalities:

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

leads to (2.4.12). The proof of Theorem 2.4.4 is completed. ■

Proof of Corollary 2.4.5. By (2.4.10) in Theorem 2.4.4 and the condition of $\min_{(i,j)\in\mathbb{E}^P}\min_{1\leq t\leq n}|\omega_{ij}(\tau_t)|\gg\lambda_3$, we have

$$\mathsf{P}\left(\min_{(i,j)\in\mathbb{E}_n^P}\min_{1\leq t\leq n}|\widehat{\omega}_{ij}(\tau_t)|\geq\lambda_3>0\right)\to 1. \tag{B.1.24}$$

Letting $\mathscr{E}_\epsilon$ and c_ϵ be defined as in the proof of Theorem2.4.4 and choosing $C_3=2c_\epsilon$ in Assumption 2.E(ii), we may prove that

$$\max_{(i,j)\notin\mathbb{E}_n^P}\max_{1\leq t\leq n}|\widehat{\omega}_{ij}(\tau_t)|\leq c_\epsilon(\nu_{n,d}^{*}+\nu_{n,d}^{\diamond})<\lambda_3 \tag{B.1.25}$$

conditional on $\mathscr{E}_\epsilon$. By virtue of (B.1.24) and (B.1.25), letting $\epsilon\to 0$, we prove (2.4.13). ■

B.2 Technical lemmas

In this appendix, we give some technical lemmas which are crucial to proofs of the main theoretical results in Appendix B.1. 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.2.1. *Suppose that Assumption 2.A is satisfied. Let*

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

where ι_1 and ρ are defined in Assumption 2.A, and C_1 is defined in (2.2.4). For any d-dimensional vector u satisfying $\|u\|=1$,

$$\max_{1\leq t\leq n}\mathsf{E}\left[\exp\left\{\iota_2\left(u^{\intercal}e_t\right)^2\right\}\right]\leq C_0<\infty, \tag{B.2.1}$$

and

$$\max_{1\leq t\leq n}\max_{1\leq i\leq d}\mathsf{E}\left[\exp\left\{\iota_3x_{t,i}^2\right\}\right]\leq C_0^{1/(1-\rho)}<\infty, \tag{B.2.2}$$

where C_0 is a positive constant defined in Assumption 2.A(iii).

Proof of Lemma B.2.1. Writing $u_t^\intercal = u^\intercal \mathbf{\Sigma}_t^{1/2}$ and using Assumption 2.A(ii)(iii), we may show that

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

completing the proof of (B.2.1).

By the time-varying linear process representation (2.2.3), we have

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

where $\Phi_{t,k,i}^\intercal$ is the i-th row vector of $\mathbf{\Phi}_{t,k}$. Without loss of generality, assume (2.2.4) for all $k \ge 0$. Letting $u_{t,k,i} = \Phi_{t,k,i}/\|\Phi_{t,k,i}\|$ and noting that

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

we may show that

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

which, together with the independence assumption over e_t and (B.2.1), indicates that

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

$$\leq \prod_{k=0}^{\infty} C_0^{\rho^k} = C_0^{1/(1-\rho)},$$

completing the proof of (B.2.2). ■

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

Lemma B.2.2. *Let $(z_t, \mathscr{F}_t)_{t\geq 1}$ be a sequence of martingale differences and $\sigma_n^2 = \sum_{t=1}^n \mathsf{E}(z_t^2|\mathscr{F}_{t-1})$. Suppose that there exists a constant $a > 0$ such that $\mathsf{P}(|z_t| \leq a|\mathscr{F}_{t-1}) = 1$ for all $t \geq 2$. Then, for all $x, y > 0$,*

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

Define

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

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.2.3 below gives the uniform asymptotic orders for the kernel-weighted quantities $L_{i,k}(\cdot)$, $k = 0, 1$.

Lemma B.2.3. *Suppose that Assumptions 2.A and 2.B are satisfied. Then we have*

$$\max_{1\leq i\leq d}\max_{1\leq t\leq n}|L_{i,k}(\tau_t)|_{\max} = O_P\left(\zeta_{n,d}\right), \quad k = 0, 1, \tag{B.2.3}$$

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

Proof of Lemma B.2.3. We only prove (B.2.3) for $k = 0$ as the proof is analogous for $k = 1$. Noting that

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

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.2.3) with $k = 0$, it is sufficient to show that

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^n e_{l,i}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left((nh)^{-1/2}\log(n \vee d)\right) \tag{B.2.4}$$

and

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^n b_{l,i}^\intercal(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left(sh^2\log(n \vee d)\right). \tag{B.2.5}$$

Define

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

and

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

where ι_2 and ι_3 are defined in Lemma B.2.1. Then, we have the following decomposition:

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

By the Bonferroni and Markov inequalities as well as (B.2.1), for any $\epsilon > 0$, we have

$$\begin{aligned}&\mathsf{P}\left(\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| > \epsilon(nh)^{-1/2}\log(n\vee d)\right) \\ \leq\ &\mathsf{P}\left(\max_{1\leq i\leq d}\max_{1\leq t\leq n}|e_{t,i}| > 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right) \\ \leq\ &\sum_{i=1}^{d}\sum_{t=1}^{n}\mathsf{P}\left(|e_{t,i}| > 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right) \\ \leq\ &\sum_{i=1}^{d}\sum_{t=1}^{n}(n\vee d)^{-4}\mathsf{E}\left(\exp\left\{\iota_2 e_{t,i}^2\right\}\right) \\ \leq\ &M(n\vee d)^{-2} = o(1). \end{aligned} \tag{B.2.6}$$

Hence, we have

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_t-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.2.7}$$

Following the proof of (B.2.7), we also have

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^{n}\overline{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right) \tag{B.2.8}$$

and

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.2.9}$$

By the Cauchy-Schwarz and Markov inequalities and (B.2.1), we may show that

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

$$\begin{aligned} &\leq \left[\mathsf{E}\left(|e_{l,i}|^2\right)\right]^{1/2}\left[\mathsf{E}\left(\exp\left\{\iota_2 e_{l,i}^2\right\}\right)\right]^{1/2}(n\vee d)^{-2} \\ &\leq M(n\vee d)^{-2}, \end{aligned}$$

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

$$\begin{aligned} \left|\frac{1}{n}\sum_{l=1}^{n}\mathsf{E}\left[\overline{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathscr{F}_{l-1}(X)\right]\right| &= \left|\frac{1}{n}\sum_{l=1}^{n}\mathsf{E}\left[\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathscr{F}_{l-1}(X)\right]\right| \\ &= O_P\left((n\vee d)^{-2}\sqrt{\log(n\vee d)}\right) \\ &= o_P\left((nh)^{-1/2}\log(n\vee d)\right), \end{aligned} \tag{B.2.10}$$

where $\mathscr{F}_l(X)=\sigma(X_t:\ t\leq l)$. With (B.2.7)–(B.2.10), we readily have that

$$\frac{1}{n}\sum_{l=1}^{n}e_{l,i}x_{l-1,j}K_h(\tau_l-\tau_t)=\frac{1}{n}\sum_{l=1}^{n}\left\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathscr{F}_{l-1}(X)\right]\right\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+o_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.2.11}$$

By the Bonferroni inequality and the Bernstein inequality in Lemma B.2.2, we prove that

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

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

$$\max_{1\leq i\leq d}\max_{1\leq j\leq d}\max_{1\leq t\leq n}\left|\frac{1}{n}\sum_{l=1}^{n}\left\{\overline{e}_{l,i}-\mathsf{E}\left[\overline{e}_{l,i}|\mathscr{F}_{l-1}(X)\right]\right\}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|=O_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.2.12}$$

By virtue of (B.2.11) and (B.2.12), we complete the proof of (B.2.4).

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

$$\begin{aligned} &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l-\tau_t) \\ =\ &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\overline{X}_{l-1}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\overline{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t) \\ &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\widetilde{X}_{l-1}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\widetilde{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t). \end{aligned} \tag{B.2.13}$$

Similarly to the proof of (B.2.11), we may show that the last three terms on the right side of (B.2.13) are of order $o_P\left(h^2\log(n\vee d)\right)$. By Assumption 2.A(i) and the Taylor expansion of $\boldsymbol{\alpha}_{i\bullet}(\cdot)$, we can

prove that the first term on the right side of (B.2.13) is of order $O_P\left(sh^2\log(n\vee d)\right)$ uniformly over i,j,t. The proof of (B.2.5) is completed. ■

Lemma B.2.4 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 (2.4.6) and (2.4.7) of Section 2.4.2.

Lemma B.2.4. *Suppose Assumptions 2.A–2.D are satisfied. Then we have*

$$\max_{1\le i\le d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}^o(\tau_t)-\alpha_{i,j}(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2\right), \tag{B.2.14}$$

and

$$\max_{1\le i\le d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}^{\prime o}(\tau_t)-\alpha_{i,j}'(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2h^{-2}\right). \tag{B.2.15}$$

Proof of Lemma B.2.4. For any $1\le i\le d$, 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 $j\in\overline{\mathscr{F}}_i$, and $w_t^o=\left(w_{1|t}^o,\cdots,w_{d|t}^o\right)^\intercal$ with $w_{j|t}^o=0$ for $j\in\overline{\mathscr{F}}_i'$. Define

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

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

$$\mathscr{Q}_i\left(\mathbf{A}_i+\sqrt{\zeta_{n,d}^*}\mathbf{V}^o,\mathbf{B}_i+\sqrt{\zeta_{n,d}^*}\mathbf{W}^o/h\right)-\mathscr{Q}_i\left(\mathbf{A}_i,\mathbf{B}_i\right)=\Pi_{i,1}^o+\Pi_{i,2}^o+\Pi_{n,3}^o, \tag{B.2.16}$$

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

$$\Pi_{i,1}^o\ =\ \sum_{t=1}^{n}\left[\mathscr{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t)+\sqrt{\zeta_{n,d}^*}v_t^o,\boldsymbol{\alpha}_{i\bullet}'(\tau_t)+\sqrt{\zeta_{n,d}^*}w_t^o/h\mid\tau_t\right)-\mathscr{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}_{i\bullet}'(\tau_t)\mid\tau_t\right)\right],$$

$$\Pi_{i,2}^o = \sum_{j=1}^{d} p'_{\lambda_2}\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right)\left(\left\|\boldsymbol{\alpha}_{i,j}+\sqrt{\zeta_{n,d}^*}\mathbf{v}_j^o\right\|-\|\boldsymbol{\alpha}_{i,j}\|\right),$$

$$\Pi_{i,3}^o = \sum_{j=1}^{d} p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right)\left(\left\|h\boldsymbol{\alpha}'_{i,j}+\sqrt{\zeta_{n,d}^*}\mathbf{w}_j^o\right\|-\left\|h\boldsymbol{\alpha}'_{i,j}\right\|\right),$$

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

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

$$\Pi_{i,1}^o = -2\sqrt{\zeta_{n,d}^*}\sum_{t=1}^{n}\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]L_i(\tau_t)+\zeta_{n,d}^*\sum_{t=1}^{n}\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]\boldsymbol{\Psi}(\tau_t)\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]^\intercal. \quad \text{(B.2.17)}$$

By the definition of $\mathscr{B}_i^*(M_*)$, Lemma B.2.3 and the Cauchy-Schwarz inequality, we prove

$$\left|\sum_{t=1}^{n}\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]L_i(\tau_t)\right| = O_P\left(\sqrt{\zeta_{n,d}^*}n^{1/2}\right)\cdot\|\mathbf{U}^o\| \quad \text{(B.2.18)}$$

uniformly over i. By the uniform restricted eigenvalue condition in Assumption 2.C(ii), we have

$$\sum_{t=1}^{n}\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]\boldsymbol{\Psi}(\tau_t)\left[(v_t^o)^\intercal,(w_t^o)^\intercal\right]^\intercal \geq \kappa_0\sum_{t=1}^{n}\left(\|v_t^o\|^2+\|w_t^o\|^2\right) = n\kappa_0 M_* \quad \text{(B.2.19)}$$

for $\mathbf{U}^o\in\mathscr{B}_i^*(M_*)$. Combining (B.2.17)–(B.2.19) and letting $M_*>0$ be sufficiently large, we have

$$\min_{1\leq i\leq d}\Pi_{i,1}^o \geq \kappa_0\zeta_{n,d}^*\|\mathbf{U}^o\|^2+O_P\left(\zeta_{n,d}^*n^{1/2}\right)\cdot\|\mathbf{U}^o\| > \frac{\kappa_0}{2}\zeta_{n,d}^*\|\mathbf{U}^o\|^2 \quad w.p.a.1. \quad \text{(B.2.20)}$$

On the other hand, by Theorem 2.4.1 and Assumption 2.D(ii), we have

$$\mathsf{P}\left(\min_{1\leq i\leq d}\min_{j\in\mathscr{I}_i}\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|>a_0\lambda_2\right)\to 1,$$

and

$$\mathsf{P}\left(\min_{1\leq i\leq d}\min_{j\in\mathscr{I}_i'}\widetilde{D}_{i,j}>a_0\lambda_2\right)\to 1.$$

As $\alpha_{i,j}(\tau_t)=0$ and $u_{1,j}^o=0$ for $j\in\overline{\mathscr{I}}_i(\tau_t)$, we thus have

$$\Pi_{i,2}(\tau_t) = \sum_{j\in\mathscr{I}_i(\tau_t)}p'_{\lambda_2}\left(|\widetilde{\alpha}_{i,j}(\tau_t)|\right)\left(\left|\alpha_{i,j}(\tau_t)+\sqrt{\zeta_{n,d}^*}(\tau_t)u_{1,j}^o\right|-|\alpha_{i,j}(\tau_t)|\right)=0 \quad w.p.a.1, \quad \text{(B.2.21)}$$

and similarly

$$\Pi_{i,3}(\tau_t) = \sum_{j\in\mathscr{I}_i(\tau_t)}p'_{\lambda_2}\left(|\widetilde{\alpha}'_{i,j}(\tau_t)|\right)\left(\left|h\alpha'_{i,j}(\tau_t)+\sqrt{\zeta_{n,d}^*}(\tau_t)u_{2,j}^o\right|-\left|h\alpha'_{i,j}(\tau_t)\right|\right)=0 \quad w.p.a.1. \quad \text{(B.2.22)}$$

Hence, by (B.2.20)–(B.2.22), letting $M_* > 0$ be large enough, we can prove that

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

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

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

for any $1\le i\le d$. The proof of Lemma B.2.4 is completed. ■

Lemma B.2.5 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.2.5. *Suppose that Assumptions 2.A–2.D are satisfied. Then we have*

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}|\widehat{\sigma}_{ij}(\tau)-\sigma_{ij}(\tau)|=O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{*}\right),\tag{B.2.23}$$

where $\sigma_{ij}(\tau)$ is the (i,j)-entry of $\mathbf{\Sigma}(\tau)$, $\nu_{n,d}^{\diamond}$ and $\nu_{n,d}^{*}$ are defined in Assumption 2.E(ii).

Proof of Lemma B.2.5. By the definition of $\widehat{\sigma}_{ij}(\tau)$ in (2.3.8), we have

$$\begin{aligned}\widehat{\sigma}_{ij}(\tau)-\sigma_{ij}(\tau) &= \left\{\frac{\sum_{t=1}^n\varpi_{n,t}(\tau)e_{t,i}e_{t,j}}{\sum_{t=1}^n\varpi_{n,t}(\tau)}-\sigma_{ij}(\tau)\right\}+\left\{\frac{\sum_{t=1}^n\varpi_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)e_{t,j}}{\sum_{t=1}^n\varpi_{n,t}(\tau)}+\right.\\ &\quad\left.\frac{\sum_{t=1}^n\varpi_{n,t}(\tau)e_{t,i}\left(\widehat{e}_{t,j}-e_{t,j}\right)}{\sum_{t=1}^n\varpi_{n,t}(\tau)}+\frac{\sum_{t=1}^n\varpi_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)\left(\widehat{e}_{t,j}-e_{t,j}\right)}{\sum_{t=1}^n\varpi_{n,t}(\tau)}\right\}\\ &=: \chi_{ij}^{\diamond}(\tau)+\chi_{ij}^{*}(\tau).\end{aligned}\tag{B.2.24}$$

We first prove that

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\chi_{ij}^{\diamond}(\tau)\right|=O_P\left(\nu_{n,d}^{\diamond}\right).\tag{B.2.25}$$

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

$$\begin{aligned}&\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\frac{\sum_{t=1}^n\varpi_{n,t}(\tau)\sigma_{ij}(\tau_t)}{\sum_{t=1}^n\varpi_{n,t}(\tau)}-\sigma_{ij}(\tau)\right|\\ \le\ &\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\sigma_{ij}''(\tau)\right|\cdot\left|\frac{\sum_{t=1}^n(\tau_t-\tau)^2\varpi_{n,t}(\tau)}{\sum_{t=1}^n\varpi_{n,t}(\tau)}\right|\\ \le\ &M\sup_{0\le\tau\le1}\left|\frac{\sum_{t=1}^n(\tau_t-\tau)^2\varpi_{n,t}(\tau)}{\sum_{t=1}^n\varpi_{n,t}(\tau)}\right|=O\left(b^2\right).\end{aligned}\tag{B.2.26}$$

Let $\overline{e}_{t,i}$ and $\widetilde{e}_{t,j}$ be defined as in the proof of Lemma B.2.3. Then, we have

$$\begin{aligned}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right) e_{t,i}e_{t,j} &= \sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right) \overline{e}_{t,i}\overline{e}_{t,j} + \sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right) \overline{e}_{t,i}\widetilde{e}_{t,j} + \\ &\quad \sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right) \widetilde{e}_{t,i}\overline{e}_{t,j} + \sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right) \widetilde{e}_{t,i}\widetilde{e}_{t,j}. \qquad \text{(B.2.27)}\end{aligned}$$

Following the proof of (B.2.11), the first term on the right side of (B.2.27) is the asymptotic leading term. Consider covering the closed interval $[0,1]$ by some disjoint intervals $\mathscr{I}_k$, $k = 1, \cdots, N$, 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.B(i), we have

$$\begin{aligned}&\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| \\ \le\ &\max_{1\le i,j\le d}\max_{1\le k\le N}\left|\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| + \\ &\max_{1\le i,j\le d}\max_{1\le k\le N}\sup_{\tau\in\mathscr{I}_k}\left|\frac{1}{nb}\sum_{k=1}^{n}\left[K\left(\frac{\tau_t-\tau}{b}\right) - K\left(\frac{\tau_t-\tau_k^*}{b}\right)\right]\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| \\ \le\ &\max_{1\le i,j\le d}\max_{1\le k\le N}\left|\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| + O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \qquad \text{(B.2.28)}\end{aligned}$$

By the Bonferroni inequality and Lemma B.2.2 as well as the condition $nb/[\log(n\vee d)]^3 \to \infty$ in Assumption 2.E(i), we may show that

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

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

$$\max_{1\le i,j\le d}\max_{1\le k\le N}\left|\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| = O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \qquad \text{(B.2.29)}$$

Combining (B.2.28) and (B.2.29), we can prove that

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{\tau_t-\tau}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j} - \mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| = O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \qquad \text{(B.2.30)}$$

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.2.1) in Lemma B.2.1,

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

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

$$\max_{1\leq i,j\leq d}\sup_{0\leq\tau\leq 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)-\sigma_{ij}(\tau_t)\right]\right|=o_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.2.31}$$

With (B.2.27), (B.2.30) and (B.2.31), we can prove that

$$\max_{1\leq i,j\leq d}\sup_{0\leq\tau\leq 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.2.32}$$

Analogously, we also have

$$\max_{1\leq i,j\leq d}\sup_{0\leq\tau\leq 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K_1\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.2.33}$$

Using (B.2.32), (B.2.33) and the definition of $\varpi_{n,t}(\tau)$, we may show that

$$\max_{1\leq i,j\leq d}\sup_{0\leq\tau\leq 1}\left|\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right), \tag{B.2.34}$$

which, together with (B.2.26), leads to (B.2.25).

Using the arguments in the proof of Lemma B.2.4, we may prove that

$$\max_{1\leq i\leq d}\max_{1\leq t\leq n}\left\|\widehat{\boldsymbol{\alpha}}_{i\bullet}^{o}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\right\|=O_P\left(\sqrt{s}\zeta_{n,d}\right), \tag{B.2.35}$$

which, together with (B.2.2) in Lemma B.2.1, indicates that

$$\max_{1\le i\le d}\max_{1\le t\le n}|\widehat{e}_{t,i}-e_{t,i}|=O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right). \tag{B.2.36}$$

By (B.2.25), (B.2.36) 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

$$\begin{aligned}
&\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\sum_{t=1}^n\varpi^*_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)e_{t,j}\right|\\
\le\ &\max_{1\le j\le d}\sup_{0\le\tau\le1}\left(\sum_{t=1}^n\left|\varpi^*_{n,t}(\tau)\right|e_{t,j}^2\right)^{1/2}\max_{1\le i\le d}\sup_{0\le\tau\le1}\left(\sum_{t=1}^n\left|\varpi^*_{n,t}(\tau)\right|\left(\widehat{e}_{t,i}-e_{t,i}\right)^2\right)^{1/2}\\
=\ &O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right)=O_P\left(\nu^*_{n,d}\right).
\end{aligned}\tag{B.2.37}$$

Similarly, we can also show that

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\sum_{t=1}^n\varpi^*_{n,t}(\tau)e_{t,i}\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_P\left(\nu^*_{n,d}\right) \tag{B.2.38}$$

and

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\sum_{t=1}^n\varpi^*_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_P\left(\left[\nu^*_{n,d}\right]^2\right)=o_P\left(\nu^*_{n,d}\right). \tag{B.2.39}$$

From (B.2.37)–(B.2.39), we readily have that

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\chi^*_{ij}(\tau)\right|=O_P\left(\nu^*_{n,d}\right),$$

which, together with (B.2.24) and (B.2.25), completes the proof of Lemma B.2.5. ■

B.3 Proofs of Propositions 2.5.1 and 2.5.2

In this appendix, we provide proofs of the convergence properties for the factor-adjusted estimators stated in Propositions 2.5.1 and 2.5.2. 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 (2.5.3) or (2.5.4), 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.2.3 to the factor-adjusted kernel-weighted quantities.

Lemma B.3.1. *Suppose that Assumptions 2.A, 2.B and 2.F(i) are satisfied. Then we have*

$$\max_{1\le i\le d}\max_{1\le t\le n}\left|\widehat{L}_{i,k}(\tau_t)\right|_{\max}=O_P\left(\zeta^{\dagger}_{n,d}\right),\quad k=0,1, \tag{B.3.1}$$

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

Proof of Lemma B.3.1. As in the proof of Lemma B.2.3, we only consider $k=0$. As

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

by Assumption 2.F(i), we may show that

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

Then, by Lemma B.2.3, we complete the proof of (B.3.1) for $k=0$. ■

Write

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

Let $\widehat{\sigma}^{\dagger}_{ij}(\tau)$ be the factor-adjusted local linear estimate $\sigma_{ij}(\tau)$, i.e., replace $\widehat{e}_{t,i}$ by $\widehat{e}^{\dagger}_{t,i}$ in (2.3.8). The following lemma extends Lemma B.2.5 to the factor-adjusted volatility function estimate.

Lemma B.3.2. *Suppose that the assumptions of Proposition 2.5.1(iii) are satisfied. Then we have*

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le1}\left|\widehat{\sigma}^{\dagger}_{ij}(\tau)-\sigma_{ij}(\tau)\right|=O_P\left(\nu^{\diamond}_{n,d}+\nu^{\dagger}_{n,d}\right), \tag{B.3.2}$$

where $\nu^{\diamond}_{n,d}$ *is defined in Assumption 2.E(ii) and* $\nu^{\dagger}_{n,d}$ *is defined in Assumption 2.F(iv).*

Proof of Lemma B.3.2. As in (B.2.24), we have

$$\begin{aligned}\widehat{\sigma}^{\dagger}_{ij}(\tau)-\sigma_{ij}(\tau) &= \left\{\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)e_{t,i}e_{t,j}}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}-\sigma_{ij}(\tau)\right\}+\left\{\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i}-e_{t,i}\right)e_{t,j}}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}+\right.\\ &\quad \left.\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)e_{t,i}\left(\widehat{e}^{\dagger}_{t,j}-e_{t,j}\right)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}+\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i}-e_{t,i}\right)\left(\widehat{e}^{\dagger}_{t,j}-e_{t,j}\right)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}\right\}\\ &=: \chi^{\diamond}_{ij}(\tau)+\chi^{\dagger}_{ij}(\tau).\end{aligned} \tag{B.3.3}$$

By (B.2.25), we only need to show

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\chi^{\dagger}_{ij}(\tau)\right| = O_P\left(\nu^{\dagger}_{n,d}\right). \tag{B.3.4}$$

Following the proof of (B.2.36), we have

$$\max_{1\le i\le d}\max_{1\le t\le n}\left|\widehat{e}^{\dagger}_{t,i} - e_{t,i}\right| = O_P\left(s\zeta^{\dagger}_{n,d}\sqrt{\log(n\vee d)}\right). \tag{B.3.5}$$

By (B.2.25), (B.3.5) and the Cauchy-Schwarz inequality, we can prove that

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i} - e_{t,i}\right)e_{t,j}\right| = O_P\left(s\zeta^{\dagger}_{n,d}\sqrt{\log(n\vee d)}\right) = O_P\left(\nu^{\dagger}_{n,d}\right), \tag{B.3.6}$$

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)e_{t,i}\left(\widehat{e}^{\dagger}_{t,j} - e_{t,j}\right)\right| = O_P\left(\nu^{\dagger}_{n,d}\right), \tag{B.3.7}$$

$$\max_{1\le i,j\le d}\sup_{0\le\tau\le 1}\left|\sum_{t=1}^{n}\varpi^{*}_{n,t}(\tau)\left(\widehat{e}^{\dagger}_{t,i} - e_{t,i}\right)\left(\widehat{e}^{\dagger}_{t,j} - e_{t,j}\right)\right| = o_P\left(\nu^{\dagger}_{n,d}\right). \tag{B.3.8}$$

With (B.3.6)–(B.3.8), we complete the proof of (B.3.4). ■

Define

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

Proof of Proposition 2.5.1. We start with the proof of

$$\min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)} u^{\intercal}\widehat{\mathbf{\Psi}}(\tau_t)u \ge \kappa_0/2, \quad w.p.a.1, \tag{B.3.9}$$

where $\mathscr{B}_i(\tau)$ is defined as in (2.4.2). In fact, combining Assumption 2.F(i) with the arguments in the proofs of Lemmas B.2.3 and B.3.1, we may show that

$$\max_{1\le t\le n}\left\|\widehat{\mathbf{\Psi}}(\tau_t) - \mathbf{\Psi}(\tau_t)\right\|_{\max} = O_P\left([\log(n\vee d)]^{1/2}\delta_X\right). \tag{B.3.10}$$

Then, using (B.3.10) and the arguments in the proof of Lemma B.4.1, we have

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

which, together with Assumptions 2.C(ii) and 2.F(i), completes the proof of (B.3.9).

The proofs of (2.5.6) and (2.5.7) are similar to the proofs of Theorems 2.4.1 and 2.4.2 but with Lemma B.2.3 and (2.4.3) replaced by Lemma B.3.1 and (B.3.9), respectively. The proof of (2.5.8) is similar to the proof of Theorem 2.4.4 but with Lemma B.2.5 replaced by Lemma B.3.2. Details are omitted here to save space. ■

Proof of Proposition 2.5.2. With Proposition 5.1(ii), the proof of (2.5.9) is similar to the proof of Corollary 4.1. With Proposition 5.1(iii), the proof of (2.5.10) is similar to the proof of Corollary 4.2. ■

B.4 Verification of Assumption 2.C(ii)

In this appendix, we verify the uniform restricted eigenvalue condition (2.4.3) for the time-varying VAR under the Gaussian assumption, i.e., $e_t \sim \mathsf{N}(\mathbf{0}_d, \mathbf{\Sigma}_t)$. Recall that

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

We first give some technical lemmas together with their proofs.

Lemma B.4.1. *Conditional on the event that*

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

we have

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

where $\mathscr{B}_i(\tau)$ is defined in Section 2.4.1 and s is defined in Assumption 2.B(ii).

Proof of Lemma B.4.1. The proof is similar to Lemma 6 in Kock and Callot (2015). Write $\mathscr{J}_{i,t} = \mathscr{J}_i(\tau_t)$ and $\mathscr{J}'_{i,t} = \mathscr{J}'_i(\tau_t)$. For $u = \left(u_1^{\intercal}, u_2^{\intercal}\right)^{\intercal} \in \mathscr{B}_i(\tau_t)$ and given $\mathscr{E}_{\mathbf{\Psi}}(\delta)$, we have

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

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

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

Taking $\min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)}$ on both sides of the above inequality, we complete the proof of Lemma B.4.1. ■

Letting

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

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 B.4.2. *Let* $\mathbf{Z}$ *be an* $n \times 1$ *vector with* $\mathbf{Z} \sim \mathsf{N}(\mathbf{0}_n, \mathbf{Q})$. *Then, for any* $\delta, m > 0$,

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

The inequality in Lemma B.4.2 is crucial to derive the probability of the event $\mathscr{E}_{\mathbf{\Psi}}(\delta)$ defined in Lemma B.4.1, as shown in the following lemma.

Lemma B.4.3. *Suppose that Assumptions 2.A and 2.B(i) are satisfied. Then, for any* $\delta, m > 0$, *we have*

$$\mathsf{P}\left(\mathscr{E}_{\mathbf{\Psi}}(\delta)\right) \leq 4nd^2\left[6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right)\right], \tag{B.4.1}$$

where $C_\diamond = \frac{2C_* C_K C_1^2}{(1-\rho)(1-\rho^2)}$, C_* *is defined in Lemma B.2.1,* C_K *is the upper bound of the kernel function* $K(\cdot)$, *and* C_1 *and* ρ *are defined in (2.2.4).*

Proof of Lemma B.4.3. Let the (i,j)-entry of $\mathbf{\Psi}(\tau_t)$ be $\Psi_{i,j}(\tau_t)$. For any $\delta > 0$, we note that

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

Hence, it suffices to show

$$\mathsf{P}\left(|\Psi_{i,j}(\tau_t) - \mathsf{E}[\Psi_{i,j}(\tau_t)]| > \delta\right) \leq 6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right). \tag{B.4.2}$$

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 B.4.2 with $\mathbf{Z} = \widetilde{\mathbf{X}}_u(\tau_t)$ and $\mathbf{Q} = \mathbf{Q}(\tau_t) = \mathsf{Cov}(\widetilde{\mathbf{X}}_u(\tau_t))$. Consider a typical entry in $\mathbf{Q}(\tau_t)$: $\mathsf{Cov}\left(u^{\intercal}\overline{X}_{K,l_1}(\tau_t), u^{\intercal}\overline{X}_{K,l_2}(\tau_t)\right)$ when $|\tau_{l_1} - \tau_t| \leq h$ and $|\tau_{l_2} - \tau_t| \leq h$, 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

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

For $1 \leq l_1, l_2 \leq n$ with $|\tau_{l_1} - \tau_t| \leq h$ and $|\tau_{l_2} - \tau_t| \leq h$, by (2.2.3) and (2.2.4),

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

Hence,

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

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 $\delta, m > 0$,

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

indicating that

$$\mathsf{P}\left(u^{\intercal}\mathbf{\Psi}(\tau_t)u - \mathsf{E}\left[u^{\intercal}\mathbf{\Psi}(\tau_t)u\right] > \delta\right) \leq 2\exp\left(\frac{-\delta^2(nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right). \tag{B.4.3}$$

Choosing u as a vector with the i-th element being one and the others being zeros, by (B.4.3), we have

$$\mathsf{P}\left(|\Psi_{i,i}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t)]| > \delta\right) \leq 2\exp\left(\frac{-\delta^2(nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right) \tag{B.4.4}$$

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

$$\begin{aligned} &\mathsf{P}\left(|\Psi_{i,j}(\tau_t) - \mathsf{E}[\Psi_{i,j}(\tau_t)]| > \delta\right) \\ \leq\ &\mathsf{P}\left(|\Psi_{i,i}(\tau_t) - 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) - 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t)]|/2 > \delta/2\right) + \\ &\mathsf{P}\left(|\Psi_{i,i}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) + \Psi_{j,j}(\tau_t)]|/2 > \delta/2\right) \\ \leq\ &\mathsf{P}\left(|\Psi_{i,i}(\tau_t) + 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t) + 2\Psi_{i,j}(\tau_t) + \Psi_{j,j}(\tau_t)]| > \delta\right) + \\ &\mathsf{P}\left(|\Psi_{i,i}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t)]| > \delta/2\right) + \mathsf{P}\left(|\Psi_{j,j}(\tau_t) - \mathsf{E}[\Psi_{j,j}(\tau_t)]| > \delta/2\right) \\ \leq\ &6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right). \end{aligned} \tag{B.4.5}$$

By virtue of (B.4.4) and (B.4.5), we complete the proof of (B.4.2). ■

The following proposition verifies the uniform restricted eigenvalue condition.

Proposition B.4.1. Suppose that Assumptions 2.A and 2.B(i) are satisfied. If

$$\min_{1\le t\le n}\inf_{u\in\mathscr{B}} u^{\intercal}\mathsf{E}[X_tX_t^{\intercal}]u \ge 2\kappa_0, \tag{B.4.6}$$

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

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

we have (2.4.3) w.p.a.1.

Proof of Proposition B.4.1. Taking $\delta=c_\circ/s$ and $m^2=\left(\frac{c_\circ^2nh}{32C_\diamond^2s^2}\right)^{1/2}$ in Lemma B.4.3 with $c_\circ$ being a proper constant to be determined later, we have

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

which converges to 0 if $s=o\left((nh)^{1/2}/\log(ndh^{1/2})\right)$. By Lemma B.4.1, we then have

$$\min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)} u^{\intercal}\mathbf{\Psi}(\tau_t)u \ge \min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)} u^{\intercal}\mathsf{E}[\mathbf{\Psi}(\tau_t)]u-18c_\circ \quad w.p.a.1. \tag{B.4.7}$$

It remains to prove that the first term on the right side of (B.4.7) has a lower bound and to find a proper value for $c_\circ$. In fact, by (B.4.6), we have

$$\begin{aligned}&\min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)} u^{\intercal}\mathsf{E}[\mathbf{\Psi}(\tau_t)]u\\ =\;&\min_{1\le i\le d}\min_{1\le t\le n}\inf_{u\in\mathscr{B}_i(\tau_t)}\frac{1}{nh}\sum_{l=1}^{n}\mathsf{E}\left[\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)^{\intercal}X_lX_l^{\intercal}\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)\right]K\left(\frac{\tau_l-\tau_t}{h}\right)\\ \ge\;&2\kappa_0\min_{1\le t\le n}\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{\tau_l-\tau_t}{h}\right)=2\kappa_0-\epsilon,\end{aligned}$$

where ϵ is an arbitrary small number. Choosing $c<(\kappa_0-\epsilon)/18$ in (B.4.7), we can complete the proof of (2.4.3). ■

B.5 Tuning parameter selection

The numerical performance of the proposed three-state shrinkage estimation procedure depends on a careful selection of the three tuning parameters: 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 (2.3.3) is essentially the weighted least squares with kernel weights $K_h(\tau_t - \tau)$. The BIC objective function is thus defined as

$$\mathsf{BIC}_i(\lambda_1;\tau) = \log\left[\frac{\mathscr{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau \mid \lambda_1), \widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau \mid \lambda_1)\right)}{\sum_{t=1}^n K_h(\tau_t - \tau)}\right] + \frac{\log(n_e)}{n_e} \cdot \left[|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau \mid \lambda_1)|_0 + \left|\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau \mid \lambda_1)\right|_0\right], \tag{B.5.1}$$

where $\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau \mid \lambda_1)$ and $\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau \mid \lambda_1)$ 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 (B.5.1) 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 *et al.* (2015a). For example, Cheng *et al.* (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

$$\mathsf{GIC}_i(\lambda_2) = \log\left[\frac{1}{n}\sum_{t=1}^n \left\{x_{t,i} - \widehat{\boldsymbol{\alpha}}_{i\bullet}^{\intercal}(\tau_t \mid \lambda_2) X_{t-1}\right\}^2\right] + \frac{\gamma_{n,d}}{n} \cdot \frac{36 s_i(\lambda_2)}{35h}, \tag{B.5.2}$$

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 (B.5.2) 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

$$\mathsf{EBIC}(\lambda_3;\tau) = -\log\left(\det(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3))\right) + \mathsf{Tr}(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3)\widehat{\boldsymbol{\Sigma}}(\tau)) + \frac{\log(n_e)}{n_e} \cdot \sum_{i<j} I(|\widehat{\omega}_{ij}(\tau \mid \lambda_3)| > 0), \tag{B.5.3}$$

where $\widehat{\mathbf{\Omega}}(\tau \mid \lambda_3) = [\widehat{\omega}_{ij}(\tau \mid \lambda_3)]_{d\times d}$ 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 (B.5.3) 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 (2.5.3) or (2.5.4), when the number of factors is set to be q, and define the sum of squared residuals as $\mathsf{V}_n(q) = \sum_{t=1}^n |\widehat{X}_t(q)|_2^2$. When we consider the approximate factor model (2.5.1), 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

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

and obtain $\widehat{q}$ as the estimated number of factors. When we consider the time-varying factor model (2.5.2), we adopt Su and Wang (2017)'s information criterion, i.e., maximise the following objective function with respect to q

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

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 *et al.* (2023) 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 2.7 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 $R(k)$, $1 \le k \le k_{\max}$. 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

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

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

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

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

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

and MCC is defined as

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

Appendix C

Appendix to Chapter 3

C.1 Proofs of the asymptotic theorems

Proof of Proposition 3.2.1. The uniform consistency results in (3.2.10) and (3.2.11) follow from Theorem 4 in Fan *et al.* (2013), Theorem 3(i) in Chen *et al.* (2018) and Lemma D.1 in Li *et al.* (2023). It remains to prove (3.2.12). Note that

$$\begin{aligned}\widehat{\varepsilon}_{t,i} - \varepsilon_{t,i} &= -\left(\widehat{\boldsymbol{\lambda}}_i^{\intercal}\widehat{\mathbf{F}}_t - \left[\left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right]^{\intercal}\left[\mathbf{R}\mathbf{F}_t\right]\right) \\ &= -\left(\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right)^{\intercal}\left(\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right) - \left[\left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right]^{\intercal}\left(\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right) \\ &\quad -\left(\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right)^{\intercal}\mathbf{R}\mathbf{F}_t. \end{aligned} \tag{C.1.1}$$

By (3.2.10), (3.2.11), Assumption 3.A and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\max_{1\le i\le N}\max_{1\le t\le T}\left|\left(\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right)^{\intercal}\left(\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right)\right| &\le \max_{1\le i\le N}\left\|\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right\|\max_{1\le t\le T}\left\|\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right\| \\ &= o_P\left(\left(\frac{\log N}{T}\right)^{1/2} + \frac{T^{2/\delta}}{N^{1/2}}\right)\end{aligned} \tag{C.1.2}$$

and

$$\begin{aligned}\max_{1\le i\le N}\max_{1\le t\le T}\left|\left[\left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right]^{\intercal}\left(\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right)\right| &\le \max_{1\le i\le N}\left\|\left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right\|\max_{1\le t\le T}\left\|\widehat{\mathbf{F}}_t - \mathbf{R}\mathbf{F}_t\right\| \\ &= O_P\left(\frac{1}{T^{1/2}} + \frac{T^{2/\delta}}{N^{1/2}}\right).\end{aligned} \tag{C.1.3}$$

By the sub-Gaussian moment condition on $\mathbf{F}_t$ in Assumption 3.A(iii), the Bonferroni and Markov inequalities, we have $\max_{1\le t\le T}\|\mathbf{F}_t\| = O_P\left((\log T)^{1/2}\right)$, which together with (3.2.11), leads to

$$\max_{1\le i\le N}\max_{1\le t\le T}\left|\left(\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right)^{\intercal}\mathbf{R}\mathbf{F}_t\right| \le \max_{1\le i\le N}\left\|\widehat{\boldsymbol{\lambda}}_i - \left(\mathbf{R}^{-1}\right)^{\intercal}\boldsymbol{\lambda}_i\right\|\max_{1\le t\le T}\|\mathbf{R}\mathbf{F}_t\|$$

$$= O_P\left((\log T)^{1/2}\left[\left(\frac{\log N}{T}\right)^{1/2}+\frac{T^{2/\delta}}{N^{1/2}}\right]\right). \quad \text{(C.1.4)}$$

With (C.1.1)–(C.1.4), we complete the proof of (3.2.12). ■

Proof of Proposition 3.3.1. Let

$$w_{t,k}(u_k)=K\left(\frac{U_{t,k}-u_k}{h}\right)\Big/\left\{\sum_{t=1}^{T-1}K\left(\frac{U_{t,k}-u_k}{h}\right)\right\},$$

and note that

$$\begin{aligned}\widehat{\sigma}_{\varepsilon,k,ij}(u_k)-\sigma_{\varepsilon,k,ij}(u_k) &= \sum_{t=1}^{T-1}w_{t,k}(u_k)\left[\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j}-\sigma_{\varepsilon,k,ij}(u_k)\right]\\ &= \sum_{t=1}^{T-1}w_{t,k}(u_k)\left(\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j}-\varepsilon_{t+1,i}\varepsilon_{t+1,j}\right)\\ &\quad+\sum_{t=1}^{T-1}w_{t,k}(u_k)\left[\varepsilon_{t+1,i}\varepsilon_{t+1,j}-\sigma_{\varepsilon,k,ij}(u_k)\right].\end{aligned}$$

By Lemma 1 in Chen *et al.* (2019), we have

$$\max_{1\le i,j\le N}\max_{1\le k\le d}\sup_{u_k\in\mathscr{U}_{k,h}}\left|\sum_{t=1}^{T-1}w_{t,k}(u_k)\left[\varepsilon_{t+1,i}\varepsilon_{t+1,j}-\sigma_{\varepsilon,k,ij}(u_k)\right]\right|=O_P\left(\zeta_{NT,2}\right). \quad \text{(C.1.5)}$$

By (3.2.12) in Proposition 3.2.1, we can show that

$$\max_{1\le i,j\le N}\max_{1\le k\le d}\sup_{u_k\in\mathscr{U}_{k,h}}\left|\sum_{t=1}^{T-1}w_{t,k}(u_k)\left(\widehat{\varepsilon}_{t+1,i}\widehat{\varepsilon}_{t+1,j}-\varepsilon_{t+1,i}\varepsilon_{t+1,j}\right)\right|=O_P\left(\zeta_{NT,1}\right). \quad \text{(C.1.6)}$$

Combining (C.1.5) and (C.1.6), we have

$$\max_{1\le i,j\le N}\max_{1\le k\le d}\sup_{u_k\in\mathscr{U}_{k,h}}\left|\widehat{\sigma}_{\varepsilon,k,ij}(u_k)-\sigma_{\varepsilon,k,ij}(u_k)\right|=O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right). \quad \text{(C.1.7)}$$

Then following the proof of Lemma 2 in Chen *et al.* (2019), we can show that

$$\max_{1\le i,j\le N}\max_{0\le k\le d}\left|\widehat{a}_{k,ij}-a^o_{k,ij}\right|=O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right). \quad \text{(C.1.8)}$$

Finally, by (C.1.7) and (C.1.8), we have

$$\begin{aligned}\widehat{\sigma}_{\varepsilon,ij}(\mathbf{u})-\sigma^o_{\varepsilon,ij}(\mathbf{u}) &= \left[\widehat{a}_{0,ij}+\sum_{k=1}^{d}\widehat{a}_{k,ij}\widehat{\sigma}_{\varepsilon,k,ij}(u_k)\right]-\left[a^o_{0,ij}+\sum_{k=1}^{d}a^o_{k,ij}\sigma_{\varepsilon,k,ij}(u_k)\right]\\ &= \left(\widehat{a}_{0,ij}-a^o_{0,ij}\right)+\sum_{k=1}^{d}\left(\widehat{a}_{k,ij}-a^o_{k,ij}\right)\sigma_{\varepsilon,k,ij}(u_k)+\sum_{k=1}^{d}a^o_{k,ij}\left[\widehat{\sigma}_{\varepsilon,k,ij}(u_k)-\sigma_{\varepsilon,k,ij}(u_k)\right]\\ &\quad+\sum_{k=1}^{d}\left(\widehat{a}_{k,ij}-a^o_{k,ij}\right)\left[\widehat{\sigma}_{\varepsilon,k,ij}(u_k)-\sigma_{\varepsilon,k,ij}(u_k)\right]\end{aligned}$$

$$= \quad O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right)$$

uniformly for $1 \leq i,j \leq N$ and $\mathbf{u} \in \mathscr{U}_h$. This completes the proof of Proposition 3.3.1. ■

Proof of Proposition 3.3.2. The proof of this proposition is analogous to that of Proposition 3.3.1, but with (3.2.10) replacing the role of (3.2.12). Details are omitted here to save space. ■

Proof of Theorem 3.3.1. By (3.3.3) in Proposition 3.3.1, we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})-\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\right\|_{\max} = O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right). \tag{C.1.9}$$

By (C.1.9), the uniform sparsity assumption (3.2.8), and the inequality $\|\mathbf{A}_1\mathbf{A}_2\|_{\max} \leq \|\mathbf{A}_1\|_{\max}\|\mathbf{A}_2\|_1$ for any two conformable matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\mathbf{I}_N-\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max} \leq \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})-\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\right\|_{\max}\left\|\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1 = O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right).$$

This together with Assumption 3.D(iii) and the constraint the CLIME estimator satisfies (see (3.2.19)), implies

$$\begin{aligned}
&\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}\\
\leq\ &\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{I}_N\right\|_{\max}+\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\mathbf{I}_N-\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max}\\
\leq\ &\rho+O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right)=O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right).
\end{aligned} \tag{C.1.10}$$

It follows from the definition of the CLIME estimator and Lemma 1 in Cai *et al.* (2011) that $\left\|\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\right\|_1 \leq \left\|\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1 \leq M$ uniformly over $\mathbf{u}\in\mathscr{U}_h$. Then, by (C.1.9) and (C.1.10), we have

$$\begin{aligned}
&\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}\\
\leq\ &\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}+\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\left[\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})-\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\right]\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}\\
\leq\ &\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}+\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Sigma}}_\varepsilon(\mathbf{u})-\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\right\|_{\max}\left\|\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1\\
=\ &O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right).
\end{aligned} \tag{C.1.11}$$

By (C.1.11) and the relation $\|\mathbf{A}_1\mathbf{A}_2\|_{\max} \leq \|\mathbf{A}_1\|_\infty\|\mathbf{A}_2\|_{\max} = \|\mathbf{A}_1\|_1\|\mathbf{A}_2\|_{\max}$ for any symmetric matrix $\mathbf{A}_1$ and a conformable matrix $\mathbf{A}_2$, we have

$$\begin{aligned}
\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max} &\leq \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1\left\|\mathbf{\Sigma}^o_\varepsilon(\mathbf{u})\left[\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})-\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right]\right\|_{\max}\\
&= O_P\left(\zeta_{NT,1}+\zeta_{NT,2}\right).
\end{aligned} \tag{C.1.12}$$

By (C.1.12) and noting that $\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})$ is the symmetrisation of $\widetilde{\mathbf{\Omega}}_\varepsilon(\mathbf{u})$ via (3.2.20), we prove (3.3.5).

We next turn to the proof of (3.3.6). Let $\omega^o_{\varepsilon,ij}(\mathbf{u})$ be the (i,j)-entry of $\mathbf{\Omega}^o_\varepsilon(\mathbf{u})$. Then by the definition of $\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})$ in (3.2.20) and Lemma 1 in Cai *et al.* (2011), we have, for any $\mathbf{u} \in \mathscr{U}_h$ and $1 \le j \le N$,

$$\sum_{i=1}^N |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \le \sum_{i=1}^N |\widetilde{\omega}_{\varepsilon,ij}(\mathbf{u})| \le \sum_{i=1}^N \left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right|.$$

Then, as $\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) = \widehat{\omega}_{\varepsilon,ji}(\mathbf{u})$ and $\omega^o_{\varepsilon,ij}(\mathbf{u}) = \omega^o_{\varepsilon,ji}(\mathbf{u})$, we have

$$\begin{aligned}
&\sum_{j=1}^N |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})|\, I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \le \rho\right) \\
= &\sum_{j=1}^N |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| - \sum_{j=1}^N |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})|\, I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) \\
\le &\sum_{j=1}^N |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| - \sum_{j=1}^N \left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right| + \sum_{j=1}^N \left|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) - \omega^o_{\varepsilon,ij}(\mathbf{u})\right| \\
\le &\sum_{j=1}^N \left|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) - \omega^o_{\varepsilon,ij}(\mathbf{u})\right|.
\end{aligned}$$

Further noticing that $\|\mathbf{A}\|_o \le \sqrt{\|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty} = \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$ for any symmetric matrix $\mathbf{A}$, we have

$$\begin{aligned}
\sup_{\mathbf{u}\in\mathscr{U}_h} \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_O &\le \sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\le i\le N} \sum_{j=1}^N \left|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) - \omega^o_{\varepsilon,ij}(\mathbf{u})\right| \\
&\le 2 \sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\le i\le N} \sum_{j=1}^N \left|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) - \omega^o_{\varepsilon,ij}(\mathbf{u})\right| \\
&\le 2 \sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\le i\le N} \sum_{j=1}^N \left|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) - \omega^o_{\varepsilon,ij}(\mathbf{u})\right| I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) \\
&\quad + 2 \sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\le i\le N} \sum_{j=1}^N \left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right| I\left(|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \le \rho\right) \\
&=: \Pi_1 + \Pi_2. \qquad \text{(C.1.13)}
\end{aligned}$$

Define the event

$$\mathscr{E} = \left\{ \sup_{\mathbf{u}\in\mathscr{U}_h} \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max} \le M_\eta(\zeta_{NT,1} + \zeta_{NT,2}) \right\},$$

where M_η is a positive constant such that $\mathsf{P}(\mathscr{E}) \ge 1 - \eta$ for any $\eta > 0$. Then, conditional on $\mathscr{E}$,

$$\Pi_1 \le 2M_\eta(\zeta_{NT,1} + \zeta_{NT,2}) \sup_{\mathbf{u}\in\mathscr{U}_h} \left[\max_{1\le i\le N} \sum_{j=1}^N I\left(|\widehat{\boldsymbol{\omega}}_{\varepsilon,ij}(\mathbf{u})| > \rho\right) \right] \qquad \text{(C.1.14)}$$

and

$$|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \le |\omega^o_{\varepsilon,ij}(\mathbf{u})| + |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) - \omega^o_{\varepsilon,ij}(\mathbf{u})| \le |\omega^o_{\varepsilon,ij}(\mathbf{u})| + M_\eta(\zeta_{NT,1} + \zeta_{NT,2}).$$

Recall that $\rho = c_5(\zeta_{NT,1} + \zeta_{NT,2})$ in Assumption 3.D(iii). By choosing $c_5 = 2M_\eta$, we can verify that, on $\mathscr{E}$, the event $\{|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| > \rho\}$ implies $\{|\omega^o_{\varepsilon,ij}(\mathbf{u})| > M_\eta(\zeta_{NT,1} + \zeta_{NT,2})\}$. Consequently, by (C.1.14), we have, on $\mathscr{E}$,

$$\begin{aligned}
\Pi_1 &\leq M_\eta(\zeta_{NT,1} + \zeta_{NT,2}) \left[\sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\leq i\leq N} \sum_{j=1}^{N} I\left(|\omega^o_{\varepsilon,ij}(\mathbf{u})| > M_\eta(\zeta_{NT,1} + \zeta_{NT,2})\right) \right] \\
&\leq M_\eta(\zeta_{NT,1} + \zeta_{NT,2}) \left[\sup_{\mathbf{u}\in\mathscr{U}} \max_{1\leq i\leq N} \sum_{j=1}^{N} \frac{\left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right|^q}{M^q_\eta(\zeta_{NT,1} + \zeta_{NT,2})^q} \right] \\
&= O\left(\varpi_N \cdot (\zeta_{NT,1} + \zeta_{NT,2})^{1-q}\right). \qquad \text{(C.1.15)}
\end{aligned}$$

On the other hand, by the triangle inequality, for any $\mathbf{u} \in \mathscr{U}_h$, we have

$$|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \geq |\omega^o_{\varepsilon,ij}(\mathbf{u})| - |\widehat{\omega}_{\varepsilon,ij}(\mathbf{u}) - \omega^o_{\varepsilon,ij}(\mathbf{u})| \geq |\omega^o_{\varepsilon,ij}(\mathbf{u})| - M_\eta(\zeta_{NT,1} + \zeta_{NT,2})$$

on $\mathscr{E}$. Then it is easy to see that $\{|\widehat{\omega}_{\varepsilon,ij}(\mathbf{u})| \leq \rho\}$ implies

$$\left\{|\omega^o_{\varepsilon,ij}(\mathbf{u})| \leq (c_5 + M_\eta)\,(\zeta_{NT,1} + \zeta_{NT,2})\right\}.$$

Hence, for Π_2, by (3.2.8) and Assumption 3.D(iii), we have

$$\begin{aligned}
\Pi_2 &\leq 2 \sup_{\mathbf{u}\in\mathscr{U}_h} \max_{1\leq i\leq N} \sum_{j=1}^{N} \left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right| I\left(|\omega^o_{\varepsilon,ij}(\mathbf{u})| \leq (c_5 + M_\eta)\,(\zeta_{NT,1} + \zeta_{NT,2})\right) \\
&\leq 2(c_5 + M_\eta)^{1-q}(\zeta_{NT,1} + \zeta_{NT,2})^{1-q} \sup_{\mathbf{u}\in\mathscr{U}} \max_{1\leq i\leq N} \sum_{j=1}^{N} \left|\omega^o_{\varepsilon,ij}(\mathbf{u})\right|^q \\
&= O_P\left(\varpi_N(\zeta_{NT,1} + \zeta_{NT,2})^{1-q}\right). \qquad \text{(C.1.16)}
\end{aligned}$$

The proof of (3.3.6) can be completed by (C.1.13), (C.1.15) and (C.1.16).

Finally, (3.3.7) is proved by (3.3.5), the proof of (3.3.6)[1] and noting that

$$\frac{1}{N}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_F^2 \leq \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_{\max} \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1.$$

■

Proof of Theorem 3.3.2. We first re-write

$$\widehat{\mathbf{\Omega}}_X(\mathbf{u}) = \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1}\widehat{\mathbf{\Lambda}}^\intercal\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) \qquad \text{(C.1.17)}$$

[1]Note that we proved (3.3.6) by showing that $\sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_O \leq \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_1 = \sup_{\mathbf{u}\in\mathscr{U}_h}\left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}^o_\varepsilon(\mathbf{u})\right\|_\infty = O_P\left(\varpi_N(\zeta_{NT,1} + \zeta_{NT,2})^{1-q}\right)$ (see (C.1.13), (C.1.15), and (C.1.16)).

and

$$\mathbf{\Omega}_X^o(\mathbf{u}) = \mathbf{\Omega}_\varepsilon^o(\mathbf{u}) - \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\mathbf{\Lambda}_R \left[\mathbf{\Omega}_{F,R}^o(\mathbf{u}) + \mathbf{\Lambda}_R^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\mathbf{\Lambda}_R\right]^{-1} \mathbf{\Lambda}_R^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u}), \tag{C.1.18}$$

where $\widehat{\mathbf{\Omega}}_F(\mathbf{u}) = \left[\widehat{\mathbf{\Sigma}}_F(\mathbf{u})\right]^{-1}$, $\mathbf{\Omega}_{F,R}^o(\mathbf{u}) = \left[\mathbf{R}\mathbf{\Sigma}_F^o(\mathbf{u})\mathbf{R}^\intercal\right]^{-1}$ and $\mathbf{\Lambda}_R = \mathbf{\Lambda}\mathbf{R}^{-1}$. By (C.1.17), (C.1.18) as well as some standard arguments, we may show that

$$\widehat{\mathbf{\Omega}}_X(\mathbf{u}) - \mathbf{\Omega}_X^o(\mathbf{u}) = -\sum_{k=1}^{6} \mathbf{\Xi}_k(\mathbf{u}),$$

where

$$\begin{aligned}
\mathbf{\Xi}_1(\mathbf{u}) &= \mathbf{\Omega}_\varepsilon^o(\mathbf{u}) - \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}), \\
\mathbf{\Xi}_2(\mathbf{u}) &= \left[\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\right] \widehat{\mathbf{\Lambda}} \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}), \\
\mathbf{\Xi}_3(\mathbf{u}) &= \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\widehat{\mathbf{\Lambda}} \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \widehat{\mathbf{\Lambda}}^\intercal \left[\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u}) - \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\right], \\
\mathbf{\Xi}_4(\mathbf{u}) &= \mathbf{\Omega}_\varepsilon^o(\mathbf{u}) \left(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}_R\right) \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \widehat{\mathbf{\Lambda}}^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u}), \\
\mathbf{\Xi}_5(\mathbf{u}) &= \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\mathbf{\Lambda}_R \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \left(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}_R\right)^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u}), \\
\mathbf{\Xi}_6(\mathbf{u}) &= \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\mathbf{\Lambda}_R\mathbf{D}(\mathbf{u})\mathbf{\Lambda}_R^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u})
\end{aligned}$$

with

$$\mathbf{D}(\mathbf{u}) = \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} - \left[\mathbf{\Omega}_{F,R}^o(\mathbf{u}) + \mathbf{\Lambda}_R^\intercal \mathbf{\Omega}_\varepsilon^o(\mathbf{u})\mathbf{\Lambda}_R\right]^{-1} =: \left[\widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1} - \left[\mathbf{\Omega}_{F,\varepsilon}^o(\mathbf{u})\right]^{-1}.$$

By Theorem 3.3.1(ii), we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \|\mathbf{\Xi}_1(\mathbf{u})\|_O = O_P\left(\varpi_N \left(\zeta_{NT,1} + \zeta_{NT,2}\right)^{1-q}\right). \tag{C.1.19}$$

Note that, with probability approaching one, $\lambda_{\min}\left(\widehat{\mathbf{\Omega}}_F(\mathbf{u})\right) = \left[\lambda_{\max}\left(\widehat{\mathbf{\Sigma}}_F(\mathbf{u})\right)\right]^{-1}$ is uniformly bounded away from zero on $\mathscr{U}_h$. Furthermore, following the proof of Lemma 15 in Fan *et al.* (2013) and by Assumption 3.A(ii), we can show that, with probability approaching one, $\lambda_{\min}\left(\widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right) \geq c_0 N$ uniformly over $\mathbf{u} \in \mathscr{U}_h$, where c_0 is a positive constant. Combining these facts, we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \left\| \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \right\|_O = O_P(1/N). \tag{C.1.20}$$

Then by Theorem 3.3.1(ii), Assumption 3.A and the uniform sparsity assumption (3.2.8), we readily have

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \|\mathbf{\Xi}_2(\mathbf{u})\|_O \leq \sup_{\mathbf{u}\in\mathscr{U}_h} \left\{ \left\|\mathbf{\Omega}_\varepsilon^o(\mathbf{u}) - \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\right\|_O \left\|\widehat{\mathbf{\Lambda}} \left[\widehat{\mathbf{\Omega}}_F(\mathbf{u}) + \widehat{\mathbf{\Lambda}}^\intercal \widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\widehat{\mathbf{\Lambda}}\right]^{-1} \widehat{\mathbf{\Lambda}}^\intercal \right\|_O \left\|\widehat{\mathbf{\Omega}}_\varepsilon(\mathbf{u})\right\|_O \right\}$$

$$= O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{1-q}\right), \tag{C.1.21}$$

and similarly,

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\|\boldsymbol{\Xi}_3(\mathbf{u})\|_O = O_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{1-q}\right). \tag{C.1.22}$$

On the other hand, by (3.2.11) in Proposition 3.2.1(ii), we have

$$\begin{aligned}\left\|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\left(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}_R\right)\right\|_O \leq& \|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\|_O\left\|\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}_R\right\|_O \leq \|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\|_1 N^{1/2}\left\|\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}_R\right\|_\infty\\ =&O_P\left(N^{1/2}\left[\left(\frac{\log N}{T}\right)^{1/2}+\frac{T^{2/\delta}}{N^{1/2}}\right]\right),\end{aligned}$$

and further by Assumption 3.A,

$$\left\|\widehat{\boldsymbol{\Lambda}}^\intercal\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\right\|_O \leq \left\|\widehat{\boldsymbol{\Lambda}}^\intercal\right\|_O\|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\|_O \leq N^{1/2}\left\|\widehat{\boldsymbol{\Lambda}}\right\|_\infty\|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\|_1 = O_P(N^{1/2}).$$

Combining the above results with (C.1.20), we have

$$\begin{aligned}\sup_{\mathbf{u}\in\mathscr{U}_h}\|\boldsymbol{\Xi}_4(\mathbf{u})\|_O &\leq \sup_{\mathbf{u}\in\mathscr{U}_h}\left\{\left\|\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\left(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}_R\right)\right\|_O\left\|\left[\widehat{\boldsymbol{\Omega}}_F(\mathbf{u})+\widehat{\boldsymbol{\Lambda}}^\intercal\widehat{\boldsymbol{\Omega}}_\varepsilon(\mathbf{u})\widehat{\boldsymbol{\Lambda}}\right]^{-1}\right\|_O\left\|\widehat{\boldsymbol{\Lambda}}^\intercal\boldsymbol{\Omega}^o_\varepsilon(\mathbf{u})\right\|_O\right\}\\ &= O_P\left(\left(\frac{\log N}{T}\right)^{1/2}+\frac{T^{2/\delta}}{N^{1/2}}\right) = o_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{1-q}\right),\end{aligned} \tag{C.1.23}$$

and similarly,

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\|\boldsymbol{\Xi}_5(\mathbf{u})\|_O = o_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{1-q}\right). \tag{C.1.24}$$

We next consider $\boldsymbol{\Xi}_6(\mathbf{u})$. Note that

$$\begin{aligned}\|\mathbf{D}(\mathbf{u})\|_O &= \left\|\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})-\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\\ &\leq \left(\left\|\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}-\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O+\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\right)\\ &\quad\times\left\|\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})-\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right\|_O\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O.\end{aligned}$$

Further, notice that

$$\begin{aligned}&-\left(\left\|\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}-\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O+\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\right)\left\|\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})-\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right\|_O\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\\ \leq\ &-\left(\left\|2\mathbf{I}-\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})-\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right\|_O+\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})-\mathbf{I}\right\|_O\right)\\ &\times\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\\ \leq\ &-\left(\left\|\mathbf{I}-\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right\|_O\right)\left\|\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\\ \leq\ &-\left\|\left[\widehat{\boldsymbol{\Omega}}_{F,\varepsilon}(\mathbf{u})\right]^{-1}-\left[\boldsymbol{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O.\end{aligned}$$

Hence,

$$\begin{aligned} & \|\mathbf{D}(\mathbf{u})\|_O \left[1 - \left\|\widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u}) - \mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right\|_O \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O\right] \\ \leq & \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O \left\|\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u}) - \widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u})\right\|_O \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O. \end{aligned}$$

This implies

$$\|\mathbf{D}(\mathbf{u})\|_O \leq \frac{\left\|\widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u}) - \mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right\|_O \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O^2}{1 - \left\|\widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u}) - \mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right\|_O \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O}. \tag{C.1.25}$$

By Proposition 3.2.1(ii), Proposition 3.3.2 and Theorem 3.3.1(ii), we can show that

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \left\|\widehat{\mathbf{\Omega}}_{F,\varepsilon}(\mathbf{u}) - \mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right\|_O = O_P\left(N\varpi_N(\zeta_{NT,1} + \zeta_{NT,2})^{1-q}\right). \tag{C.1.26}$$

Similar to (C.1.20), we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \left\|\left[\mathbf{\Omega}^o_{F,\varepsilon}(\mathbf{u})\right]^{-1}\right\|_O = O_P(1/N). \tag{C.1.27}$$

By virtue of (C.1.25)–(C.1.27), we can prove that

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \|\mathbf{D}(\mathbf{u})\|_O = O_P\left(N^{-1}\varpi_N(\zeta_{NT,1} + \zeta_{NT,2})^{1-q}\right). \tag{C.1.28}$$

By (3.2.8) and Assumption 3.A, $\|\mathbf{\Omega}^o_\varepsilon(\mathbf{u})\mathbf{\Lambda}_R\|_O = O_P(N^{1/2})$ uniformly over $\mathbf{u} \in \mathscr{U}_h$, which together with (C.1.28), implies

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \|\mathbf{\Xi}_6(\mathbf{u})\|_O = O_P\left(\varpi_N\left(\zeta_{NT,1} + \zeta_{NT,2}\right)^{1-q}\right). \tag{C.1.29}$$

With (C.1.19), (C.1.21)–(C.1.24) and (C.1.29), we complete the proof of (3.3.11).

We next prove (3.3.12). By Theorem 3.3.1(iii), we have

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \frac{1}{N}\|\mathbf{\Xi}_1(\mathbf{u})\|_F^2 = O_P\left(\varpi_N\left(\zeta_{NT,1} + \zeta_{NT,2}\right)^{2-q}\right). \tag{C.1.30}$$

Notice that for any two compatible matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, it holds that

$$\|\mathbf{A}_1\mathbf{A}_2\|_F \leq \|\mathbf{A}_1\|_F\|\mathbf{A}_2\|_O, \qquad \|\mathbf{A}_1\mathbf{A}_2\|_F \leq \|\mathbf{A}_1\|_O\|\mathbf{A}_2\|_F. \tag{C.1.31}$$

Then by (C.1.20) and (C.1.30), we can show that

$$\sup_{\mathbf{u}\in\mathscr{U}_h} \frac{1}{N}\left(\|\mathbf{\Xi}_2(\mathbf{u})\|_F^2 + \|\mathbf{\Xi}_3(\mathbf{u})\|_F^2\right) = O_P\left(\varpi_N\left(\zeta_{NT,1} + \zeta_{NT,2}\right)^{2-q}\right). \tag{C.1.32}$$

By (C.1.31), (3.2.8), (C.1.20), Proposition 3.2.1(ii) and Assumption 3.A, we have

$$\begin{aligned}\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\|\boldsymbol{\Xi}_4(\mathbf{u})\|_F^2 &\leq \frac{1}{N}\sup_{\mathbf{u}\in\mathscr{U}_h}\left\{\|\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\|_O^2\left\|\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}_R\right\|_F^2\left\|\left[\widehat{\boldsymbol{\Omega}}_F(\mathbf{u})+\widehat{\boldsymbol{\Lambda}}^\intercal\widehat{\boldsymbol{\Omega}}_\varepsilon(\mathbf{u})\widehat{\boldsymbol{\Lambda}}\right]^{-1}\right\|_O^2\left\|\widehat{\boldsymbol{\Lambda}}^\intercal\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\right\|_O^2\right\}\\ &= O_P\left(\frac{1}{N}\left\{N\left[\left(\frac{\log N}{T}\right)^{1/2}+\frac{T^{2/\delta}}{N^{1/2}}\right]^2\frac{1}{N^2}\cdot N\right\}\right)\\ &= o_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{2-q}\right),\end{aligned}\tag{C.1.33}$$

and similarly,

$$\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\|\boldsymbol{\Xi}_5(\mathbf{u})\|_F^2 = o_P\left(\varpi_N\left(\zeta_{NT,1}+\zeta_{NT,2}\right)^{2-q}\right).\tag{C.1.34}$$

For $\boldsymbol{\Xi}_6(\mathbf{u})$, by (3.2.8), (C.1.28), and Assumption 3.A, we have

$$\begin{aligned}\sup_{\mathbf{u}\in\mathscr{U}_h}\frac{1}{N}\|\boldsymbol{\Xi}_6(\mathbf{u})\|_F^2 &\leq \frac{1}{N}\sup_{\mathbf{u}\in\mathscr{U}_h}\|\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\boldsymbol{\Lambda}_R\mathbf{D}(\mathbf{u})\|_O^2\left\|\boldsymbol{\Lambda}_R^\intercal\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\right\|_F^2\\ &\leq \frac{1}{N}\sup_{\mathbf{u}\in\mathscr{U}_h}\|\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\|_O^2\|\boldsymbol{\Lambda}_R\|_O^2\|\mathbf{D}(\mathbf{u})\|_O^2\left\|\boldsymbol{\Lambda}_R^\intercal\right\|_F^2\|\boldsymbol{\Omega}_\varepsilon^o(\mathbf{u})\|_O^2\\ &= O_P\left(\frac{\varpi_N^2(\zeta_{NT,1}+\zeta_{NT,2})^{2-2q}}{N}\right).\end{aligned}\tag{C.1.35}$$

Combining (C.1.30) and (C.1.32)–(C.1.35), we prove (3.3.12). ■

Bibliography

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics* **216**, 86–105.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* **201**, 384–399.

Aït-Sahalia, Y. and Xiu, D. (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association* **114**, 287–303.

Anatolyev, S. and Mikusheva, A. (2022). Factor models with many assets: Strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics* **229**, 103–126.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**, 885–905.

Ando, T. and Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association* **115**, 266–279.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**, 135–171.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics* **40**, 436–465.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica* **72**, 1127–1177.

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. *arXiv preprint arXiv:2201.06110* .

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: Recovering the market volatility shocks. *The Econometrics Journal* **19**, 33–60.

Barigozzi, M. and Hallin, M. (2017). Generalized dynamic factor models and volatilities: Estimation and forecasting. *Journal of Econometrics* **201**, 307–321.

Barigozzi, M., Lippi, M., and Luciani, M. (2020). Cointegration and error correction mechanisms for singular stochastic vectors. *Econometrics* **8**, 3.

Barigozzi, M., Lippi, M., and Luciani, M. (2021). Large-dimensional dynamic factor models: Estimation of impulse–response functions with $I(1)$ cointegrated factors. *Journal of Econometrics* **221**, 455–482.

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. *The Journal of Machine Learning Research* **16**, 417–453.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics* **37**, 1705–1732.

Bollerslev, T., Meddahi, N., and Nyawa, S. (2019). High-dimensional multivariate realized volatility estimation. *Journal of Econometrics* **212**, 116–136.

Briggs, N. E. and MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research* **38**, 25–56.

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., 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, T. T., Hu, J., Li, Y., and Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics* **214**, 482–494.

Cai, T. T., Ren, Z., and Zhou, H. H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10**, 1–59.

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

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance* **52**, 57–82.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–1304.

Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. (2012). Latent variable graphical model selection via convex optimization. *The Annals of Statistics* **40**, 1935–1967.

Chen, E. Y., Fan, J., and Zhu, X. (2023). Community network auto-regression for high-dimensional time series. *Journal of Econometrics* **235**, 1239–1256.

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

Chen, J., Li, D., and Linton, O. (2019). A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics* **212**, 155–176.

Chen, J., Li, D., Linton, O., and Lu, Z. (2016). Semiparametric dynamic portfolio choice with multiple conditioning variables. *Journal of Econometrics* **194**, 309–318.

Chen, J., Li, D., Linton, O., and Lu, Z. (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association* **113**, 919–932.

Chen, J., Li, D., Wei, L., and Zhang, W. (2021a). Nonparametric homogeneity pursuit in functional-coefficient models. *Journal of Nonparametric Statistics* **33**, 387–416.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021b). Quantile factor models. *Econometrica* **89**, 875–910.

Cheng, M.-Y., 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.

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.

Choi, J. Y., Salandro, D., and Shastri, K. (1988). On the estimation of bid-ask spreads: Theory and evidence. *Journal of Financial and Quantitative Analysis* **23**, 219–230.

Chordia, T., Roll, R., and Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics* **56**, 3–28.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* **159**, 116–133.

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.

Dai, C., Lu, K., and Xiu, D. (2019). Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics* **208**, 43–79.

Davis, R. A., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* **25**, 1077–1096.

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

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

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* **182**, 119–134.

Diebold, F. X. and Yılmaz, K. (2015). *Financial and macroeconomic connectedness: A network approach to measurement and monitoring*. Oxford University Press, USA.

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.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance* **47**, 427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**, 3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**, 1–22.

Fan, J., Feng, Y., Jiang, J., and Tong, X. (2016a). Feature augmentation via nonparametrics and selection (FANS) in high-dimensional classification. *Journal of the American Statistical Association* **111**, 275–287.

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., Furger, A., and Xiu, D. (2016b). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics* **34**, 489–503.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Springer.

Fan, J. and Kim, D. (2018). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association* **113**, 1268–1283.

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., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* **107**, 412–428.

Fan, J. and Liao, Y. (2022). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *Journal of the American Statistical Association* **117**, 909–924.

Fan, J., Liao, Y., and Liu, H. (2016c). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal* **19**, C1–C32.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of statistics* **39**, 3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 603–680.

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory* **57**, 5467–5484.

Fan, J., Ma, Y., and Dai, W. (2014a). Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *Journal of the American Statistical Association* **109**, 1270–1284.

Fan, J., Masini, R., and Medeiros, M. C. (2021). Bridging factor and sparse models. *arXiv preprint arXiv:2102.11341* .

Fan, J., Xue, L., and Zou, H. (2014b). Strong oracle optimality of folded concave penalized estimation. *The Annals of statistics* **42**, 819.

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**, 1–9.

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

Freyaldenhoven, S. (2022). Factor models with local factors—determining the number of relevant factors. *Journal of Econometrics* **229**, 80–102.

Giglio, S., Kelly, B., and Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics* **14**, 337–368.

Giglio, S., Xiu, D., and Zhang, D. (2021). Test assets and weak factors. Chicago Booth research paper available at SSRN http://dx.doi.org/10.2139/ssrn.3768081 .

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

Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* **159**, 55–73.

Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics* **163**, 29–41.

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.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory* **13**, 529–557.

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *The Review of Financial Studies* **6**, 191–212.

Hasbrouck, J. and Ho, T. S. (1987). Order arrival, quote behavior, and the return-generating process. *The Journal of Finance* **42**, 1035–1048.

Hasbrouck, J. and Seppi, D. J. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics* **59**, 383–411.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Forecasting systemic impact in financial networks. *International Journal of Forecasting* **30**, 781–794.

Herwartz, H. and Lütkepohl, H. (2011). Generalized least squares estimation for cointegration parameters under conditional heteroskedasticity. *Journal of Time Series Analysis* **32**, 281–291.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* **119**, 2249–2276.

Janková, J. and van de Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* **9**, 1205–1229.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Kalnina, I. and Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics* **147**, 47–59.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications* **126**, 3527–3577.

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. P. (2010). Estimating time-varying networks. *The Annals of Applied Statistics* **4**, 94–123.

Kong, X.-B. (2017). On the number of common factors with high-frequency data. *Biometrika* **104**, 397–410.

Kong, X.-B. (2018). On the systematic and idiosyncratic volatility with large panel high-frequency data. *The Annals of Statistics* **46**, 1077–1108.

Kong, X.-B., Lin, J.-G., Liu, C., and Liu, G.-Y. (2023). Discrepancy between global and local principal component analysis on large-panel high-frequency data. *Journal of the American Statistical Association* **118**, 1333–1344.

Koo, B. and Linton, O. (2012). Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics* **170**, 210–233.

Krampe, J. and Margaritella, L. (2021). Factor models with sparse var idiosyncratic components. *arXiv preprint arXiv:2112.07149* .

Kunitomo, N. and Kurisu, D. (2021). Detecting factors of quadratic variation in the presence of market microstructure noise. *Japanese Journal of Statistics and Data Science* **4**, 601–641.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics* **37**, 4254–4278.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics* **40**, 694–726.

Li, D., Chen, J., and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal* **14**, 387–408.

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

Li, D., Linton, O., and Lu, Z. (2015b). A flexible semiparametric forecasting model for time series. *Journal of Econometrics* **187**, 345–357.

Li, Y.-N., Li, D., and Fryzlewicz, P. (2023). Detection of multiple structural breaks in large covariance matrices. *Journal of Business & Economic Statistics* **41**, 846–861.

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

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40**, 2293–2326.

Liu, H. and Wang, L. (2017). TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics* **11**, 241–294.

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.

Liu, L. and Zhang, D. (2021). Robust estimation of high-dimensional vector autoregressive models. *arXiv preprint arXiv:2109.10354* .

Loh, P.-L. and Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *The Annals of Statistics* **41**, 3022–3049.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

McCracken, M. and Ng, S. (2020). FRED-QD: A quarterly database for macroeconomic research. Working paper available at https://www.nber.org/papers/w26872 .

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

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields* **151**, 435–474.

Miao, K., Phillips, P. C., and Su, L. (2023). High-dimensional vars with common factors. *Journal of Econometrics* **233**, 155–183.

Motta, G., Hafner, C. M., and von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory* **27**, 1279–1319.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* **45**, 7–38.

Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical Review E* **66**, 016128.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* **92**, 1004–1016.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics* **208**, 23–42.

Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference* **136**, 1237–1257.

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**, 487–504.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance* **39**, 1127–1139.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13**, 341–360.

Safikhani, A. and Shojaie, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *Journal of the American Statistical Association* **117**, 251–264.

Seo, B. (2007). Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity. *Journal of Econometrics* **137**, 68–111.

Serrat, O. (2017). *Social network analysis*. Springer.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management* **21**, 49–58.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association* **83**, 1097–1107.

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.

Sun, T. and Zhang, C.-H. (2013). Sparse matrix inversion with scaled lasso. *The Journal of Machine Learning Research* **14**, 3385–3418.

Tang, C. Y., Fan, Y., and Kong, Y. (2020). Precision matrix estimation by inverse principal orthogonal decomposition. *Communications in Mathematical Research* **36**, 68–92.

Tao, M., Wang, Y., and Chen, X. (2013a). Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory* **29**, 838–856.

Tao, M., Wang, Y., Yao, Q., and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association* **106**, 1025–1040.

Tao, M., Wang, Y., and Zhou, H. H. (2013b). Optimal sparse volatility matrix estimation for high-dimensional itô processes with measurement errors. *The Annals of Statistics* **41**, 1816–1864.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**, 267–288.

Uematsu, Y. and Yamagata, T. (2023a). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics* **41**, 213–227.

Uematsu, Y. and Yamagata, T. (2023b). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics* **41**, 126–139.

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models **42**, 1166–1202.

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*, volume 48. Cambridge university press.

Wang, D., Yu, Y., and Rinaldo, A. (2021a). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics* **49**, 203–232.

Wang, H., Peng, B., Li, D., and Leng, C. (2021b). Nonparametric estimation of large covariance matrices with conditional sparsity. *Journal of Econometrics* **223**, 53–72.

Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association* **104**, 747–757.

Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association* **103**, 1556–1569.

Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics* **38**, 943–978.

Wu, C., Zhao, H., Fang, H., and Deng, M. (2017). Graphical model selection with latent variables. *Electronic Journal of Statistics* **11**, 3485–3521.

Xu, M., Chen, X., and Wu, W. B. (2020). Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* **22**, 55.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics* **40**, 2541–2571.

Yan, Y., Gao, J., and Peng, B. (2020). A class of time-varying vector moving average (∞) models: Nonparametric kernel estimation and application. *arXiv preprint arXiv:2010.01492* .

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The 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, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 217–242.

Zhang, D. and Wu, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics* **49**, 233–254.

Zhang, R., Robinson, P., and Yao, Q. (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association* **114**, 916–927.

Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *The Annals of Statistics* **40**, 1376–1402.

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. *The Annals of Statistics* **36**, 1509–1533.