Research Collection School Of Economics

School of Economics

High-dimensional VARs with common factors

Ke MIAO

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Liangjun SU

Citation

HIGH-DIMENSIONAL VARS WITH COMMON FACTORS

By

Ke Miao, Peter C.B. Phillips, and Liangjun Su

July 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2252

High-Dimensional VARs with Common Factors[*]

Ke Miao[a], Peter C.B. Phillips[b] and Liangjun Su[c]

[a] School of Economics, Fudan University

[b] Yale Univeristy, University of Auckland,

University of Southampton, & Singapore Management University

[c] School of Economics and Management, Tsinghua University

July 7, 2020

Abstract

This paper studies high-dimensional vector autoregressions (VARs) augmented with common factors that allow for strong cross section dependence. Models of this type provide a convenient mechanism for accommodating the interconnectedness and temporal co-variability that are often present in large dimensional systems. We propose an ℓ_1-nuclear-norm regularized estimator and derive non-asymptotic upper bounds for the estimation errors as well as large sample asymptotics for the estimates. A singular value thresholding procedure is used to determine the correct number of factors with probability approaching one. Both the LASSO estimator and the conservative LASSO estimator are employed to improve estimation precision. The conservative LASSO estimates of the non-zero coefficients are shown to be asymptotically equivalent to the oracle least squares estimates. Simulations demonstrate that our estimators perform reasonably well in finite samples given the complex high dimensional nature of the model with multiple unobserved components. In an empirical illustration we apply the methodology to explore the dynamic connectedness in the volatilities of financial asset prices and the transmission of 'investor fear'. The findings reveal that a large proportion of connectedness is due to common factors. Conditional on the presence of these common factors, the results still document remarkable connectedness due to the interactions between the individual variables, thereby supporting a common factor augmented VAR specification.

JEL Classification: C13; C33; C38; C51

Keywords: Common factors; Connectedness; Cross-sectional dependence; Investor fear; High-dimensional VAR; Nuclear-norm regularization.

1   Introduction

In a pathbreaking study, Mann and Wald (1943) introduced vector autoregressions (VARs) and developed the limit theory for estimation and inference.[1] The VAR approach was further developed and promoted for empirical macroeconomic research in an influential paper by Sims (1980). Since then, the methodology has become one of the most heavily used tools in the applied finance and macroeconomics literature. It offers a simple and useful method of capturing rich dynamics and interconnectedness in multiple time series. Unrestricted VARs can be efficiently estimated by least squares regressions, which makes them particularly attractive in applied research. But low dimensional VARs often suffer from the notorious omitted variable bias problem, which makes the approach vulnerable to misleading inferences on both coefficients and impulse responses. In a series of articles (e.g., Sims (1992), Sims (1993), and Leeper et al. (1996)) Sims and his coauthors have explored whether to include more variables in VAR formulations to improve the forecasting performance.

In the absence of restrictions, the number of VAR coefficients increases quadratically, making the VAR estimation inevitably a high dimensional problem as the number of variables increases. The dynamic factor model (DFM), introduced by Geweke (1977), provides a synthetic tool to summarize useful information from a large number of time series while avoiding some of the problems of high dimensionality. Since then, a large literature has emerged on DFMs. Examples of theoretical work include Forni et al. (2000), Bai and Ng (2002), Bai (2003), and Hallin and Liška (2007). In applied finance and macroeconomics, various studies document the useful capacity of DFMs in capturing comovements among macroeconomic or financial time series; see Fama and French (1993), Stock and Watson(1999, 2002), Giannone et al.(2004), Ludvigson and Ng (2007), and Cheng and Hansen (2015), among many others. In an important work, Bernanke et al. (2005) propose a factor-augmented VAR (FAVAR) model to assist in making structural inferences while avoiding the problem of information sparsity that occurs in low dimensional VAR systems. Although the presence of common factors helps to capture additional variation and co-variation in the data, there is still evidence to suggest that misspecification continues to play a role in applied work with DFMs, particularly in forecasting. Stock and Watson (2005), for instance, test the ability of cross variation in forecasting, namely, whether observations on another variable such as x_{jt} help in predicting x_{it} given lagged values of x_{it} and common factors using 132 U.S. macroeconomic time series. Their results suggest that exclusion of other variables like x_{jt} from the regression equation for x_{it} involves misspecifications that can impair forecasting performance. A systematic approach to dealing with potential misspecifications

---

[1]The extension to the structural VAR (SVAR) case was developed in the final section of Mann and Wald (1943); but this seems largely to have been ignored in the vast literature on SVAR. For further discussions, see Hurn et al. (2020).

of this type is to employ modern machine learning methods that rely on regularized estimation. The present paper seeks to attain this goal in the context of large dimensional FAVAR systems.

Regularized estimation has recently received intense attention in both econometrics and statistics. In the cross-sectional framework, among the most influential works are Tibshirani (1996), Zhao and Yu (2006), Zou (2006), Candes and Tao (2007) and Huang et al. (2008). Inspired by the methods developed in these papers a growing body of literature on high dimensional autoregressive models has emerged. Haufe et al. (2010) propose a group-LASSO-based method to discover causal effects in multivariate time series. Basu and Michailidis (2015) study deviation bounds for Gaussian processes and investigate the ℓ_1 regularized estimation of transition matrices in sparse VAR models. Kock and Callot (2015) establish oracle inequalities for high dimensional VAR models. Han et al. (2015) propose a generalized Dantzig selector in high dimensional VARs. Guo et al. (2016) study a class of VAR models with banded coefficient matrices. These methods have opened up new avenues for handling high dimensional VAR models in practical work. In particular, regularized estimation has now been employed in various empirical applications in economic and financial analyses. For example, Smeekes and Wijler (2018) study forecasting capabilities of penalized regressions in cases where the generating process is a factor model; Medeiros et al. (2019) consider inflation forecasting with machine learning methods; Uematsu and Tanaka (2019) examine high-dimensional forecasting and variable selections via folded-concave penalized regressions; and Barigozzi and Brownlees (2019), Barigozzi and Hallin (2017), and Demirer et al. (2018) adopt high dimensional VARs to estimate networks and construct measures of financial sector connectedness.

All the aforementioned studies assume that the model's idiosyncratic errors exhibit at most weak cross-sectional dependence (c.f., Chudik et al., 2011). However, the vast literature on the DFM indicates that this assumption is fragile in empirical applications. In response to this limitation, the present paper proposes a new high dimensional VAR model in which some common factors (CFs) feature in the determination of each time series besides the idiosyncratic errors and lagged values of the time series themselves. In an earlier work, Chudik and Pesaran (2011) consider a factor-augmented infinite dimensional VAR model. For simplicity, they construct a model in which the factor-induced strong cross section dependence is explicitly separated from other sources of cross section dependence. They mention the possibility of using high dimensional VAR models with CFs but do not explicitly analyze the model. The FAVAR system in the present paper allows for serial correlation among the CFs, which in turn leads to correlation between the CFs and the lagged time series. To properly control for the presence of CFs in this FAVAR system it is necessary to estimate the factors, factor loadings, and transition matrices simultaneously. Practical implementation also requires the determination of the number of factors and lag length.

To estimate the high dimensional VAR model with CFs, we consider a three-step procedure.

In the first step, we consider an ℓ_1-nuclear-norm regularized least squares estimation problem that minimizes the sum of squared residuals with an ℓ_1-norm penalty imposed on the transition matrices and a nuclear norm penalty on the low rank matrix Θ representing the common component. Imposing the ℓ_1-norm penalty helps to estimate sparse transition matrices. The nuclear norm penalty helps to estimate the low rank matrix arising from the CFs and factor loadings. The ℓ_1-norm regularization has become standard in statistics and econometrics since the pioneering work of Tibshirani (1996). The nuclear norm regularization has recently become popular in the estimation of low rank matrices in statistics and econometrics; see, Negahban and Wainwright (2011), Rohde and Tsybakov (2011), Negahban et al. (2012), Chernozhukov et al. (2018), Belloni et al. (2019), Fan et al. (2019), Feng (2019), Koltchinskii et al. (2019), Moon and Weidner (2019), and Ma et al. (2020b), among others. All these previous works focus on the error bounds (in Frobenius norm) for the nuclear norm regularized estimates, except Chernozhukov et al. (2018), Moon and Weidner (2019) and Ma et al. (2020b) who study inference in linear or nonlinear panel data models with a low-rank structure. Like the latter authors, we simply use the nuclear norm regularization to obtain consistent initial estimates. Under some regularity conditions, we establish the nonasymptotic bounds for the estimation error of the transition matrices and the low rank matrix Θ. Applying a singular value thresholding (SVT) procedure on the singular values of the estimate of Θ, we obtain an estimate of the number of factors. We also show that the true number of factors can be estimated correctly with probability approaching one (w.p.a.1). Then, given the estimated factor number, preliminary estimates of the CFs can be obtained.

In the second step, we include the estimated CFs as regressors and consider a generalized LASSO estimator to obtain an updated estimate of the transition matrices. We show that the estimation errors can be uniformly controlled, which facilitates the construction of weights for subsequent estimation by conservative LASSO in the third step. Under some regularity conditions, we show that this third step conservative LASSO estimator of the transition matrices achieves sign consistency (see Zhao and Yu, 2006). Besides, the third step estimator of the transition matrices, factors and factor loadings are asymptotically equivalent to the corresponding oracle least squares estimators that are obtained by using detailed information about the form of the true regression model. We also study the asymptotic properties of these oracle least squares estimators and find that they perform as well as if the true common factors were known.

We illustrate the usefulness of our methodology through a real-data example. We revisit the financial connectedness measures proposed by Diebold and Yilmaz (2014) and document strong evidence of the existence of CFs in the volatilities of 23 sector exchange traded funds (ETFs). The findings show that CFs account for a large proportion of the variation in these volatilities; and, conditional on the CFs, a high level of connectedness remains present among the idiosyncratic components. This

empirical application demonstrates the particular usefulness of our high dimensional VAR model with CFs in its ability to allow for time series with strong cross section dependence while distinguishing variations that originates from different sources.

The remainder of the paper is organized as follows. In Section 2, we introduce our model and conduct a stationarity analysis. Section 3 introduces the estimation methods and examines their theoretical properties. In Section 4, we conduct Monte Carlo experiments to evaluate the finite sample performance of our estimators. We apply the model and methods to study financial connectedness in Section 5. Section 6 concludes. Proofs of the main results in the paper are given in the Appendix. Further technical details are provided in the online Supplementary Material.

1.1 Notation

To proceed, we introduce some notation. Let $A = (a_{ij}) \in \mathbb{R}^{M \times N}$ and $v = (v_1, ..., v_N)' \in \mathbb{R}^N$ be a matrix and a vector, respectively. We denote v_I as the subvector of v whose entries are indexed by a set $I \subset [N] \equiv \{1, ..., N\}$. We denote $A_{I,J}$ as the submatrix of A whose rows and columns are indexed by I and J, respectively. Let $A_{*,J} \equiv A_{[N],J}$ be the submatrix of A whose columns are indexed by J, $A_{I,*} \equiv A_{I,[M]}$ be the submatrix of A whose rows are indexed by I. For notational simplicity, we also write the individual columns and rows of A respectively as $A_{*,j} = A_{*,\{j\}}$ for $j \in [N]$ and $A_{i,*} = A_{\{i\},*}$ for $i \in [M]$.

For a random variable or vector x, we denote its expectation and ℓ_p-norm as $E(x)$ and $|||x|||_p \equiv [E(|x|_p^p)]^{1/p}$. We define the ℓ_0, ℓ_q ($q \geq 1$), and ℓ_∞ norms of a vector v to be

$$|v|_0 \equiv \sum_{i=1}^N \mathbf{1}(v_i \neq 0), \ |v|_q \equiv \left(\sum_{i=1}^N |v_i|^q \right)^{1/q}, \text{ and } |v|_\infty \equiv \max_{1 \leq i \leq N} |v_i|,$$

where $\mathbf{1}(\cdot)$ is the indicator function. In the special case $q = 2$, $|\cdot|_2$ is the Euclidean norm of v. We write $|v| \equiv |v|_2$ for notational simplicity.

For $1 \leq q < \infty$, we define the ℓ_q, $\ell_{\max}$, Frobenius (F), and nuclear ($*$) norms of the matrix A to be:

$$||A||_q \equiv \max_{||v||_q=1} ||Av||_q, \ ||A||_{\max} \equiv \max_{i,j} |a_{ij}|, \ ||A||_F \equiv \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2} \text{ and } ||A||_* = \sum_{k=1}^{\min(N,M)} \psi_k(A),$$

where $\psi_k(\cdot)$ is the kth largest singular value of A for $k = 1, ..., \min(N, M)$. We also denote the largest and smallest singular value of A as $\psi_{\max}(A)$ and $\psi_{\min}(A)$. In the special case $q = 2$, the ℓ_2 matrix norm is given by $||A||_2 = ||A||_{\mathrm{op}} \equiv \psi_1(A)$.

For a full rank $T \times R$ matrix F with $T > R$, we denote the corresponding orthogonal projection

matrices as $\mathbb{P}_F = F(F'F)^{-1}F'$ and $\mathbb{M}_F = I_T - \mathbb{P}_F$, where I_T denotes the $T \times T$ identity matrix. Let vec$(\cdot)$ denote the (columnwise) vectorization operator, and $\otimes$ be the (right hand) Kronecker operator. Let $\vee$ and $\wedge$ denote max and min operators, viz., $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

2 Model

For a N-dimensional vector-valued time series $\{Y_t\} = \{(y_{1t}, ..., y_{Nt})'\}$, the high-dimensional vector autoregression model of order p with CFs is given by:

$$Y_t = \sum_{j=1}^{p} A_j^0 Y_{t-j} + \Lambda^0 f_t^0 + u_t, \quad t = 1, ..., T, \tag{2.1}$$

where $A_1^0, ..., A_p^0$ are $N \times N$ transition matrices, $\Lambda^0 = (\lambda_1^0, ..., \lambda_N^0)'$ is an $N \times R^0$ factor loading matrix, f_t^0 is an R^0-dimensional vector of common factors, and $u_t \equiv (u_{1t}, ..., u_{Nt})'$ is an N-dimensional vector of unobserved idiosyncratic errors. Throughout this paper we use the superscript 0 to denote true values. The coefficients of interest are A_j^0's, Λ^0, and $F^0 \equiv (f_1^0, ..., f_T^0)'$. In practice, we need to determine the number of factors and the VAR order p. We propose a method to consistently determine p in Section 3. The number of factors can be determined in the first step of our estimation procedure introduced in Section 3. We consider the framework that both the number of cross-sectional units N and the time periods T go to infinity. The estimation is a natural high-dimensional problem with the number of parameters, $N^2 p + R^0 N + R^0 T$, growing linearly with T and quadratically with N.

It is convenient to reformulate model (2.1) as a multivariate regression problem in the form

$$\underbrace{\begin{bmatrix} Y_1' \\ \vdots \\ Y_T' \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} Y_0' & \cdots & Y_{1-p}' \\ \vdots & \ddots & \vdots \\ Y_{T-1}' & \cdots & Y_{T-p}' \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} A_1^{0\prime} \\ \vdots \\ A_p^{0\prime} \end{bmatrix}}_{B^0} + \underbrace{\begin{bmatrix} f_1^{0\prime} \\ \vdots \\ f_T^{0\prime} \end{bmatrix}}_{F^0} \underbrace{\begin{bmatrix} \lambda_1^{0\prime} \\ \vdots \\ \lambda_N^{0\prime} \end{bmatrix}'}_{\Lambda^{0\prime}} + \underbrace{\begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix}}_{\mathbf{U}}, \tag{2.2}$$

where $\mathbf{Y} \in \mathbb{R}^{T \times N}$, $\mathbf{X} \in \mathbb{R}^{T \times Np}$, $B^0 \in \mathbb{R}^{Np \times N}$, and $\mathbf{U} \in \mathbb{R}^{T \times N}$. A key observation here is that $\Theta^0 \equiv F^0 \Lambda^{0\prime}$ is a low rank matrix. However, due to the correlation between $\mathbf{X}B^0$ and Θ^0, the direct use of principal component analysis (PCA) on $\mathbf{Y}$ cannot deliver a consistent estimate of the common factors. Note that under some regularity conditions, both $||\mathbf{X}B^0||_{\text{op}}$ and $||\Theta^0||_{\text{op}}$ are $O_P(\sqrt{NT})$ and $||\mathbf{U}||_{\text{op}} = O_P(\sqrt{N} + \sqrt{T})$. We cannot separate the low rank matrix Θ^0 from $\mathbf{Y}$ without information about B^0. Besides, when the common factors are themselves serially correlated, pure VAR(p) estimation generally suffers from the endogeneity bias issues.

6

2.1 Stationarity analysis

Let $X_t' \equiv \mathbf{X}_{t,*}$. The N-dimensional VAR(p) process $\{Y_t\}$ can be rewritten in a companion form as an Np-dimensional VAR(1) process with CFs, viz.,

$$
\underbrace{\begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix}}_{X_{t+1}} = \underbrace{\begin{bmatrix} A_1^0 & A_2^0 & \cdots & A_{p-1}^0 & A_p^0 \\ I_N & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_N & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & I_N & \mathbf{0} \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix}}_{X_t} + \underbrace{\begin{bmatrix} \Lambda^0 f_t^0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\mathcal{F}_t} + \underbrace{\begin{bmatrix} u_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\mathcal{U}_t}. \tag{2.3}
$$

If one treats $\mathcal{F}_t + \mathcal{U}_t$ as an impulse at period t, the process $\{X_{t+1}\}$ in (2.3) can be regarded as a high-dimensional VAR(1) process. We can write the reverse characteristic polynomial (Lükepohl, 2005) of Y_t as

$$
\mathcal{A}(z) \equiv I_N - \sum_{j=1}^{p} A_j^0 z^p.
$$

In the low-dimensional framework, the process is stationary if $\mathcal{A}(z)$ has no roots in and on the complex unit circle, or equivalently the largest modulus of the eigenvalues of Φ is less than 1. To achieve identification, we need to study the Gram or signal matrix $S_X \equiv \mathbf{X}'\mathbf{X}/T$ and $\Sigma_X = E(X_t X_t')$. Basu and Michailidis (2015; hereafter BM) study the deviation bounds for the Gram matrix, using a Gaussianity assumption and boundedness of the spectral density function. Following their lead, we impose some conditions that will ensure S_X to be well behaved.

To proceed, we write X_{t+1} as a moving average process of infinite order (MA(∞)):

$$
X_{t+1} = \sum_{j=0}^{\infty} \Phi^j (\mathcal{F}_{t-j} + \mathcal{U}_{t-j}) \equiv X_{t+1}^{(f)} + X_{t+1}^{(u)}, \tag{2.4}
$$

where $X_{t+1}^{(f)} \equiv \sum_{j=0}^{\infty} \Phi^j \mathcal{F}_{t-j}$ and $X_{t+1}^{(u)} \equiv \sum_{j=0}^{\infty} \Phi^j \mathcal{U}_{t-j}$. Then we can study the stationarity of Y_t by studying $X_{t+1}^{(f)}$ and $X_{t+1}^{(u)}$, respectively. First, we consider $X_{t+1}^{(f)}$, which is the component due to the common factors. Note that the covariance matrix of $\mathcal{F}_t$ is a high-dimensional matrix with rank R^0 and explosive nonzero eigenvalues. Even if the largest modulus of the eigenvalues of Φ is smaller than 1, the variances of entries of $X_{t+1}^{(f)}$ are not assured to be uniformly bounded. Specifically, we consider $y_{it}^{(f)}$, which is the ith entry of $X_{t+1}^{(f)}$. Let $e_{j,M}$ be the jth unit M-dimensional vector. Noting

7

that $y_{it}^{(f)} = (e_{1,p} \otimes e_{i,N})'X_{t+1}^{(f)}$, we can write $y_{it}^{(f)}$ as an MA(∞) process

$$y_{it}^{(f)} = \sum_{j=0}^{\infty} (e_{1,p} \otimes e_{i,N})' \Phi^j (e_{1,p} \otimes \Lambda^0) f_{t-j}^0 \equiv \sum_{j=0}^{\infty} \alpha_{iN}^{(f)}(j) f_{t-j}^0,$$

where f_t^0 are allowed to be serially correlated. To ensure $y_{it}^{(f)} = O_P(1)$, we need to require the coefficients $\alpha_{iN}^{(f)}(j)$ to be well-behaved. Note that we generally do not have $||\Phi||_{\mathrm{op}} \leq 1$, as explained in the supplement of BM (2015). In Assumption A.1 below, we impose sufficient conditions that ensure $y_{it}^{(f)}$ are well-behaved. The online supplementary material provides a discussion of these conditions.

For the process $\{X_{t+1}^{(u)}\}$, stationarity is assured if we assume the covariance matrix of u_t is well-behaved and u_t is serially uncorrelated as in BM (2015) and Kock and Callot (2015; hereafter KC). Similarly to $y_{it}^{(f)}$, we define $y_{it}^{(u)}$ such that

$$y_{it} \equiv y_{it}^{(f)} + y_{it}^{(u)}, \tag{2.5}$$

where

$$y_{it}^{(u)} = \sum_{j=0}^{\infty} \alpha_{iN}^{(u)}(j) u_{t-j} \text{ and } \alpha_{iN}^{(u)}(j) \equiv (e_{1,p} \otimes e_{i,N})' \Phi^j (e_{1,p} \otimes I_N).$$

Again, imposing zero serial correlation and weak cross-sectional correlation across u_{it}'s is not enough to ensure $y_{it}^{(u)} = O_P(1)$ uniformly.

Let $\underline{c}$ and $\bar{c}$ denote generic constants that may vary across occurrences. Throughout the paper, we will treat Λ^0 as nonrandom. To ensure the stationarity of $\{Y_t\}$, we impose the following assumption.

Assumption A.1. (i) $u_t = C^{(u)} \epsilon_t^{(u)}$, where $\epsilon_t^{(u)} = (\epsilon_{1,t}^{(u)}, ..., \epsilon_{m,t}^{(u)})'$, $\epsilon_{i,t}^{(u)}$'s are i.i.d. random variables across (i,t) with mean zero and variance 1, and $C^{(u)}$ is an $N \times m$ matrix such that $C^{(u)} C^{(u)\prime} = \Sigma_u$ and $\underline{c} \leq \psi_{\min}(\Sigma_u) \leq \psi_{\max}(\Sigma_u) \leq \bar{c}$;

(ii) $\{f_t^0\}$ follows a strictly stationary linear process:

$$f_t^0 - \mu_f = \sum_{j=0}^{\infty} C_j^{(f)} \epsilon_{t-j}^{(f)},$$

where $\epsilon_t^{(f)} \equiv (\epsilon_{1,t}^{(f)}, ..., \epsilon_{R^0,t}^{(f)})'$ are i.i.d. with mean 0 and covariance matrix I_{R^0} across t, $\sup_{m \geq 1}(m + 1)^{\alpha} \sum_{j=m}^{\infty} ||C_j^{(f)}||_{\max} \leq \bar{c} < \infty$ for some constant $\alpha > 1$;

(iii) $\max_{1 \leq r \leq R^0} |||\epsilon_{r,t}^{(f)}|||_q < \bar{c}$ and $\max_{1 \leq i \leq m} |||\epsilon_{i,t}^{(u)}|||_q < \bar{c}$ for some $q > 4$;

(iv) $\{\epsilon_t^{(u)}\}$ is independent of $\{\epsilon_t^{(f)}\}$;

(v) the largest modulus of the eigenvalues of Φ is bounded by some constant $\rho \in (0,1)$;

(vi) $||(\Phi^j)_{[N],[N]}||_{\mathrm{op}} \leq \bar{c}\rho^j$ and $|\alpha_{iN}^{(f)}(j)| < \bar{c}\rho^j$;

(vii) $\max_{|z|=1} \psi_{\max}(\mathcal{A}^*(z)\mathcal{A}(z)) \leq \bar{c}$, *where $|z|$ denotes the modulus of z in the complex plane, and $\mathcal{A}^*(z)$ denotes the conjugate transpose of $\mathcal{A}(z)$.*

Assumption A.1(i) is frequently made in high dimensional time series analysis; see, e.g., Bai (1996), Chen and Qin (2010) and Ma et al. (2020a). It requires that u_t be independent over t and weakly dependent across i. At the cost of more complicated notations, one can allow $\psi_{\min}(\Sigma_u)$ to converge to zero and $\psi_{\max}(\Sigma_u)$ to diverge to infinity, both at a slow rate. Assumption A.1(ii) assumes the common factors to be stationary and allows for weak serial correlation. The factors can have nonzero mean so that y'_{it}s can also have nonzero mean. Assumption A.1(iii) requires that both $\epsilon_{i,t}^{(u)}$ and $\epsilon_{i,t}^{(f)}$ have finite qth order moments, which is a weak assumption compared to the Gaussian distribution assumption of BM (2015) and KC (2015). Assumption A.1(iv) requires independence between $\{\epsilon_t^{(u)}\}$ and $\{\epsilon_t^{(f)}\}$, which facilitates separate study of $y_{it}^{(f)}$ and $y_{it}^{(u)}$. Assumption A.1(v) is a standard assumption to ensure stationarity. Assumption A.1(vi) is a high level condition to ensure that $E(y_{it}^2)$ is uniformly bounded. Assumption A.1(vii) helps to bound the minimum eigenvalue of Σ_X. By the inequalities

$$\max_{|z|=1} \Lambda_{\max}(\mathcal{A}^*(z)\mathcal{A}(z)) \leq (\max_{|z|=1} ||\mathcal{A}(z)||_{\mathrm{op}})^2 \leq 1 + \sum_{k=1}^{p} ||A_j^0||_{\mathrm{op}},$$

we can see that requiring all the A_j^0's to have finite operator norms is a sufficient condition.

The online Supplementary Material provides further discussion on Assumption A.1(vi)-(vii). The following proposition ensures the stationarity of the process $\{y_{it}\}$ and establishes a lower bound for $\psi_{\min}(\Sigma_X)$.

Proposition 2.1 *Suppose that Assumption A.1 holds. (i) Then Y_t is a stationary process, $\sup_i E(y_{it}^2) < \infty$, and*

$$\psi_{\min}(\Sigma_X) \geq \frac{\psi_{\min}(\Sigma_u)}{\max_{|z|=1} \psi_{\max}(\mathcal{A}^*(z)\mathcal{A}(z))}.$$

(ii) Let $\Sigma_{XF} \equiv E(X_t f_t^{0\prime})$, and $\Sigma \equiv \Sigma_X - \Sigma_{XF}\Sigma_F^{-1}\Sigma'_{XF}$. We again have $\psi_{\min}(\Sigma) \geq \frac{\psi_{\min}(\Sigma_u)}{\max_{|z|=1} \psi_{\max}(\mathcal{A}^*(z)\mathcal{A}(z))}$.

3 Estimation method and theoretical results

This section develops an estimation procedure for the model and establishes its properties, both asymptotic and non-asymptotic. The procedure assumes at this point that the VAR order p is known and that R^0 is unknown. In practice, we can determine p via the data-driven method introduced in Section 3.5. The number of factors can be determined consistently in the first estimation step.

3.1 First-step estimator

In the first step, we propose an ℓ_1-nuclear norm regularized estimator to estimate the coefficient matrix B^0 and the low rank matrix Θ^0 simultaneously. We impose a sparsity condition on B^0 and use ℓ_1-norm regularization to achieve the selection of regressors. Like Moon and Weidner (2019; hereafter MW) and Chernozhukov et al. (2018) we adopt nuclear norm regularized estimation to obtain initial consistent estimation of the low rank matrix Θ^0. The first step estimator is given by the following procedure.

First-step estimator: *Let* $\gamma_1 = \gamma_1(N,T) = c_1 T^{-1/2} \log N$ *and* $\gamma_2 = \gamma_2(N,T) = c_2(N^{-1/2} + T^{-1/2})$ *for some constants c_1 and c_2.*

1. *Estimate the coefficient matrix B^0 and the low rank matrix Θ^0 by running the following ℓ_1-nuclear norm regularized regression:*

$$
\begin{aligned}
(\tilde{B}, \tilde{\Theta}) &= \operatorname{argmin}_{(B,\Theta)} \mathcal{L}(B, \Theta), \text{ where} \\
\mathcal{L}(B, \Theta) &\equiv \frac{1}{2NT} \|\mathbf{Y} - \mathbf{X}B - \Theta\|_F^2 + \frac{\gamma_1}{N} |\mathrm{vec}(B)|_1 + \frac{\gamma_2}{\sqrt{NT}} \|\Theta\|_*.
\end{aligned}
\tag{3.1}
$$

2. *Estimate the number of factors R^0 by the singular value thresholding (SVT) as:*

$$
\hat{R} = \sum_{i=1}^{N \wedge T} \mathbf{1}\{\psi_i(\tilde{\Theta}) \geq (\gamma_2 \sqrt{NT} \|\tilde{\Theta}\|_{\mathrm{op}})^{1/2}\}.
$$

3. *Obtain a preliminary estimate of F^0. Let the singular value decomposition (SVD) of $\tilde{\Theta}$ be $\tilde{\Theta} = \tilde{U}\tilde{D}\tilde{V}'$, where $\tilde{D} = diag(\psi_1(\tilde{\Theta}), ..., \psi_{N \wedge T}(\tilde{\Theta}))$. Set $\tilde{F} = \sqrt{T}\tilde{U}_{*,[\hat{R}]}$.*

Remark 3.1. The objective function $\mathcal{L}(B, \Theta)$ is the sum of squared residuals with both the nuclear norm regularization on Θ and ℓ_1-regularization on B. To obtain the numerical solution, we can apply an EM type algorithm. In the E-step, we fix B and update the estimate of Θ. The solution can be obtained following the result of Lemma 1 of MW (2019).[2] In the M-step, we fix Θ and update B. The optimization problem can be decomposed to N LASSO-type linear regression problems.

3.1.1 Non-asymptotic results for the first-step estimator

In this subsection we establish the non-asymptotic properties of the first step estimator. In particular, for $\tilde{B}$ and $\tilde{\Theta}$, we establish a non-asymptotic inequality for their estimation errors. For $\hat{R}$, we show that $\hat{R} = R^0$ w.p.a.1.

---

[2]Let the SVD of A be $A = USV'$, where $S = diag(s_1, ..., s_q)$, with $q = rank(A)$. Then $\operatorname{argmin}_{\Theta}\left(\frac{1}{2}\|A - \Theta\|_F^2 + \gamma\|\Theta\|_*\right)$ is given by $U \cdot diag((s_1 - \gamma)_+, ..., (s_q - \gamma)_+) \cdot V'$, where $(s)_+ = \max(0, s)$.

To proceed, we introduce some notation and assumptions. We first introduce a key invertibility condition for the linear operator $(\Delta^{(1)}, \Delta^{(2)}) :\to \mathbf{X}\Delta^{(1)} + \Delta^{(2)}$ when $(\Delta^{(1)}, \Delta^{(2)})$ is restricted to lie in a 'cone'. A similar condition is imposed in MW (2019) and Chernozhukov et al. (2018). Following their lead, we refer to the condition as '*restricted strong convexity*'. To define the 'cone', let $J_i \subset [Np]$ be an index set such that $j \in J_i$ if and only if $B_{ji}^0 \neq 0$. Let $J_i^c = [Np]\backslash J_i$. Let the SVD of Θ^0 be $\Theta^0 = U^0 D^0 V^{0\prime}$. For a $T \times N$ matrix $\Delta^{(2)}$, define the operators

$$\mathcal{P}(\Delta^{(2)}) \equiv U_{*,[R^0]}^0 U_{*,[R^0]}^{0\prime} \Delta^{(2)} V_{*,[R^0]}^0 V_{*,[R^0]}^{0\prime} \text{ and } \mathcal{M}(\Delta^{(2)}) \equiv \Delta^{(2)} - \mathcal{P}(\Delta^{(2)}).$$

Hence, the operator $\mathcal{P}(\cdot)$ projects a matrix onto a 'low-rank' space which contains Θ^0. For some $c > 0$, the 'cone' $\mathcal{C}_{NT}(c) \subset \mathbb{R}^{Np \times N} \times \mathbb{R}^{T \times N}$ is a set of $(\Delta^{(1)}, \Delta^{(2)})$ satisfying the restriction:

$$\frac{\gamma_1 \sum_{i=1}^N |\Delta_{J_i^c,i}^{(1)}|_1}{N} + \frac{\gamma_2 \left\|\mathcal{M}(\Delta^{(2)})\right\|_*}{\sqrt{NT}} \leq c\frac{\gamma_1 \sum_{i=1}^N |\Delta_{J_i,i}^{(1)}|_1}{N} + c\frac{\gamma_2 \left\|\mathcal{P}(\Delta^{(2)})\right\|_*}{\sqrt{NT}}.$$

We impose the following condition.

Assumption A.2 (Restricted strong convexity) *If* $(\Delta^{(1)}, \Delta^{(2)}) \in \mathcal{C}_{NT}(c)$ *for some* $c > 0$, *then there exist constants* κ_c *and* κ_c' *such that*

$$\left\|\mathbf{X}\Delta^{(1)} + \Delta^{(2)}\right\|_F^2 \geq T \cdot \kappa_c' \left\|\Delta^{(1)}\right\|_F^2 + \kappa_c \left\|\Delta^{(2)}\right\|_F^2.$$

Let $k_i = |J_i|$, $K_J \equiv \sup_i k_i$ and $K_a \equiv \sum_{i=1}^N k_i/N$. The next assumption involves a regularity condition on the errors and a sparsity condition on the transition matrix.

Assumption A.3 (i) $\|\mathbf{U}\|_{\mathrm{op}}/\sqrt{NT} \leq \gamma_2/2$, *where* γ_2 *is the tuning parameter for the nuclear norm regularization;.*

(ii) $K_a = o(T\left(N^{-1/2} + T^{-1/2}\right)/(\log N)^2)$.

Assumption A.3(i) requires the idiosyncratic error matrix to have an operator norm of order $O_P(\sqrt{N} + \sqrt{T})$. This condition has become standard in the literature; see, e.g., Lu and Su (2016), Moon and Weidner (2017), Su and Wang (2017), Chernozhukov et al. (2018), and MW (2019). Moon and Weidner (2017) provide examples of conditions that ensure the above assumption. In particular, it holds if $\epsilon_{it}^{(u)}$'s are i.i.d. sub-Gaussian (see, e.g., Vershynin, 2018).

Assumption A.3(ii) impose some sparsity conditions on the transition matrix. We allow K_a (and thus K_J) to diverge to infinity at a rate slower than $T\left(N^{-1/2} + T^{-1/2}\right)/(\log N)^2$ for some of the results below. Such a sparsity condition can be relaxed to the approximate sparsity condition as in Belloni et al. (2012) but that extension is not pursued here.

Theorem 3.1 *Suppose that Assumptions A.1-A.3(i) hold. Then we have*

$$N^{-1/2} \left\| \tilde{B} - B^0 \right\|_F \leq \bar{c}(\gamma_1 \sqrt{K_a} \vee \gamma_2) \text{ and } (NT)^{-1/2} \left\| \tilde{\Theta} - \Theta^0 \right\|_F \leq \bar{c}(\gamma_1 \sqrt{K_a} \vee \gamma_2),$$

with probability at least $1 - \bar{c}'(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$ *for some finite positive constants $\underline{c}$, $\bar{c}$, and $\bar{c}'$.*

Theorem 3.1 establishes the non-asymptotic inequalities for the estimation errors of $\tilde{B}$ and $\tilde{\Theta}$ in terms of Frobenius norm. The inequalities are valid when both $N^2 T^{1-q/4}(\log N)^{-q/2}$ and $N^{2-\underline{c}\log N}$ are small. In general, the first term dominates the second one for finite q and divergent (N, T). If the error terms are sub-exponential, we can allow q to diverge to infinity in which case the second term could dominate the first one. To prove the above theorem, we need to establish a bound for $T^{-1}\|\mathbf{U}'\mathbf{X}\|_{\max}$. Specifically, we need to find a sharp probability bound for a partial sum like $T^{-1}\sum_{t=1}^{T} y_{i,t-k}u_{jt}$. We resort to a Nagaev-type inequality, as introduced by Wu (2005) and Wu and Wu (2016), allowing for both dependence among summands and non-Gaussianity. The summand $y_{i,t-k}u_{jt}$ has a nonlinear Wold presentation $y_{i,t-k}u_{jt} = g_{ijk}(\ldots, \epsilon_{t-1}, \epsilon_t)$, where $\epsilon_t \equiv (\epsilon_t^{(u)\prime}, \epsilon_t^{(f)\prime})'$ is i.i.d. random variables under Assumption A.1. Then one can verify that the *dependence-adjusted norm* (see Wu and Wu, 2016) of $y_{i,t-k}u_{jt}$ is well bounded so that one can obtain a sharp probability bound using the Nagaev-type inequality for nonlinear processes.

Despite the fact that Theorem 3.1 is a non-asymptotic result, it is interesting to examine its asymptotic implications under Assumption A.3(ii). Note that Assumption A.3(ii) implies that $\gamma_1 \sqrt{K_a} = o(N^{-1/4} + T^{-1/4})$. Consequently, Theorem 3.1 implies that both $N^{-1/2}\|\tilde{B} - B^0\|_F$ and $(NT)^{-1/2}\|\tilde{\Theta} - \Theta^0\|_F$ are $o_P(N^{-1/4} + T^{-1/4})$. This rate can be improved to $O_P(N^{-1/2} + T^{-1/2}\log N)$ if we restrict our attention to the case where $K_a = O(1)$.

Next, we impose an assumption on the common factor and the factor loadings.

Assumption A.4 (i) *There exists an $\bar{N}$ such that for all $N > \bar{N}$, $\|\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda\|_{\max} \leq \bar{c}N^{-1/2}$ for an $R^0 \times R^0$ matrix Σ_Λ and $\|\Lambda^0\|_{\max} \leq \bar{c}$;*

(ii) *Let $\Sigma_F = E(f_t^0 f_t^{0\prime})$. There are constants $s_1 > \cdots > s_{R^0} > 0$ so that s_j equals the jth largest eigenvalue of $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$.*

Assumption A.4 requires that the factors and the factor loadings are strong/pervasive with well-behaved sample second moments. Assumption A.4(ii) requires distinct eigenvalues of $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$ in order to identify the corresponding eigenvectors.

The next theorem establishes the consistency of $\hat{R}$ and the mean-square convergence rate of $\tilde{F}$.

Theorem 3.2 *Suppose Assumptions A.1-A.4 hold. There exist positive constants $\underline{c}$, $\bar{c}$ and $\bar{c}'$, and a random matrix $\tilde{H}$ depending on (F^0, Λ^0) such that (i) $\hat{R} = R^0$ and (ii) $||\tilde{F} - F^0\tilde{H}||_F/\sqrt{T} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2)$, both with probability larger than $1 - \bar{c}'(N^2T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$.*

Theorem 3.2(i) establishes the consistency of $\hat{R}$ and the mean-square convergence rate of $\tilde{F}$. Intuitively, since $\tilde{\Theta}$ is a consistent estimator of $\Theta^0 \equiv F^0\Lambda^{0\prime}$ with well-controlled estimation errors, we expect the first R^0 singular values of $\tilde{\Theta}$ to be $O_P(\sqrt{NT})$ and the other singular values to be $O_P[\sqrt{NT}(\gamma_1 \vee \gamma_2)]$. Then the hard SVT procedure can distinguish the $\sqrt{NT}$-order singular values from those of smaller order. Alternatively, given the consistency of $\tilde{B}$ established in Theorem 3.1, we can regard the 'residual' $\mathbf{Y} - \mathbf{X}\tilde{B}$ as an approximation of $F^0\Lambda^{0\prime} + \mathbf{U}$. It is easy to see that one can also apply the methods of Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) to determine the number of factors. Theorem 3.2(ii) establishes the convergence rate of $\tilde{F}$. The $R \times R$ transformation matrix $\tilde{H}$ is similar to the matrix H in Bai (2003).

3.2 Second-step estimator

In this subsection, we introduce the second-step estimator. The second-step estimator is a generalization of the LASSO estimator, which includes the estimated factor matrix $\tilde{F}$ as regressors. Our goal is to obtain an estimator of B^0 whose elements uniformly converge to the true values. Then the second-step estimator can be utilized to construct adaptive- or conservative-LASSO weights in the third step.

Second-step estimator: *Let $\gamma_3 = c_3(\gamma_1\sqrt{K_a} \vee \gamma_2)$ for some constant c_3. For each $i \in [N]$, solve the minimization problem:*

$$(\dot{B}'_{*,i}, \dot{\lambda}'_i)' = \mathrm{argmin}_{(v', \lambda')' \in \mathbb{R}^{Np+R^0}} \frac{1}{2T}||\mathbf{Y}_{*,i} - \mathbf{X}v - \tilde{F}\lambda||_F^2 + \gamma_3|v|_1, \qquad (3.2)$$

where the LASSO penalty is only imposed on the coefficients of $\mathbf{X}$. Then the second-step estimators of B^0 and Λ^0 are given by $\dot{B} = (\dot{B}_{*,1}, ..., \dot{B}_{*,N})$ and $\dot{\Lambda} = (\dot{\lambda}_1, ..., \dot{\lambda}_N)'$, respectively.

Remark 3.2. Note that the ℓ_1-norm penalty is only imposed on the coefficients of $\mathbf{X}$. In the proof of Theorem 3.3 below, we show that $\dot{B}_{*,i}$ solves the LASSO problem with dependent variable $\mathbb{M}_{\tilde{F}}\mathbf{Y}_{*,i}$ and regressors $\mathbb{M}_{\tilde{F}}\mathbf{X}$.

Below, we establish the non-asymptotic properties of $\dot{B}$.

3.2.1 Non-asymptotic results for the second step estimator

Recall that $\Sigma \equiv \Sigma_X - \Sigma_{XF}\Sigma_F^{-1}\Sigma'_{XF}$. Let $\tilde{\Sigma} = \mathbf{X}'\mathbb{M}_{\tilde{F}}\mathbf{X}/T$. By Proposition 2.1, $\psi_{\min}(\Sigma)$ is bounded below from 0. Nevertheless, there is no guarantee that the sample matrix $\tilde{\Sigma}$ be positive definite.

In fact, if $Np > T$, $\tilde{\Sigma}$ is always singular, which leads to $\min_{|v| \neq 0} \frac{v' \tilde{\Sigma} v}{|v|^2} = 0$. In this case, we follow Bickel et al. (2009) and KC (2015) to establish the *restricted eigenvalue condition*. Specifically, we replace the above minimum by a minimum over a smaller set. Let $J \subset [Np]$ be an index set and $J^c = [Np] \backslash J$. We say the restricted eigenvalue condition is satisfied for some $K \in [Np]$ if

$$\min_{\substack{|J| \leq K}} \min_{\substack{|v| \neq 0 \\ |v_{J^c}|_1 \leq 3|v_J|_1}} \frac{v' \tilde{\Sigma} v}{|v_J|^2} \equiv \kappa_{\tilde{\Sigma}}(K) > 0, \tag{3.3}$$

where $|J|$ denotes the cardinality of J. In (3.3), the minimum is restricted to those vectors that $|v_{J^c}|_1 \leq 3|v_J|_1$, where J has cardinality no bigger than K. In this restricted space, we can show that (3.3) is satisfied with a high probability for $K = K_J$.

The following theorem establishes the $\ell_{\max}$-norm bound for the estimation error of $\dot{B}$.

Theorem 3.3 *Suppose that Assumptions A.1-A.4 hold. Suppose that (3.3) is satisfied. Then*

$$||\dot{B} - B^0||_{max} \leq \max_{1 \leq i \leq N} |\dot{B}_{*,i} - B^0_{*,i}|_1 \leq \frac{48}{[\psi_{\min}(\Sigma_X)]^2} K_J \gamma_3$$

with probability larger than $1 - \bar{c}(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c} \log N})$ for some finite positive constants $\underline{c}$ and $\bar{c}$.

3.3   Third-step estimator

In the first and second steps, we impose penalties on all elements in the coefficient matrix B, which introduces asymptotic biases into the estimators of the transition matrices. Zou (2006) proposes an adaptive LASSO technique in a linear regression framework, which penalizes the true zero parameters more than the non-zero ones. Then he shows that the adaptive LASSO estimator is asymptotically equivalent to the oracle least-squares estimator, which is obtained with the information of relevant regressors. Kock and Callot (2015) also explore the adaptive LASSO method in the high-dimensional VAR framework.

In practice, the regressors with zero estimates in the preliminary stage, which are usually plain LASSO estimates, are excluded in the adaptive LASSO. Hence, any incorrect regressor exclusion by the preliminary stage estimates directly leads to wrong regressor selection of adaptive LASSO. To solve this problem, the conservative LASSO, which gives regressors that are excluded by the initial estimator a second chance, is introduced (e.g., Caner and Kock, 2018). In this subsection, we extend the conservative LASSO estimator to the framework of high dimensional VAR with CFs.

Third-step estimator (**Conservative LASSO**): *Implement the following procedure:*

1. (Set weights) *Let* $\gamma_4 = \gamma_4(N, T)$. *Let* W *be a* $Np \times N$ *matrix with entries*

$$w_{ki} = \begin{cases} 1 & \text{if } |\dot{B}_{ki}| < \alpha\gamma_4, \\ 0 & \text{if } |\dot{B}_{ki}| \geq \alpha\gamma_4, \end{cases} \tag{3.4}$$

   where $k \in [Np]$, $i \in [N]$, *and* $\alpha > 0$. *Set* $\hat{F}^{(0)} = \tilde{F}$.

2. (Update $\hat{B}^{(\ell)}$) *For integer* $\ell \geq 1$, *update the estimate of* B *and* Λ:

$$(\hat{B}_{*,i}^{(\ell)\prime}, \hat{\lambda}_i^{(\ell)\prime})' = \mathrm{argmin}_{(v', \lambda')' \in \mathbb{R}^{NP+\hat{R}}} \frac{1}{2T} \left\| \mathbf{Y}_{*,i} - \mathbf{X}v - \hat{F}^{(\ell-1)}\lambda \right\|_F^2 + \gamma_4 \sum_{k=1}^{pN} w_{ki} |v_k|,$$

   where v_k *is the kth entry of* v, $i \in [N]$. *Let* $\hat{B}^{(\ell)} \equiv (\hat{B}_{*,1}^{(\ell)}, ..., \hat{B}_{*,N}^{(\ell)})$.

3. (Update $\hat{F}^{(\ell)}$) *Obtain the SVD of* $\mathbf{Y} - \mathbf{X}\hat{B}^{(\ell)}$ *as* $\mathbf{Y} - \mathbf{X}\hat{B}^{(\ell)} = \hat{U}^{(\ell)}\hat{D}^{(\ell)}\hat{V}^{(\ell)\prime}$. *Obtain an updated estimate of* F^0 *as* $\hat{F}^{(\ell)} = \sqrt{T}\hat{U}_{*,[\hat{R}]}^{(\ell)}$. *Set* $\ell = \ell + 1$.

4. *Iterate steps 2-3 until numerical convergence. Denote the final estimators as* $\hat{B}$, $\hat{F}$ *and* $\hat{\Lambda}$.

Remark 3.3. Note that the weights do not change with iterations in the above procedure. It is worth mentioning that the weights w_{ki} can take various forms. For example, Caner and Kock (2018) also consider $w_{ki} \equiv \frac{\gamma_{\mathrm{prec}}}{|\dot{B}_{ki}| \vee \gamma_{\mathrm{prec}}}$, where $\gamma_{\mathrm{prec}} = \alpha\gamma_4$.

3.3.1 Asymptotic properties of the third-step estimator

In this subsection, we will establish two results: (i) the conservative LASSO estimator $\hat{B}^{(\ell)}$ has the variable-selection consistency; (ii) $\hat{B}$ is asymptotically equivalent to the oracle least squares estimator.

First, we introduce some notation. Following Zhao and Yu (2006) and Huang et al. (2008), we say that $\hat{B}^{(\ell)} =_s B^0$, or $\hat{B}^{(\ell)}$ is sign-consistent for B^0, if and only if $\mathrm{sgn}(\hat{B}_{*,i}^{(\ell)}) = \mathrm{sgn}(B_{*,i}^0)$ for all $i \in [N]$, where

$$\mathrm{sgn}(B_{*,i}) \equiv [\mathrm{sgn}(B_{1,i}), ..., \mathrm{sgn}(B_{Np,i})]', \text{ and } \mathrm{sgn}(B_{ki}) \equiv \begin{cases} 1 & \text{if } B_{ki} > 0 \\ 0 & \text{if } B_{ki} = 0 \\ -1 & \text{if } B_{ki} < 0 \end{cases}.$$

Assumption A.5 (i) *As* $(N, T) \to \infty$, *the magnitude of nonzero coefficients are of larger asymptotic order than* γ_4: $\gamma_4 = o(\min_{i \in [N]} \min_{k \in J_i} |B_{ki}^0|)$;

   (ii) $(K_J^{3/2} T^{-1/2} \log N + K_J^{1/2} N^{-1/2}) = o(\gamma_4)$ *and* $K_J K_a [T^{-1}(\log N)^2 + N^{-1}] = o(T^{-1/2})$;

   (iii) $N^2 T^{1-q/4} (\log N)^{-q/2} \to 0$ *and* $T/N^2 \to 0$ *as* $(N, T) \to \infty$.

Assumption A.5(i) assumes the nonzero entries of B^0 cannot be too small, which is a standard assumption in the adaptive LASSO literature. The lower bound $\min_{i\in[N]} \min_{k\in J_i} |B_{ki}^0|$ has to be larger than γ_4 in order to separate the nonzero entries from zeros. By Assumption A.5(i) and Theorem 3.3, we can show that $\max_{k\in J_i} w_{ki} = 0$ and $\min_{k\in J_i^c} w_{ki} = 1$ w.p.a.1. In this case, we only put penalty on the true zero entries. Assumption A.5(ii) imposes some conditions on K_J and K_a. This assumption ensures that $\|\mathbf{X}(\hat{B}^{(\ell)} - B^0)\|_F$ has a desired convergence rate. Assumption A.5(iii) imposes some conditions on the relative rates at which N and T pass to infinity and they depend on the number (q) of moments for the innovation processes in the error term and factors. In the special case where N and T pass to infinity at the same rate, this condition requires $q \geq 12$. This requirement greatly relaxes the sub-Gaussian assumption imposed on the error terms in the early literature.

The following theorem establishes the variable selection consistency of $\hat{B}^{(\ell)}$ and the preliminary convergence rates of $\hat{B}^{(\ell)}$ and $\hat{F}^{(\ell)}$.

Theorem 3.4 *Suppose that Assumptions A.1-A.5 hold. Then*
(i) $P(\hat{B}^{(\ell)} =_s B^0) \to 1$, as $(N, T) \to \infty$;
(ii) $\|\mathbf{X}(\hat{B}^{(\ell)} - B^0)\|_F / \sqrt{NT} = O_P(\gamma_1\sqrt{K_a} + \gamma_2)$;
(iii) $\|\hat{F}^{(\ell)} - F^0\tilde{H}\|_F / \sqrt{T} = O_P(\gamma_1\sqrt{K_a} + \gamma_2)$.

Theorem 3.4(i) shows that $\hat{B}^{(\ell)}$ has the oracle property in that it selects the correct variables w.p.a.1. Due to the presence of common factors and the possibly divergent number (k_i) of nonzero coefficients in $B_{*,i}^0$, we can only obtain a preliminary rate $O_P(\gamma_1\sqrt{K_a} + \gamma_2)$ in Theorem 3.4(ii)-(iii). Apparently, this rate depends on the average number (K_a) of nonzero coefficients in $B_{*,i}^0$'s.

To improve the rate of convergence, we study the final estimators $\hat{B}$, $\hat{F}$ and $\hat{\Lambda}$. Now, $\hat{F}$ corresponds to the first $\hat{R}$ eigenvectors of $(\mathbf{Y} - \mathbf{X}\hat{B})(\mathbf{Y} - \mathbf{X}\hat{B})'$, rescaled by $\sqrt{T}$, and one can expand $\hat{F} - F^0\tilde{H}$ following the lead of Bai and Ng (2002) and Bai (2009). By looking at the product of $\hat{F} - F^0\tilde{H}$ with other terms, we can derive a sharper bound for some intermediate estimates. Finally we can improve the probability order of each element in $\hat{B}_{J_i,i} - B_{J_i,i}^0$ to $O(T^{-1/2})$.

The following theorem reports the asymptotic distribution of $\hat{B}_{J_i,i}$.

Theorem 3.5 *Suppose that Assumptions A.1-A.5 hold. Let S_i denote an $L \times k_i$ selection matrix such that $\|S_i\|_F$ is finite and L is a fixed integer. Conditional on the event $\{\hat{B} =_s B^0\}$, for each $i \in [N]$, we have $\sqrt{T}S_i(\hat{B}_{J_i,i} - B_{J_i,i}^0) \xrightarrow{d} N(\mathbf{0}, \sigma_i^2 S_i(\Sigma_{J_i,J_i})^{-1}S_i')$ where $\sigma_i^2 = E\left(u_{it}^2\right)$.*

Note that we specify a selection matrix S_i in Theorem 3.5 that is not needed if k_i is fixed. Intuitively, we allow k_i to diverge to infinity as $(N, T) \to \infty$ and we cannot derive the asymptotic normality of $\hat{B}_{J_i,i}$ directly when $k_i \to \infty$. Instead, we follow standard practice on estimation and

inference with a divergent number of parameters (see, e.g., Fan and Peng, 2004, Lam and Fan, 2008, and Qian and Su, 2016) and prove the asymptotic normality for an arbitrary but finite number of linear combinations of the elements of $\hat{B}_{J_i,i}$. In the special case where k_i is fixed, we can take $S_i = I_{|J_i|}$ and obtain the usual joint asymptotic normal distribution for all elements of $\hat{B}_{J_i,i}$.

3.4   Tuning parameter selection

In practice, we need to select the tuning parameters γ_ℓ, for $\ell = 1, ..., 4$. For γ_2, which is the tuning parameter for the nuclear norm penalty, we adopt a simple plug-in approach similar to that introduced in Chernozhukov et al. (2018). An ideal tuning parameter for γ_2 is one such that

$$||\mathbf{U}||_{\mathrm{op}}/\sqrt{NT} \le (1-c)\gamma_2$$

for some $c > 0$ with high probability. Suppose $\mathbf{U}$ is a random matrix with i.i.d. sub-Gaussian entries that have mean zero and variance σ_u^2, its operator norm is bounded by $C\sigma_u(\sqrt{N} + \sqrt{T})$ for some $C > 0$ with high probability (see Vershynin, 2018). One can first use $\gamma_2 = \frac{\hat{\sigma}_y}{C}(\sqrt{N} + \sqrt{T})$ for some $C > 1$ and $\hat{\sigma}_y$ is the sample standard deviation of Y. After obtaining an estimate $\hat{\sigma}_u$ of σ_u, we can calculate a suitable γ_2 via simulation. Specifically, we can simulate the random matrices $\mathbf{U}$ with i.i.d. $N(0, \hat{\sigma}_u^2)$. Then we let $\gamma_2 = Q(||\mathbf{U}||_{\mathrm{op}}, 0.95)$, where $Q(x, \alpha)$ denote the αth quantile of x.

For γ_1, γ_3, and γ_4, we propose to use the five-fold cross validation (CV) process. Let $\gamma = (\gamma_1, \gamma_3, \gamma_4)'$. For the first-step estimation, the procedure goes as follows:

1. Partition the data into 5 separate sets along the time dimension: $\mathbf{T}_1, ..., \mathbf{T}_5 \subset [T]$;

2. For $k = 1, ..., 5$, fit the model to the training set by excluding the kth fold data. Denote the estimators by $\tilde{B}^{(\gamma,k)}$ and $\tilde{\Lambda}^{(\gamma,k)}$, where $\tilde{\Lambda}^{(\gamma,k)}$ is a $N \times R$ matrix containing the first R right singular vectors of $\tilde{\Theta}$. Calculate the sum of squared prediction errors

$$cv(\gamma, k) = \mathrm{tr}[(\mathbf{Y}_{\mathbf{T}_k,*} - \mathbf{X}_{\mathbf{T}_k,*}\tilde{B}^{(\gamma,k)})\mathbb{M}_{\tilde{\Lambda}^{(\gamma,k)}}(\mathbf{Y}_{\mathbf{T}_k,*} - \mathbf{X}_{\mathbf{T}_k,*}\tilde{B}^{(\gamma,k)})'];$$

3. Compute the CV error for a fixed tuning parameter by $CV(\gamma) = \sum_{k=1}^{5} cv(\gamma, k)$.

4. Select $\gamma^* = \operatorname{argmin}_\gamma CV(\gamma)$.

Remark 3.4. Once the sample $\mathbf{T}_k$ is excluded, we cannot obtain an estimate of $F_{\mathbf{T}_k,*}$. Hence we cannot obtain the residuals by deducting the estimate of $F_{\mathbf{T}_k,*}\Lambda'$. For this reason, we multiply $\mathbf{Y}_{\mathbf{T}_k,*} - \mathbf{X}_{\mathbf{T}_k,*}\tilde{B}^{(\gamma,k)}$ by $\mathbb{M}_{\tilde{\Lambda}^{(\gamma,k)}}$ to project out $F_{\mathbf{T}_k,*}\Lambda'$ in the above procedure.

For the second and third step estimators, the CV procedure can be constructed similarly.

3.5 Lag length selection

In the above estimation procedure, we have so far assumed that the lag length p is known. In practice, the lag length p is usually unknown and requires estimation. In this subsection, we propose a procedure to determine the lag length p. Suppose we estimate the model with $p_{\max} \geq p^0$, where we use the superscript '0' to denote the true parameter. The model with $p_{\max}$ lags continues to be a correctly specified model except that $A_k^0 = \mathbf{0}$ for $k > p^0$. Due to the LASSO regularization, the elements of the estimator $\hat{A}_p$ for $p > p^0$ should converge to zero. For this reason, we propose to determine the lag length by the following procedure:

1. Given $p_{\max}$, obtain the estimates $\hat{A}_k$ for $k \in [\, p_{\max}]$;

2. Calculate $a_k = ||\hat{A}_k||_{\mathrm{F}}^2 \vee c$ for some constant c and $k \in [p_{\max}]$;

3. The criterion function we consider is given by the ratio

$$GR(p) = \frac{\sum_{k=p}^{p_{\max}} a_k}{\sum_{k=p+1}^{p_{\max}} a_k}, \ \ p = 1, ..., p_{\max} - 1.$$

   The term GR refers to the growth ratio of $\sum_{k=p}^{p_{\max}} a_k$.

4. Obtain the estimator of p^0 as $\hat{p} = \mathrm{argmax}_{1 \leq k < p_{\max}} GR(k)$.

Remark 3.5. We make some remarks in order. First, one can also simply run an ℓ_1-nuclear penalized regression with $p_{\max}$, which is the first step of the estimation procedure given in Section 3.1. We only require that $||\hat{A}_k - A_k^0||_{\mathrm{F}}$ converge to zero at a certain rate. Second, in practice one may obtain a very small or even zero value for $||\hat{A}_k||_{\mathrm{F}}^2$ when $k > p^0$. In this case, if we directly use $a_k = ||\hat{A}_k||_{\mathrm{F}}^2$, the growth ratio may possibly choose a larger p than p^0. To solve this problem, we bound a_k below by some constant $c > 0$. Third, the $GR(p)$ criterion function is constructed to allow some A_k^0 with $k < p^0$ to be a matrix of zeros. If we believe all A_k^0's are nonzero matrices for $k \in [p^0]$, one can also consider the criterion function $FR(p) = a_p/a_{p+1}$, where the term FR refers to Frobenius norm ratio.

4 Monte Carlo Simulations

This section reports the results of set of Monte Carlo experiments designed to evaluate the finite sample performance of the proposed estimation procedure.

4.1 Data generating processes

We consider three cases with $p = 1$. For each data generating process (DGP), we generate the data from the following high dimensional VAR(1) process with CFs:

$$Y_t = A_1^0 Y_{t-1} + \Lambda^0 f_t^0 + u_t, \tag{4.1}$$

where A_1^0 varies across different DGPs, $\Lambda^0 = (\lambda_1^0, ..., \lambda_N^0)'$. The factor loadings λ_{ri}^0, for $r = 1, ..., R^0$, are independently and identically distributed (i.i.d.) standard normal random variables. The factors f_{tr}^0, for $r = 1, ..., R^0$, follow an autoregressive process:

$$f_{tr}^0 = \rho_f \cdot f_{t-1,r}^0 + \epsilon_{tr}^{(f)},$$

where $\rho_f = 0.6$ and $\epsilon_{tr}^{(f)}$ are i.i.d. $N(0,1)$. The idiosyncratic error terms are generated as $u_{it} = s \cdot \epsilon_{it}^{(u)}$, where s controls the signal-to-noise ratio, and $\epsilon_{it}^{(u)}$ are i.i.d. $N(0,1)$.

DGP 1 (Tridiagonal transition matrix): $(A_1^0)_{ij} = 0.3 \cdot \mathbf{1}(|i - j| \leq 1)$.

DGP 2 (Block-diagonal transition matrix): We generate a block-diagonal matrix $A_1^0 = \text{bdiag}(S_1, ..., S_K)$, where the S_k's are 5×5 random matrices. The diagonal entries of S_k are fixed with $(S_k)_{i,i} = 0.3$. In each column of S_k, we randomly choose 2 out of 4 off-diagonal entries and set them to be -0.3.

DGP 3 (Random transition matrix): We fix the diagonal entries of A_1^0 to be 0.3 (i.e. $(A_1^0)_{ii} = 0.3$). In each row of A_1^0, we randomly choose 3 out of $N - 1$ entries and set them to be -0.3.

FIGURE 1 around here

Figure 1 illustrates the structure of the random transition matrices used in our simulation. For each DGP, we consider $N = 30, 60$, and $T = 100, 200, 400$, leading to six combinations of cross-sectional and time series dimensions. The number of replications is set to 500.

4.2 Implementation and estimation results

For each DGP, we consider the feasible estimator proposed in this paper and the oracle least squares estimator. The oracle estimators are obtained by using the information of the number of factors and the true regressors.

Table 1 reports the model selection accuracy. For each combination of N and T in each DGP, the fourth and fifth columns report the under- and over-estimation rate of $\hat{R}$, respectively. The TPR (true positive rate) columns report the average shares of relevant variables included. The FPR (false positive rate) columns report the average shares of irrelevant variables included. We summarize some important findings from Table 1. First, the proposed hard singular value thresholding (SVT)

19

procedure can correctly determine the number of factors for each case. Second, with N fixed, the TPR increases with T in all cases as expected. All three-step estimators can include almost all the true regressors when $T = 400$. Third, among the three estimators, the third-step conservative LASSO estimator includes the least irrelevant regressors in almost all settings. In addition, only conservative LASSO estimators tend to exclude more irrelevant regressors as T increases, while the FPRs of the first and second step estimators increase as T grows.

TABLE 1 around here

Table 2 reports the estimation error of both the feasible estimators and the oracle least squares estimator. We report the root mean squared errors (RMSEs) for all entries and nonzero entries, respectively. We summarize some important findings from Table 2. First, as expected, the oracle least squares estimator uniformly outperforms the feasible estimators. This is mainly due to the fact that the FPRs of feasible estimators were never zero. Second, the RMSE of the oracle estimator for nonzero entries decreases with T at the $\sqrt{T}$-rate and alters with N slightly. This is consistent with our theoretical prediction that the oracle least squares estimator converges to the true values at the $\sqrt{T}$-rate. Third, the conservative LASSO outperforms the other two feasible estimators in terms of RMSEs in all cases.

TABLE 2 around here

For all DGPs, we also consider estimation of a misspecified VAR(1) model, $Y_t = A_1^0 Y_{t-1} + u_t$, where the common factors are ignored. We first estimate the model with LASSO as in KC (2015). Then we construct the weights as in (3.4) and use conservative LASSO to estimate the misspecified model. Table 3 reports the performance of these two estimators. We summarize some findings from Table 3. First, the FPRs for both estimators are quite high. This indicates that the misspecification may lead to non-sparse estimates of the transition matrices when the presence of strong cross-sectional dependence is not properly accounted for. Second, the estimators for the misspecified model also have higher RMSEs. Third, in many cases, the conservative LASSO estimator performs even worse than the LASSO estimator in terms of RMSEs. So it is important to take into account the factor structure in the estimation of a VAR with CFs.

TABLE 3 around here

5 Empirical application

5.1 Evaluating a network of financial assets volatilities

In recent years, financial asset connectedness has been an active topic in financial econometrics. Examples of contributions to this literature include Barigozzi and Brownlees (2019; hereafter BB), Barigozzi and Hallin (2017), Billio et al. (2012), Diebold and Yilmaz (2014; hereafter DY), Diebold and Yilmaz (2015), and Hautsch et al. (2014). Some of these authors directly model the large panel of time series as a VAR process without the potential presence of common factors. A LASSO-type method is employed to estimate the transition matrices. However, Barigozzi and Hallin (2017) and BB (2019) document evidence for the existence of a factor structure in volatility. Barigozzi and Hallin (2017) consider controlling for the presence of common factors by means of a dynamic factor model. BB (2019) use the regression residuals of individual volatilities on observed factors (e.g., market volatility or sector-specific volatility) to represent the idiosyncratic components of the volatilities. Neither of these papers provides theoretical justifications.

In this empirical application, we extend the measure of connectedness of DY and study the connectedness of financial assets. Specifically, we study the connectedness in a panel of volatility measures. As remarked by DY, the volatilities of financial assets can be interpreted as a form of 'investor fear'. Then volatility connectedness represents 'fear connectedness' across assets. In this scenario, it is natural to take into account common factors, which reflect confidence in the market. Spillover effects across assets is another reason for connectedness. We use the econometric methodology derived in the present work to analyze a panel of return volatilities of 23 sector ETF funds. The findings show that common factors account for 56.1% of the overall variability. Conditioning on these factors, the interdependence across individuals still captures a relatively high proportion of the variation.

<div align="center">Table 4 around here</div>

5.1.1 Data description and empirical framework

We collect the weekly 'open price', 'close price', 'high price' and 'low price' of a series of sector ETF funds from Yahoo finance. The fund names and tickers are listed in Table 4. They fall into several categories. The 'Energy', 'Financial' and 'Consumer cyclical' are three large categories, each of which contains three to four funds. Each of the other categories contain at most two funds. The sample spans July 2007 to August 2019, which corresponds to 688 weeks. As volatility is unobserved, we use the observed price data to estimate it. Specifically, we follow Garman and Klass (1980) and Alizadeh

et al. (2002) to measure the asset volatility as follows:

$$\tilde{\sigma}_{it}^2 = 0.511(H_{it} - L_{it})^2 - 0.019[(C_{it} - O_{it})(H_{it} + L_{it} - 2O_{it}) - 2(H_{it} - O_{it})(L_{it} - O_{it})]$$
$$- 0.383(C_{it} - O_{it})^2,$$

where O_{it}, C_{it}, H_{it}, and L_{it} are natural logarithms of weekly 'open price', 'close price', 'high price' and 'low price', respectively. We present the descriptive statistics of volatilities in Table 5. The kurtosis of each time series is quite large. We follow DY (2014) to normalize the data by taking natural logarithms and then centering each time series. That is, our y_{it} is given by $\log(\tilde{\sigma}_{it}^2) - \overline{\log(\tilde{\sigma}_{i\cdot}^2)}$.

Table 5 around here

Given the panel of volatilities, we fit the data to our VAR model with CFs in (2.1). By the decomposition (2.5), $y_{it} = y_{it}^{(f)} + y_{it}^{(u)}$, where $y_{it}^{(f)}$ is due to the common factors and $y_{it}^{(u)}$ is due to the idiosyncratic errors. Then $\nu_i \equiv \text{var}(y_{it}^{(f)})/\text{var}(y_{it})$ measures the proportion of variance in y_{it} that is due to common factors and $\bar{\nu} \equiv \sum_{i=1}^N \text{var}(y_{it}^{(f)})/\sum_{i=1}^N \text{var}(y_{it})$ measures the corresponding object in all time series.

For the idiosyncratic component $y_{it}^{(u)}$, we can calculate the measure of connectedness proposed by DY (2014). As discussed in Section 2, we have $y_{it}^{(u)} = \sum_{j=0}^\infty \alpha_{iN}^{(u)}(j)C^{(u)}\epsilon_{t-j}^{(u)}$, where $\alpha_{iN}^{(u)}(j) = (e_{1,p} \otimes e_{i,N})'\Phi^j(e_{1,p} \otimes I_N)$ and $\epsilon_t^{(u)} \sim (0, I_m)$. For simplicity, suppose that $m = N$. Then one can treat $\epsilon_{it}^{(u)}$ as the idiosyncratic shock to individual i. The variance of the H-step ahead prediction error due to $\{\epsilon_{j,t+h}^{(u)}\}_{h=1}^H$ is $s_{ij}^H = \sum_{h=0}^{H-1}([\alpha_{iN}^{(u)}(h)C^{(u)}]_j)^2$. If we can identify both Φ and $C^{(u)}$, we can easily estimate the variance decomposition matrix $\check{D}^H$ with (i,j)th entry $s_{ij}^H/\sum_{k=1}^N s_{ik}^H$. However, $C^{(u)}$ is not identified without further assumption. Although we cannot identify $C^{(u)}$, the matrix $\Sigma_u = C^{(u)}C^{(u)'}$ is identified. DY (2014) propose to calculate the H-step generalized variance decomposition matrix $D^H = [d_{ij}^H]_{N \times N}$, where

$$d_{ij}^H = \frac{\sigma_{jj}^{-1}\sum_{h=0}^{H-1}(\alpha_{iN}^{(u)}(h)\Sigma_u e_{j,N})^2}{\sum_{h=0}^{H-1}\alpha_{iN}^{(u)}(h)\Sigma_u\alpha_{iN}^{(u)}(h)'}, \text{ and } e_{j,N} \text{ is } j\text{th column of } I_N.$$

Unlike $\check{D}^H$, the row sums of D^H are not necessarily unity. We normalize D^H to $\tilde{D}^H$ with (i,j)th entry $\tilde{d}_{ij}^H = d_{ij}^H/\sum_{k=1}^N d_{ik}^H$ so that $\sum_{j=1}^N \tilde{d}_{ij}^H = 1$ and $\sum_{i,j=1}^N \tilde{d}_{ij}^H = N$. Hence, the overall connectedness in the $y_{it}^{(u)}$'s can be measured as $\tilde{d}^H = \sum_{i\neq j}\tilde{d}_{ij}^H/N$. In addition, we let $\tilde{d}_{i\leftarrow}^H \equiv \sum_{j\neq i}\tilde{d}_{ij}^H$. Following DY (2014), we call $\tilde{d}_{i\leftarrow}^H$ the 'FROM' index, as it measures the proportion of generalized variance decomposition that is due to other individuals. Similarly, we let $\tilde{d}_{\leftarrow j}^H \equiv \sum_{i\neq j}\tilde{d}_{ij}^H$ and call this the 'TO' index.

5.1.2 Estimation results

We use the procedure proposed in Section 3.4 to determine the lag length with $p_{\max} = 8$. The result gives $\hat{p} = 4$. When we run the regression with $p = 4$, the number of factors is determined to be one ($\hat{R} = 1$).

<div align="center">Figure 2 around here</div>

Figure 2 reports the heat map which represents the estimates of the $\hat{A}_k$'s. The element value is represented by scaled color. In total, 330 out of 2116 entries are nonzero. There are three interesting findings. First, most of the nonzero entries are estimated to be positive. The positive coefficients represent the propagation of investor fear across assets. Second, the diagonal elements of $\hat{A}_k$'s are mostly nonzero. The magnitude of the diagonal elements is larger than that of the off diagonal elements on average. Third, the number of nonzero coefficients in $\hat{A}_k$ decreases as k increases and the average magnitude of the entries also decreases. More recent investor fear causes greater present investor fear.

<div align="center">Table 6 around here</div>

Next, we calculate the statistics introduced in the last subsection. The upper panel of Table 6 provides the estimates of ν_i, $\tilde{d}^H_{i\leftarrow}$, and $\tilde{d}^H_{\leftarrow j}$. Almost all the ν_i's are above 50%, and the overall variation due to the common factors is $\bar{\nu} = 56.1\%$. The market level investor fear is playing a dominant roll in investor trading behavior. After conditioning on the factors, we consider the idiosyncratic part by looking at $\tilde{d}^H_{i\leftarrow}$, $\tilde{d}^H_{\leftarrow j}$ and the H-step generalized variance decomposition matrix $\tilde{D}^H$. The 'FROM' index ranges between 27.7% and 71.7%. Interestingly, the 'energy' and 'finance' funds have higher 'FROM' index compared to other funds. A similar observation applies for the 'TO' index. Specifically, the 'TO' index of XLE and IYE are close to 100% and both are 'energy' funds. The energy industry therefore transmits considerable investor fear to the entire market. This finding is intuitive as the oil price has been extremely volatile in recent years and the energy price affects all industries. The fund GDX (VanEck Vectors Gold Miners ETF) has the least connectedness. It receives only 27.7% connectedness from other assets and transmits only 19.1% connectedness to others. The overall connectedness measure is 49.8%. Conditioning on the factors, there is still substantive transmission of investor fear across individuals. Figure 3 reports the heat map of the H-step generalized variance decomposition matrix $\tilde{D}^H$ at $H = 12$. We observe that the interconnections within the same category is high, whereas connectedness across categories is relatively low.

<div align="center">Figure 3 around here</div>

The lower panel of Table 6 provides the measure of connectedness with the pure VAR model es-

timation as in Demirer et al. (2018). Without controlling for the common factors, the 'FROM' and 'TO' index of each fund becomes much larger. However, we observe little heterogeneity across categories. In this case, all the connectedness due to common factors is interpreted as the individual level connectedness, which potentially leads to wrong inference.

In sum, our framework extends the traditional VAR analysis of financial asset connectedness to control for the presence of common factors in the determination of volatility. We have found that common factors account for more than a half of the variation in the data. In addition to the connectedness that is due to common factors there is still a remarkable degree of connectedness that arises from spillover channels that operate among the assets themselves.

6   Conclusion

In this paper we propose a methodology to study regularized estimation of high dimensional VARs with unobserved common factors. The presence of common factors introduces strong cross sectional dependence into the process. Incorporating such dependence is particularly important in high dimensional disaggregated data where connectedness between the variables may arise through different channels. This dependence and connectedness seem to be especially relevant in studying the transmission of investor fear across financial assets in our study of asset price volatility.

In practical work our procedure can be implemented in three steps as follows. First, given the order p of the VAR process, which can be estimated via a growth ratio criterion, we can obtain preliminary estimates of the transition matrices and common component via ℓ_1-nuclear norm regularizations, with which one can estimate the number of factors consistently and obtain a preliminary consistent estimate of the common factors. Second, the model is estimated using a generalized LASSO procedure by including the preliminary estimate of the common factors as regressors. Third, conservative LASSO is then used to obtain the final estimates, which are shown to be asymptotically equivalent to the oracle least squares estimates

The methods and results in this paper open up multiple avenues for further research. First, following Barigozzi and Brownlees (2019) it may be useful in practice to impose some sparsity assumptions on the large dimensional error variance matrix and develop estimation methods to achieve this. Second, frequency domain methods can be used to estimate the common factor components. Third, the model studied here does not allow for structural change in the transition matrices or the factor loadings (c.f., Su and Wang, 2017). It will also be interesting and challenging to study high dimensional VAR models with common factors that may involve time-varying transition matrices and factor loadings, which can help to capture empirically evolution in institutional and regulatory frameworks.

APPENDIX

A Proofs of the main results

Proof of Proposition 2.1: (i) By Assumption A.1(iv), $y_{it}^{(u)}$'s and $y_{it}^{(f)}$'s are mutually independent. It suffices to study them separately. By Assumption A.1(i), we can write $y_{it}^{(u)}$ as a linear process:

$$y_{it}^{(u)} = \sum_{j=0}^{\infty} \alpha_{iN}^{(u)}(j) u_{t-j} = \sum_{j=0}^{\infty} \alpha_{iN}^{(u)}(j) C^{(u)} \epsilon_{t-j}^{(u)} \equiv \sum_{j=0}^{\infty} C_j^{(i,u)} \epsilon_{t-j}^{(u)},$$

where $C_j^{(i,u)} \equiv \alpha_{iN}^{(u)}(j) C^{(u)}$. Under Assumption A.1(vi), one can bound $|(e_{1,p} \otimes e_{i,N})' \Phi^j|$ by $\psi_{\max}([\Phi^j]_{[N],[N]})$ $\leq \bar{c}\rho^j$. It follows that $|\alpha_{iN}^{(u)}(j)| \leq \bar{c}\rho^j$. Then the MA($\infty$) representation of $y_{it}^{(u)}$ is valid with $E(y_{it}^{(u)}) = 0$ and $\mathrm{Var}(y_{it}^{(u)}) = \sum_{j=0}^{\infty} \alpha_{iN}^{(u)}(j) \Sigma_u \alpha_{iN}^{(u)}(j)' < \infty$.

Under Assumption A.1(vi), we can also show that $E(|y_{it}^{(f)}|) \leq \sum_{j=0}^{\infty} |\alpha_{iN}^{(f)}(j)| |\mu_f| < \infty$. The MA($\infty$) representation of $y_{it}^{(f)}$ is

$$y_{it}^{(f)} = E(y_{it}^{(f)}) + \sum_{j=0}^{\infty} \alpha_{iN}^{(f)}(j)(f_{t-j}^0 - \mu_f) = E(y_{it}) + \sum_{j=0}^{\infty} C_j^{(i,f)} \epsilon_{t-j}^{(f)},$$

where $C_j^{(i,f)} \equiv \sum_{k=0}^{j} \alpha_{iN}^{(f)}(k) C_{j-k}^{(f)}$. Under Assumption A.1(vi), $|C_j^{(i,f)}| \leq \sum_{k=0}^{j} |\alpha_{iN}^{(f)}(k)| \cdot ||C_{j-k}^{(f)}||_{\mathrm{op}}$. In addition, by Assumption A.1(ii),

$$\sum_{j=0}^{\infty} \sum_{k=0}^{j} \rho^k ||C_{j-k}^{(f)}||_{\max} = \sum_{k=0}^{\infty} \rho^k \sum_{j=k}^{\infty} ||C_{j-k}^{(f)}||_{\max} \leq \bar{c} \sum_{k=0}^{\infty} \rho^k (k+1)^{-\alpha},$$

for some constant $\bar{c} < \infty$. Hence $C_j^{(i,f)}$ is absolutely summable, $\mathrm{Var}(y_{it}^{(f)}) = \sum_{j=0}^{\infty} C_j^{(i,f)} C_j^{(i,f)\prime} < \infty$, and the MA($\infty$) representation of $y_{it}^{(f)}$ is valid.

Similar to the decomposition (2.5), we can write $X_t = X_t^{(u)} + X_t^{(f)}$. For Σ_X, due to the independence between $X_t^{(u)}$ and $X_t^{(f)}$, we can also write it as $\Sigma_X = \Sigma_X^{(f)} + \Sigma_X^{(u)}$, where $\Sigma_X^{(u)} \equiv E(X_t^{(u)} X_t^{(u)\prime})$ and $\Sigma_X^{(f)} \equiv E(X_t^{(f)} X_t^{(f)\prime})$. By the fact that $\Sigma_X^{(f)}$ is positive semi-definite, we have $\psi_{\min}(\Sigma_X) \geq \psi_{\min}(\Sigma_X^{(u)})$. It suffices to show $\psi(\Sigma_X^{(u)})$ is bounded below. By Proposition 2.3 of BM (2015), we have

$$\psi_{\min}(\Sigma_X^{(u)}) \geq \frac{\psi_{\min}(\Sigma_u)}{\max_{|z|=1} \psi_{\max}(\mathcal{A}^*(z)\mathcal{A}(z))}.$$

Given Assumption A.1(vii), we have that $\psi_{\min}(\Sigma_X^{(u)})$ is bounded below by some constant.

(ii) By the independence between $X_t^{(u)}$ and $X_t^{(f)}$, one can also show that $\psi_{\min}(\Sigma) \geq \psi_{\min}(\Sigma_X^{(u)})$. ∎

A.1 Theoretical analysis of the first-step estimators

To prove Theorem 3.1, we need two lemmas whose proofs are in the online supplement.

Lemma A.1 *For the $T \times N$ matrices Θ^0 and Δ, we have*

(i) $\left\|\Theta^0 + \mathcal{M}(\Delta)\right\|_* = \left\|\Theta^0\right\|_* + \left\|\mathcal{M}(\Delta)\right\|_*$;

(ii) $\|\Delta\|_F^2 = \|\mathcal{M}(\Delta)\|_F^2 + \|\mathcal{P}(\Delta)\|_F^2$;

(iii) $rank(\mathcal{P}(\Delta)) \leq 2R^0$;

(iv) $\|\Delta\|_F^2 = \sum_j \psi_j(\Delta)^2$ and $\|\Delta\|_*^2 \leq \|\Delta\|_F^2 rank(\Delta)$;

For any conformable matrices M_1 and M_2, the following statement holds:

(v) $|tr(M_1 M_2)| \leq \|M_1\|_{max} |vec(M_2)|_1$ and $|tr(M_1 M_2)| \leq \|M_1\|_{op} \|M_2\|_*$.

Lemma A.2 *Suppose that Assumption A.1 holds. There exist absolute constants c, $\underline{c}$, $\bar{c} \in (0, \infty)$ such that*

(i) $\|\mathbf{U}'\mathbf{X}\|_{max}/T \leq \gamma_1/2$ with probability greater than $1 - \bar{c}(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$;

(ii) $\|\mathbf{U}'\mathbb{P}_{F^0}\mathbf{X}\|_{max}/T \leq c \cdot \gamma_1$ with probability greater than $1 - \bar{c}(NT^{1-q/4}(\log N)^{-q/2} + N^{1-\underline{c}\log N})$.

Proof of Theorem 3.1. Let $\tilde{\Delta}^{(1)} = \tilde{B} - B^0$ and $\tilde{\Delta}^{(2)} = \tilde{\Theta} - \Theta^0$. Define the event

$$\mathcal{E}_{NT}^{(1)} = \{\|\mathbf{U}'\mathbf{X}\|_{\max}/T \leq \gamma_1/2, \|\mathbf{U}\|_{op}/\sqrt{NT} \leq \gamma_2/2\}.$$

By Lemma A.2(i) and Assumption A.3(i), $\mathcal{E}_{NT}^{(1)}$ holds with probability at least $1 - \bar{c}[N^2 T^{1-q/2}(\log N)^{-q/2} + N^{2-\underline{c}\log N}]$. By the definition of $(\tilde{B}, \tilde{\Theta})$, we have that

$$\begin{aligned}
0 &\geq \mathcal{L}(\tilde{B}, \tilde{\Theta}) - \mathcal{L}(B^0, \Theta^0) \\
&= \frac{1}{2NT}(\|\mathbf{Y} - \mathbf{X}\tilde{B} - \tilde{\Theta}\|_F^2 - \|\mathbf{U}\|_F^2) + \frac{\gamma_1}{N}(|vec(\tilde{B})|_1 - |vec(B^0)|_1) + \frac{\gamma_2}{\sqrt{NT}}(\|\tilde{\Theta}\|_* - \|\Theta^0\|_*) \\
&\equiv d_1 + d_2 + d_3.
\end{aligned} \tag{A.1}$$

To establish the asymptotic properties of $\tilde{B}$ and $\tilde{\Theta}$, we study the three terms d_1, d_2 and d_3 in order.

First, we consider d_1. By the identity $\mathbf{Y} = \mathbf{X}B^0 + \Theta^0 + \mathbf{U}$, we have

$$\left\|\mathbf{Y} - \mathbf{X}\tilde{B} - \tilde{\Theta}\right\|_F^2 - \|\mathbf{U}\|_F^2 = \left\|\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}\right\|_F^2 - 2tr[\mathbf{U}'(\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)})].$$

For $tr[\mathbf{U}'(\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)})]$, conditional on $\mathcal{E}_{NT}^{(1)}$, we apply the triangle inequality and Lemma A.1(v) to obtain

$$\begin{aligned}
\frac{1}{NT}|tr[\mathbf{U}'(\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)})]| &\leq \frac{1}{NT}\|\mathbf{U}'\mathbf{X}\|_{max}|vec(\tilde{\Delta}^{(1)})|_1 + \frac{1}{NT}\|\mathbf{U}\|_{op}\|\tilde{\Delta}^{(2)}\|_* \\
&\leq \frac{\gamma_1}{2N}|vec(\tilde{\Delta}^{(1)})|_1 + \frac{\gamma_2}{2\sqrt{NT}}\|\tilde{\Delta}^{(2)}\|_*.
\end{aligned}$$

It follows that

$$
\begin{aligned}
d_1 &\geq \frac{1}{2NT}||\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}||_\mathrm{F}^2 - \frac{\gamma_1}{2N}|\mathrm{vec}(\tilde{\Delta}^{(1)})|_1 - \frac{\gamma_2}{2\sqrt{NT}}||\tilde{\Delta}^{(2)}||_* \\
&\geq \frac{1}{2NT}||\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}||_\mathrm{F}^2 - \frac{\gamma_1}{2N}\sum_{i=1}^{N}\left(|\tilde{\Delta}_{J_i,i}^{(1)}|_1 + |\tilde{\Delta}_{J_i^c,i}^{(1)}|_1\right) \\
&\quad - \frac{\gamma_2}{2\sqrt{NT}}\left(||\mathcal{P}(\tilde{\Delta}^{(2)})||_* + ||\mathcal{M}(\tilde{\Delta}^{(2)})||_*\right).
\end{aligned}
\tag{A.2}
$$

Next, we consider d_2. By the identities that $|\tilde{B}_{*,i}|_1 = |\tilde{B}_{J_i,i}|_1 + |\tilde{B}_{J_i^c,i}|_1$ and $|B_{*,i}^0|_1 = |B_{J_i,i}^0|_1$, we have

$$
d_2 = \frac{\gamma_1}{N}\sum_{i=1}^{N}(|\tilde{B}_{J_i,i}|_1 + |\tilde{B}_{J_i^c,i}|_1 - |B_{J_i,i}^0|_1) \geq \frac{\gamma_1}{N}\sum_{i=1}^{N}(|\tilde{\Delta}_{J_i^c,i}^{(1)}|_1 - |\tilde{\Delta}_{J_i,i}^{(1)}|_1),
\tag{A.3}
$$

where we use the fact that $|\tilde{B}_{J_i,i}|_1 + |\tilde{\Delta}_{J_i,i}^{(1)}|_1 \geq |B_{J_i,i}^0|_1$ by the triangle inequality and that $|\tilde{B}_{J_i^c,i}|_1 = |\tilde{\Delta}_{J_i^c,i}^{(1)}|_1$ as $B_{J_i^c,i}^0 = 0$.

Now, we consider d_3. By the triangle inequality and Lemma A.1(i), we have

$$
\begin{aligned}
||\tilde{\Theta}||_* &= ||\tilde{\Delta}^{(2)} + \Theta^0||_* = ||\Theta^0 + \mathcal{P}(\tilde{\Delta}^{(2)}) + \mathcal{M}(\tilde{\Delta}^{(2)})||_* \\
&\geq ||\Theta^0 + \mathcal{M}(\tilde{\Delta}^{(2)})||_* - ||\mathcal{P}(\tilde{\Delta}^{(2)})||_* \\
&= ||\Theta^0||_* + ||\mathcal{M}(\tilde{\Delta}^{(2)})||_* - ||\mathcal{P}(\tilde{\Delta}^{(2)})||_*.
\end{aligned}
$$

It follows that

$$
d_3 \geq \frac{\gamma_2}{\sqrt{NT}}(||\mathcal{M}(\tilde{\Delta}^{(2)})||_* - ||\mathcal{P}(\tilde{\Delta}^{(2)})||_*).
\tag{A.4}
$$

Combining the results in (A.1)-(A.4), we have

$$
\begin{aligned}
&\frac{1}{2NT}||\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}||_\mathrm{F}^2 + \frac{\gamma_1}{2N}\sum_{i=1}^{N}||\tilde{\Delta}_{J_i^c,i}^{(1)}||_1 + \frac{\gamma_2}{2\sqrt{NT}}||\mathcal{M}(\tilde{\Delta}^{(2)})||_* \\
&\leq \frac{3\gamma_1}{2N}\sum_{i=1}^{N}||\tilde{\Delta}_{J_i,i}^{(1)}||_1 + \frac{3\gamma_2}{2\sqrt{NT}}||\mathcal{P}(\tilde{\Delta}^{(2)})||_*.
\end{aligned}
\tag{A.5}
$$

The above inequality indicates that $(\tilde{\Delta}^{(1)}, \tilde{\Delta}^{(2)}) \in \mathcal{C}_{NT}(3)$. By Assumption A.2, we obtain that

$$
\frac{1}{N}||\tilde{\Delta}^{(1)}||_\mathrm{F}^2 + \frac{1}{NT}||\tilde{\Delta}^{(2)}||_\mathrm{F}^2 \leq \bar{\kappa}_3 \frac{1}{NT}||\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}||_\mathrm{F}^2,
\tag{A.6}
$$

where $\bar{\kappa}_3 = (\kappa_3 \wedge \kappa_3')^{-1}$. By the inequality (A.5), we have

$$
\begin{aligned}
\frac{1}{NT}||\mathbf{X}\tilde{\Delta}^{(1)} + \tilde{\Delta}^{(2)}||_{\mathrm{F}}^2 &\leq \frac{3\gamma_1}{N}\sum_{i=1}^{N}|\tilde{\Delta}_{J_i,i}^{(1)}|_1 + \frac{3\gamma_2}{\sqrt{NT}}||\mathcal{P}(\tilde{\Delta}^{(2)})||_* \\
&\leq 3\gamma_1\sqrt{K_a}\frac{||\tilde{\Delta}^{(1)}||_{\mathrm{F}}}{\sqrt{N}} + 3\sqrt{2R^0}\gamma_2\frac{||\tilde{\Delta}^{(2)}||_{\mathrm{F}}}{\sqrt{NT}} \\
&\leq 3\sqrt{2}(\gamma_1\sqrt{K_a} \vee (\sqrt{2R^0}\gamma_2))\sqrt{\frac{1}{N}||\tilde{\Delta}^{(1)}||_{\mathrm{F}}^2 + \frac{1}{NT}||\tilde{\Delta}^{(2)}||_{\mathrm{F}}^2}, \quad \text{(A.7)}
\end{aligned}
$$

where the second inequality holds by Lemma A.1(ii)-(iv) and the fact that $\sum_{i=1}^{N}|\tilde{\Delta}_{J_i,i}^{(1)}|_1 \leq \sqrt{NK_a}(\sum_{i=1}^{N}|\tilde{\Delta}_{J_i,i}^{(1)}|^2)^{1/2} \leq \sqrt{NK_a}\left\|\tilde{\Delta}^{(1)}\right\|_{\mathrm{F}}$, where recall that $K_a = N^{-1}\sum_{i=1}^{N}k_i$ and $k_i \equiv |J_i|$ denotes the cardinality of the set J_i. Combining (A.6)-(A.7) yields

$$
\frac{1}{N}||\tilde{\Delta}^{(1)}||_{\mathrm{F}}^2 + \frac{1}{NT}||\tilde{\Delta}^{(2)}||_{\mathrm{F}}^2 \leq 3\sqrt{2}\bar{\kappa}_3[(\gamma_1\sqrt{K_a}) \vee (\sqrt{2R^0}\gamma_2)]\sqrt{\frac{1}{N}||\tilde{\Delta}^{(1)}||_{\mathrm{F}}^2 + \frac{1}{NT}||\tilde{\Delta}^{(2)}||_{\mathrm{F}}^2},
$$

which implies that $\frac{1}{\sqrt{N}}||\tilde{\Delta}^{(1)}||_{\mathrm{F}} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2)$ and $\frac{1}{\sqrt{NT}}||\tilde{\Delta}^{(2)}||_{\mathrm{F}} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2)$ with $\bar{c} = 3\sqrt{2}\bar{\kappa}_3(1 \vee \sqrt{2R^0}) < \infty$. This completes the proof. ∎

To prove Theorem 3.2, we need the following lemma whose proof is in the online supplement.

Lemma A.3 *Suppose that Assumptions A.1 and A.3 holds. Let $S_F \equiv F^{0\prime}F^0/T$. Then for any $x > 0$,*

$$
P(T^{1/2}||S_F - \Sigma_F||_{max} > x) \leq C_1 x^{-q/2}T^{1-q/4} + C_2\exp\left(-C_3 x^2\right)
$$

for some absolute constants C_ℓ, $\ell = 1, 2, 3$.

Proof of Theorem 3.2. We operate conditional on the event that

$$
\mathcal{E}_{NT}^{(2)} = \{\left\|\mathbf{U}'\mathbf{X}\right\|_{\max}/T \leq \gamma_1/2, \left\|\mathbf{U}\right\|_{\mathrm{op}}/\sqrt{NT} \leq \gamma_2/2 \text{ and } ||S_F - \Sigma_F||_{\mathrm{op}} \leq c\sqrt{\log N}T^{-1/2}\},
$$

where c is a large positive constant. One can verify that for some positive constants $\bar{c}'$ and $\underline{c}$,

$$
P(\mathcal{E}_{NT}^{(2)}) \geq 1 - \bar{c}'(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})
$$

by Lemmas A.2-A.3. With Theorem 3.1, we have with probability at least $1 - \bar{c}'(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$,

$$
(NT)^{-1/2}||\tilde{\Theta} - \Theta^0||_{\mathrm{op}} \leq (NT)^{-1/2}||\tilde{\Theta} - \Theta^0||_{\mathrm{F}} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2).
$$

Next, we show that $\mathcal{E}_{NT}^{(2)}$ implies the desired results.

Step 1: Bound the eigenvalues.

Let $S_\Lambda = \Lambda^{0\prime}\Lambda^0/N$ and $S_F = F^{0\prime}F^0/T$. Let $\hat{s}_1 \geq \cdots \geq \hat{s}_{R^0}$ be the R^0 nonzero eigenvalues of $\frac{1}{NT}\Theta^0\Theta^{0\prime} = \frac{1}{T}F^{0\prime}S_\Lambda F^0$. Note that $\hat{s}_1, ..., \hat{s}_{R^0}$ are the same as the eigenvalues of $S_F^{1/2}S_\Lambda S_F^{1/2}$. Conditional on the event $\mathcal{E}_{NT}^{(2)}$ and by Assumption A.4(i)-(ii), we have

$$|\hat{s}_j - s_j| \leq \bar{c}(\sqrt{\log N}T^{-1/2} + N^{-1/2}) \text{ for some } \bar{c} < \infty \text{ and } j \in [R^0].$$

This also implies that $||\Theta^0||_{\mathrm{op}} = \sqrt{(s_1 + o_P(1))NT}$. For $j > R^0$, simply define $\hat{s}_j = s_j = 0$.

Let $\tilde{s}_1 \geq \cdots \geq \tilde{s}_{N\wedge T}$ be the eigenvalues of $\frac{1}{NT}\tilde{\Theta}\tilde{\Theta}'$. Again by the Weyl's theorem, we have for $j = 1, 2, ...$

$$
\begin{aligned}
|\tilde{s}_j - s_j| &\leq |\tilde{s}_j - \hat{s}_j| + |\hat{s}_j - s_j| \\
&\leq \frac{1}{NT}||\tilde{\Theta}\tilde{\Theta}' - \Theta^0\Theta^{0\prime}||_{\mathrm{op}} + |\hat{s}_j - s_j| \\
&\leq \frac{2}{NT}||\Theta^0||_{\mathrm{op}}||\tilde{\Theta} - \Theta^0||_{\mathrm{op}} + \frac{1}{NT}||\tilde{\Theta} - \Theta^0||_{\mathrm{op}}^2 + |\hat{s}_j - s_j|,
\end{aligned}
$$

implying $|\tilde{s}_j - s_j| \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2)$ for $j = 1, 2, ...$ Then for $j \in [R^0]$, w.p.a.1,

$$
\begin{aligned}
|\hat{s}_{j-1} - \tilde{s}_j| &\geq |\hat{s}_{j-1} - \hat{s}_j| - |\hat{s}_j - \tilde{s}_j| \geq (s_{j-1} - s_j)/2 \text{ and} \\
|\tilde{s}_j - \hat{s}_{j+1}| &\geq |\hat{s}_j - \hat{s}_{j+1}| - |\tilde{s}_j - \hat{s}_j| \geq (s_j - s_{j+1})/2,
\end{aligned}
\tag{A.8}
$$

with $\hat{s}_{R^0+1} = s_{R^0+1} = 0$.

Step 2: Prove the consistency of $\hat{R}$.

Note that $\psi_r(\tilde{\Theta}) = \sqrt{NT\tilde{s}_r}$. By the result in Step 1, we have that $\psi_r(\tilde{\Theta}) \geq \sqrt{[s_{R^0} - o_P(1)]NT}$ for all $r \leq R^0$, and

$$\psi_{R^0+1}(\tilde{\Theta}) \leq \psi_{R^0+1}(\Theta^0) + \left\|\tilde{\Theta} - \Theta^0\right\|_{\mathrm{op}} \leq \left\|\tilde{\Theta} - \Theta^0\right\|_{\mathrm{F}} \leq \sqrt{NT}\bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2) = \sqrt{NT}o(\gamma_2^{1/2})$$

where we use the condition that $\gamma_1\sqrt{K_a} = o(\gamma_2^{1/2})$ under Assumption A.3(ii). These results, in conjunction with the fact that $(\gamma_2\sqrt{NT}||\tilde{\Theta}||_{\mathrm{op}})^{1/2} \asymp \sqrt{NT}\sqrt{\gamma_2}$ with $\gamma_2 = c_2(N^{-1/2} + T^{-1/2})$,[3] implies that

$$\min_{r \leq R^0} \psi_r(\tilde{\Theta}) \geq (\gamma_2\sqrt{NT}||\tilde{\Theta}||_{\mathrm{op}})^{1/2} \text{ and } \psi_{R^0+1}(\tilde{\Theta}) < (\gamma_2\sqrt{NT}||\tilde{\Theta}||_{\mathrm{op}})^{1/2}$$

with probability at least $1 - \bar{c}'(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$ for sufficiently large (N, T). Then we have $\hat{R} = R^0$ with probability at least $1 - \bar{c}'(N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$ for sufficiently large (N, T).

---

[3]Write $a \asymp b$ to denote that both a/b and b/a are stochastically bounded.

Step 3: Characterize the eigenvectors.

Next, we show that there is an $R^0 \times R^0$ matrix $\tilde{H}$, so that the columns of $\frac{1}{\sqrt{T}}F^0\tilde{H}$ are the first R^0 eigenvectors of $\Theta^0\Theta^{0\prime}$. Let v be the $R^0 \times R^0$ matrix whose columns are the eigenvectors of $S_F^{1/2}S_\Lambda S_F^{1/2}$. Then $D = v'S_F^{1/2}S_\Lambda S_F^{1/2}v$ is a diagonal matrix of the eigenvalues of $S_F^{1/2}S_\Lambda S_F^{1/2}$ that are distinct by Assumption A.4(ii). Let $\tilde{H} = S_F^{-1/2}v$. Then

$$
\begin{aligned}
\frac{1}{NT}\Theta^0\Theta^{0\prime}F^0\tilde{H} &= \frac{1}{T}F^0S_\Lambda F^{0\prime}F^0\tilde{H} = F^0S_\Lambda S_F\tilde{H} = F^0S_\Lambda S_F^{1/2}v \\
&= F^0S_F^{1/2}S_F^{-1/2}S_\Lambda S_F^{1/2}v = F^0S_F^{1/2}vv'S_F^{-1/2}S_\Lambda S_F^{1/2}v \\
&= F^0\tilde{H}D.
\end{aligned}
$$

In addition, we have $(F^0\tilde{H})'F^0\tilde{H}/T = v'S_F^{-1/2}\frac{F^{0\prime}F^0}{T}S_F^{-1/2}v = v'v = I_{R^0}$. So the columns of $\frac{1}{\sqrt{T}}F^0\tilde{H}$ are the eigenvectors of $\Theta^0\Theta^{0\prime}$, with corresponding eigenvalues in D.

Step 4: Prove the convergence.

We bound $\left\| \tilde{F} - F^0\tilde{H} \right\|_{\mathrm{F}}$ conditional on the event $\hat{R} = R^0$. By the Davis-Kahan $\sin(\Theta)$ theorem (see, e.g., Yu et al., 2014) and (A.8),

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\tilde{F} - F^0\tilde{H}\|_{\mathrm{F}} &\leq \frac{\frac{1}{NT}\|\tilde{\Theta}\tilde{\Theta}' - \Theta^0\Theta^{0\prime}\|_{\mathrm{op}}}{\min_{j\leq R^0}\min\{|\hat{s}_{j-1} - \tilde{s}_j|, |\tilde{s}_j - \hat{s}_{j+1}|\}} \\
&\leq \bar{c}\frac{1}{NT}\|\tilde{\Theta}\tilde{\Theta}' - \Theta^0\Theta^{0\prime}\|_{\mathrm{op}} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2).
\end{aligned}
$$

Next we have

$$
\begin{aligned}
\left\|\mathbb{P}_{\tilde{F}} - \mathbb{P}_{F^0}\right\|_{\mathrm{F}} &= \left\|\frac{\tilde{F}\tilde{F}'}{T} - \mathbb{P}_{F^0}\right\|_{\mathrm{F}} \leq 2\bar{c}\left\|\frac{1}{\sqrt{T}}\tilde{F} - \frac{1}{\sqrt{T}}F^0\tilde{H}\right\|_{\mathrm{F}} + \left\|\frac{F^0\tilde{H}\tilde{H}'F^{0\prime}}{T} - \mathbb{P}_{F^0}\right\|_{\mathrm{F}} \\
&\leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2),
\end{aligned}
$$

where the second equality is by the fact $\tilde{H}\tilde{H}' = S_F^{-1/2}vv'S_F^{-1/2} = S_F^{-1}$. This proves the second result in the theorem. ∎

A.2 Theoretical analysis of the second-step estimators

To prove Theorem 3.3, we need to add a lemma.

Lemma A.4 *Suppose that Assumptions A.1-A.3 hold. Let $\tilde{\Sigma} \equiv T^{-1}\mathbf{X}'\mathbf{X} - T^{-2}\mathbf{X}'\tilde{F}\tilde{F}'\mathbf{X}$. Then there exist some constants $\underline{c}$, $\bar{c}$ and $\bar{c}'$ such that with probability larger than $1 - \bar{c}'(N^2T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$ we have*

(i) $\|\tilde{H}\|_{max} \leq \|\tilde{H}\|_\infty \leq \bar{c}$ and $\|\tilde{H}^{-1}\|_F \leq \bar{c}$;

(ii) $\max_{1\leq j\leq pN} |\mathbf{X}_{*,j}|/\sqrt{T} < \bar{c}$ and $\max_{1\leq j\leq N} |\mathbf{U}_{*,j}|/\sqrt{T} < \bar{c}$;

(iii) $\|F^{0\prime}\mathbf{U}\|_{max}/T \leq T^{-1/2}\log[N/(8\bar{c}^2)]$ and $\left\|T^{-1}\mathbf{X}'F^0 - \Sigma_{XF}\right\|_{max} \leq \bar{c}T^{-1/2}\log N$;

30

(iv) $||\tilde{\Sigma} - \Sigma||_{max} \le \gamma_3$;

(v) Suppose $16K_J\gamma_3 \le \psi_{\min}(\Sigma)/2$. Then $\tilde{\Sigma}$ satisfies the restricted eigenvalue condition for K_J in (3.3) and $\kappa_{\tilde{\Sigma}}(K_J) \ge \psi_{\min}(\Sigma)/2$.

Proof of Theorem 3.3. Fix $\bar{c}$ as in Lemma A.4. In this proof, we choose a large enough constant c_3 such that $\gamma_3 = c_3(\gamma_1\sqrt{K_a} \vee \gamma_2)$ with $c_3 \ge 2 \vee (16\bar{c}^2) \vee (16\bar{c}^4)$. Let $\mathcal{E}_{NT}^{(3)}$ be the joint event of

(1) $T^{-1}\|\mathbf{U}'\mathbf{X}\|_{\max} \le \gamma_3/4$;   (2) $\max_{1\le j\le pN}|\mathbf{X}_{*,j}|/\sqrt{T} \le \bar{c}$;

(3) $\max_{1\le j\le N}|\mathbf{U}_{*,j}|/\sqrt{T} \le \bar{c}$;   (4) $\|\tilde{F} - F^0\tilde{H}\|_{\mathrm{F}}/\sqrt{T} \le \gamma_3/(16\bar{c}^3)$;

(5) $\|F^{0\prime}\mathbf{U}\|_{\max}/T \le \gamma_3/(16\bar{c}^2)$;   (6) $\|\tilde{H}\|_\infty \vee \|\tilde{H}^{-1}\|_{\mathrm{F}} \le \bar{c}$;

(7) $\hat{R} = R^0$;

and (8) $\tilde{\Sigma}$ satisfies the restricted eigenvalue condition for K_J in (3.3) with $\kappa_{\tilde{\Sigma}}(K_J) \ge \psi_{\min}(\Sigma)/2$. Under Assumptions A.1-A.3, by Lemmas A.2 and A.4, $\mathcal{E}_{NT}^{(3)}$ holds with probability larger than $1 - \bar{c}'(N^2T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c}\log N})$. Conditional on the event $\mathcal{E}_{NT}^{(3)}$, we also have that

$$
\begin{aligned}
(9) \quad T^{-1}\|\tilde{F}'\mathbf{U}\|_{\max} &\le T^{-1}\|(\tilde{F} - F^0\tilde{H})'\mathbf{U}\|_{\max} + T^{-1}\|\tilde{H}'F^{0\prime}\mathbf{U}\|_{\max} \\
&\le T^{-1}\|\tilde{F} - F^0\tilde{H}\|_{\mathrm{F}} \cdot \max_{1\le j,N}\|\mathbf{U}_{*,j}\| + \|\tilde{H}'\|_\infty T^{-1}\|F^{0\prime}\mathbf{U}\|_{\max} \\
&\le \gamma_3/(8\bar{c}),
\end{aligned}
$$

and

$$
\begin{aligned}
(10) \quad \max_{1\le i\le N} T^{-1/2}|\lambda_i^{0\prime}F^{0\prime}\mathbb{M}_{\tilde{F}}| &\le \max_{1\le i\le N}|\lambda_i^0| \cdot T^{-1/2}\|(F^0 - \tilde{F}\tilde{H}^{-1})'\mathbb{M}_{\tilde{F}}\|_{\mathrm{F}} \\
&\le \bar{c}T^{-1/2}\|\tilde{F} - F^0\tilde{H}\|_{\mathrm{F}}\left\|\tilde{H}^{-1}\right\|_{\mathrm{F}} \le \gamma_3/(8\bar{c}).
\end{aligned}
$$

Conditional on the event $\mathcal{E}_{NT}^{(3)}$, we establish the bound of $|\dot{\Delta}_{*,i}|_1 \equiv |\dot{B}_{*,i} - B_{*,i}^0|_1$ for $i \in [N]$.

Step 1. Concentrate out λ.

The objective function (3.2) is a least squares objective function with respect to λ. Given $\dot{B}_{*,i}$, we have that

$$
\dot{\lambda}_i = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'(\mathbf{Y}_{*,i} - \mathbf{X}\dot{B}_{*,i}) = T^{-1}\tilde{F}'(\mathbf{Y}_{*,i} - \mathbf{X}\dot{B}_{*,i}),
$$

where the second equality holds by the identity $\tilde{F}'\tilde{F}/T = I_T$. After concentrating out λ_i, the optimization problem becomes

$$
\dot{B}_{*,i} = \operatorname*{argmin}_{v\in\mathbb{R}^{Np}} \frac{1}{2T}\|\mathbb{M}_{\tilde{F}}(\mathbf{Y}_{*,i} - \mathbf{X}v)\|_{\mathrm{F}}^2 + \gamma_3|v|_1, \tag{A.9}
$$

where $\mathbb{M}_{\tilde{F}} = I_T - \tilde{F}\tilde{F}'/T$.

Step 2. Compare the objective functions at $\dot{B}_{*,i}$ and $B_{*,i}^0$.

By the identity $\mathbf{Y}_{*,i} = \mathbf{X}B_{*,i}^0 + F^0\lambda_i^0 + \mathbf{U}_{*,i}$ and the definition of $\dot{B}_{*,i}$, we have

$$
\begin{aligned}
0 \;\geq\; & \frac{1}{2T}[||\mathbb{M}_{\tilde{F}}(\mathbf{Y}_{*,i} - \mathbf{X}\dot{B}_{*,i})||_{\mathrm{F}}^2 - ||\mathbb{M}_{\tilde{F}}(F^0\lambda_i^0 + \mathbf{U}_{*,i})||_{\mathrm{F}}^2] + \gamma_3(|\dot{B}_{*,i}|_1 - |B_{*,i}^0|_1) \\
=\; & \frac{1}{2T}||\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}||_{\mathrm{F}}^2 - \frac{1}{T}\mathrm{tr}[(F^0\lambda_i^0 + \mathbf{U}_{*,i})'\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}] + \gamma_3(|\dot{B}_{*,i}|_1 - |B_{*,i}^0|_1),
\end{aligned}
$$

where $\dot{\Delta} \equiv \dot{B} - B^0$ and $\dot{\Delta}_{*,i}$ denotes the ith column of $\dot{\Delta}$. Then by Lemma A.1(v), we have

$$
\begin{aligned}
\frac{1}{T}||(F^0\lambda_i^0 + \mathbf{U}_{*,i})'\mathbb{M}_{\tilde{F}}\mathbf{X}||_{\max}|\dot{\Delta}_{*,i}|_1 \;\geq\; & \frac{1}{T}\mathrm{tr}[(F^0\lambda_i^0 + \mathbf{U}_{*,i})'\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}] \\
\geq\; & \frac{1}{2T}||\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}||_F^2 + \gamma_3(|\dot{B}_{*,i}|_1 - |B_{*,i}^0|_1) \\
\geq\; & \frac{1}{2T}||\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}||_F^2 + \gamma_3|\dot{\Delta}_{J_i^c,i}|_1 - \gamma_3|\dot{\Delta}_{J_i,i}|_1.
\end{aligned}
$$

where the last inequality follows because

$$
\begin{aligned}
|\dot{B}_{*,i}|_1 - |B_{*,i}^0|_1 \;=\; & |\dot{\Delta}_{*,i} + B_{*,i}^0|_1 - |B_{*,i}^0|_1 = |\dot{\Delta}_{J_i^c,i}^0|_1 + |\dot{\Delta}_{J_i,i} + B_{*,i}^0|_1 - |B_{*,i}^0|_1 \\
\geq\; & |\dot{\Delta}_{J_i^c,i}|_1 - |\dot{\Delta}_{J_i,i}|_1.
\end{aligned}
$$

Step 3. Bound $T^{-1}\max_i[||(F^0\lambda_i^0 + U_{*,i})'\mathbb{M}_{\tilde{F}}X||_{\max}]$, conditional on the event $\mathcal{E}_{NT}^{(3)}$.

By the triangle and Cauchy Schwartz inequalities and the fact that $T^{-1/2}||\tilde{F}||_{\mathrm{op}} = 1$, we have

$$
\begin{aligned}
& T^{-1}||(F^0\lambda_i^0 + \mathbf{U}_{*,i})'\mathbb{M}_{\tilde{F}}\mathbf{X}||_{\max} \\
\leq\; & T^{-1}||\lambda_i^{0\prime}F^{0\prime}\mathbb{M}_{\tilde{F}}\mathbf{X}||_{\max} + T^{-1}||\mathbf{U}_{*,i}'\mathbb{M}_{\tilde{F}}\mathbf{X}||_{\max} \\
\leq\; & \max_{1\leq j\leq Np} T^{-1/2}|\mathbf{X}_{*,j}| \cdot T^{-1/2}|\lambda_i^{0\prime}F^{0\prime}\mathbb{M}_{\tilde{F}}| + \max_{1\leq j\leq Np} T^{-1}|\mathbf{U}_{*,i}'\mathbf{X}_{*,j}| + T^{-2}||\mathbf{U}_{*,i}'\tilde{F}\tilde{F}'\mathbf{X}||_{\max} \\
\leq\; & \max_{1\leq j\leq Np} T^{-1}|\mathbf{U}_{*,i}'\mathbf{X}_{*,j}| + \left\{ T^{-1}|\mathbf{U}_{*,i}'\tilde{F}| + T^{-1/2}|\lambda_i^{0\prime}F^{0\prime}\mathbb{M}_{\tilde{F}}| \right\} \max_{1\leq j\leq Np} T^{-1/2}|\mathbf{X}_{*,j}|.
\end{aligned}
$$

Combining events (1), (9) and (10), the right hand side of the above inequality is bounded by $\gamma_3/2$ conditional on the event $\mathcal{E}_{NT}^{(3)}$.

Step 4. Obtain the final bound for $|\dot{B}_{*,i} - B_{*,i}^0|_1$.

Combining the results in Steps 2-3 and using the identity $|\dot{\Delta}_{*,i}|_1 = |\dot{\Delta}_{J_i,i}|_1 + |\dot{\Delta}_{J_i^c,i}|_1$, we have that conditional on the event $\mathcal{E}_{NT}^{(3)}$,

$$
3\gamma_3|\dot{\Delta}_{J_i,i}|_1 \geq \frac{1}{T}||\mathbb{M}_{\tilde{F}}\mathbf{X}\dot{\Delta}_{*,i}||_{\mathrm{F}}^2 + \gamma_3|\dot{\Delta}_{J_i^c,i}|_1
$$

It follows that $|\dot{\Delta}_{J_i^c,i}|_1 \leq 3|\dot{\Delta}_{J_i,i}|_1$ and conditional on $\mathcal{E}_{NT}^{(3)}$,

$$
\dot{\Delta}_{*,i}'\tilde{\Sigma}\dot{\Delta}_{*,i} \leq 3\gamma_3|\dot{\Delta}_{J_i,i}|_1 \leq 3\gamma_3\sqrt{k_i}|\dot{\Delta}_{J_i,i}| \leq \frac{6\sqrt{k_i}}{\psi_{\min}(\Sigma)}\gamma_3\sqrt{\dot{\Delta}_{*,i}'\tilde{\Sigma}\dot{\Delta}_{*,i}},
$$

where the last inequality holds by event (8) in $\mathcal{E}_{NT}^{(3)}$. It follows that $\sqrt{\dot{\Delta}'_{*,i}\tilde{\Sigma}\dot{\Delta}_{*,i}} \leq \frac{6\sqrt{k_i}}{\psi_{\min}(\Sigma)}\gamma_3$ and $|\dot{\Delta}_{J_i,i}|_1 \leq \frac{2\sqrt{k_i}}{\psi_{\min}(\Sigma)}\sqrt{\dot{\Delta}_{*,i}\tilde{\Sigma}\dot{\Delta}_{*,i}} \leq \frac{12k_i}{(\psi_{\min}(\Sigma))^2}\gamma_3$. Consequently, we have established that

$$|\dot{\Delta}_{*,i}|_1 = |\dot{\Delta}_{J_i,i}|_1 + |\dot{\Delta}_{J_i^c,i}|_1 \leq 4|\dot{\Delta}_{J_i,i}|_1 \leq \frac{48}{(\psi_{\min}(\Sigma))^2}k_i\gamma_3.$$

Then the conclusion in Theorem 3.3 follows. ∎

A.3 Theoretical analysis of the third-step estimators

To prove Theorems 3.4 and 3.5, we need the following lemma.

Lemma A.5 *Suppose that Assumptions A.1-A.5 hold. Then*
  (i) For $i = 1, ..., N$, $\psi_{\min}(\tilde{\Sigma}_{J_i,J_i}) \geq \underline{c}$ w.p.a.1 for some finite constant $\underline{c}$;
  (ii) $||\tilde{\Sigma}_{J_i^c,J_i}||_{max} \leq \bar{c}$ w.p.a.1 for some finite constant $\bar{c}$.

Proof of Theorem 3.4: For any n-dimensional vector $v = (v_1, ..., v_n)'$, denote

$$\text{abs}(v) = (|v_1|, ..., |v_n|)'.$$

We say that $v < \tilde{v}$ if and only if $v_i < \tilde{v}'_i$ for all $i \in [n]$. Let $W^{(i)} = \text{diag}(w_{1i}, ..., w_{Np,i})$, $W^{(1,i)} = W^{(i)}_{J_i,J_i}$ and $W^{(0,i)} = W^{(i)}_{J_i^c,J_i^c}$.

The following proof is done by induction. Based on the error bounds for $\hat{F}^{(\ell)}$'s, we show that results (i)-(iii) hold for the $(\ell+1)$th-step estimators. Then the results follows as we already have $||\hat{F}^{(0)} - F^0\tilde{H}||_F/\sqrt{T} = O_P(\gamma_1\sqrt{K_a} + \gamma_2)$.

For notational simplicity, let $\tilde{\Sigma}^{(\ell)}$ denote $T^{-1}\mathbf{X}'\mathbb{M}_{\hat{F}^{(\ell)}}\mathbf{X}$ for $\ell = 0, 1, 2, ...$

(i) For all (k,i)'s such that $B_{ki}^0 = 0$, $\sup_{(k,i):B_{ki}^0=0}|\dot{B}_{ki}| \leq ||\dot{B} - B^0||_{\max} \leq O_P(K_J\gamma_3) = o_P(\gamma_4)$. It follows that $W^{(0,i)} = I_{|J_i^c|}$ with probability approaching one (w.p.a.1). For all (k,i)'s such that $B_{ki}^0 \neq 0$,

$$\min_{k,i:B_{ki}^0\neq 0}|\dot{B}_{ki}| > \min_{i\in[N]}\min_{k\in J_i}|B_{ki}^0| - ||\dot{B} - B^0||_{\max} = \min_{i\in[N]}\min_{k\in J_i}|B_{ki}^0| - o_P(\gamma_4) \geq \alpha\gamma_4 \text{ w.p.a.1}$$

by Assumption A.5(i). It follows that $W^{(1,i)} = \mathbf{0}$ w.p.a.1. For each $i \in [N]$, the estimator $\hat{B}_{*,i}^{(\ell)}$ can be written as

$$\hat{B}_{*,i}^{(\ell)} = \text{argmin}_{v\in\mathbb{R}^{NP}}\mathcal{L}^{(i)}(v, \hat{F}^{(\ell-1)}),$$

where $\mathcal{L}^{(i)}(v, F) \equiv \frac{1}{2T}(\mathbf{Y}_{*,i} - \mathbf{X}v)'\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{Y}_{*,i} - \mathbf{X}v) + \gamma_4\sum_{k=1}^{pN}w_{ki}|v_k|$. Following the proof of Proposition 1 of Zhao and Yu (2006), $\text{sgn}(\hat{B}_{*,i}^{(l)}) = \text{sgn}(B_{*,i}^0)$ is implied by event $\mathcal{E}_{i,1} \cap \mathcal{E}_{i,2}$, where

$$\mathcal{E}_{i,1} \equiv \left\{\text{abs}[T^{-1/2}\tilde{\Sigma}_{J_i,J_i}^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{U}_{*,i} + F^0\lambda_i^0)] < T^{1/2}\text{abs}(B_{J_i,i}^0) - T^{1/2}\gamma_4\text{abs}[\tilde{\Sigma}_{J_i,J_i}^{-1}W^{(1,i)}\text{sgn}(B_{J_i,i}^0)]\right\}$$

33

and

$$\begin{aligned}
\mathcal{E}_{i,2} &\equiv \{\text{abs}[T^{-1/2}(-\tilde{\Sigma}_{J_i^c, J_i}\tilde{\Sigma}_{J_i, J_i}^{-1} \cdot \mathbf{X}'_{*, J_i} + \mathbf{X}'_{*, J_i^c})\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{U}_{*,i} + F^0 \lambda_i^0)] \\
&< T^{1/2}\gamma_4 W^{(0,i)} \cdot \iota_{|J_i^c|} - T^{1/2}\gamma_4 \text{abs}[\tilde{\Sigma}_{J_i^c, J_i}\tilde{\Sigma}_{J_i, J_i}^{-1} W^{(1,i)}\text{sgn}(B_{J_i, i}^0)]\}.
\end{aligned}$$

We prove (i) by showing that $\mathcal{E}_{i,1}$ and $\mathcal{E}_{i,2}$ hold w.p.a.1.

First, we consider $\mathcal{E}_{i,1}$. It suffices to show that each entry of $T^{-1/2}\text{abs}[\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{U}_{*,i} + F^0 \lambda_i^0)]$ is $o_P(\sqrt{T}\min_i \min_{k \in J_i} |B_{ki}^0|)$. Applying the triangle inequality, one has

$$\begin{aligned}
&T^{-1/2}\text{abs}[\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{U}_{*,i} + F^0 \lambda_i^0)] \\
&\leq T^{-1/2}\text{abs}(\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}\mathbf{U}_{*,i}) + T^{-1/2}\text{abs}(\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}F^0 \lambda_i^0) \\
&\leq T^{-1/2}\text{abs}(\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i}) + T^{-1/2}\text{abs}[\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}(\mathbb{P}_{F^0} - \mathbb{P}_{\hat{F}^{(\ell-1)}})\mathbf{U}_{*,i}] \\
&\quad + T^{-1/2}\text{abs}[\tilde{\Sigma}_{J_i, J_i}^{-1}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}(\hat{F}^{(\ell-1)} - F^0 \tilde{H})\tilde{H}^{-1}\lambda_i^0]. \quad\quad\quad (A.10)
\end{aligned}$$

Note that $\max_i ||\tilde{\Sigma}_{J_i, J_i}^{-1}||_{\text{op}} \leq \bar{c}$ w.p.a.1 by Lemma A.5(i). This, in conjunction with Lemma A.2(i)-(ii), implies that the first term on the right hand side of (RHS) of (A.10) is uniformly $O_P(\log N)$. With $||\hat{F}^{(\ell-1)} - F^0 \tilde{H}||_{\text{F}}/\sqrt{T} = O_P(\gamma_1\sqrt{K_a} + \gamma_2) = O_P((\log N)T^{-1/2}\sqrt{K_a} + N^{-1/2})$,[4] we have $||\mathbb{P}_{F^0} - \mathbb{P}_{\hat{F}^{(\ell-1)}}||_{\text{op}} = O_P((\log N)T^{-1/2}\sqrt{K_a} + N^{-1/2})$. Note that Lemma A.4(ii) ensures $\max_{1 \leq j \leq pN}||\mathbf{X}_{*,j}||/\sqrt{T}$ and $\max_{1 \leq j \leq N}||\mathbf{U}_{*,j}||/\sqrt{T}$ are both bounded by an absolute constant. It follows that each entry of the second term on the RHS of (A.10) is $O_P(\log N \cdot \sqrt{K_a} + \sqrt{T/N})$. Similarly, each entry of the third term on the RHS is $O_P(\log N \cdot \sqrt{K_a} + \sqrt{T/N})$. These results, along with the fact that $\log N \cdot T^{-1/2}\sqrt{K_a} = o(\min_i \min_{k \in J_i} |B_{ki}^0|)$ and $N^{-1/2} = o(\min_i \min_{k \in J_i} |B_{ki}^0|)$ in Assumption A.5 imply that $P(\mathcal{E}_{i,1}) \to 1$.

Next, we consider $\mathcal{E}_{i,2}$. Similar to the analysis for $\mathcal{E}_{i,1}$, we can use Lemma A.5(ii) to show that each entry of $T^{-1/2}(-\tilde{\Sigma}_{J_i^c, J_i}\tilde{\Sigma}_{J_i, J_i}^{-1} \cdot \mathbf{X}'_{*, J_i} + \mathbf{X}'_{*, J_i^c})\mathbb{M}_{\hat{F}^{(\ell-1)}}(\mathbf{U}_{*,i} + F^0 \lambda_i^0)$ is $O_P(\log N \cdot \sqrt{K_a} + \sqrt{T/N}) = o(\sqrt{T}\gamma_3)$. By the fact that $\gamma_3 = o(\gamma_4)$, we have $P(\mathcal{E}_{i,2}) \to 1$, as $(N, T) \to \infty$.

(ii) Conditional on the event $\{\hat{B}^{(\ell)} =_s B^0\}$, we can follow the proof of Lemma 1 in Zhao and Yu (2006) to establish the first order condition that

$$\tilde{\Sigma}_{J_i, J_i}(\hat{B}_{J_i, i}^{(\ell)} - B_{J_i, i}^0) = \frac{1}{T}\mathbf{X}'_{*, J_i}\mathbb{M}_{\hat{F}^{(\ell-1)}}(F^0 \lambda_i^0 + \mathbb{M}_{\hat{F}^{(\ell-1)}}\mathbf{U}_{*,i}),$$

---

[4]This claim holds for $\ell = 1$ by Theorem 3.2. Given this claim, we can show that $||\hat{F}^{(\ell)} - F^0 \tilde{H}||_{\text{F}}/\sqrt{T} = O_P((\log N)T^{-1/2}\sqrt{K_a} + N^{-1/2})$ for each ℓ using the results below.

for $i \in [N]$. Then

$$
\begin{aligned}
\left| \hat{B}_{J_i,i}^{(\ell)} - B_{J_i,i}^0 \right| &= \left| \tilde{\Sigma}_{J_i,J_i}^{-1} \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} (F^0 \lambda_i^0 + \mathbf{U}_{*,i}) \right| \\
&\leq \underline{c}^{-1} \left| \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} F^0 \lambda_i^0 \right| + \underline{c}^{-1} \left| \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} \mathbf{U}_{*,i} \right| \equiv \underline{c}^{-1} (A_{1i} + A_{2i}),
\end{aligned}
$$

where we use the fact that $\max_i \left\| \tilde{\Sigma}_{J_i,J_i}^{-1} \right\|_{op} \leq \underline{c}^{-1}$ w.p.a.1 by Lemma A.5(i). Note that uniform in $i \in [N]$,

$$
\begin{aligned}
A_{1i}^2 &= \frac{1}{T^2} \left| \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} (\hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0) \lambda_i^0 \right|^2 \\
&= \frac{1}{T^2} \mathrm{tr} \left(\lambda_i^{0\prime} (\hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0)' \mathbb{M}_{\hat{F}^{(\ell-1)}} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} (\hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0) \lambda_i^0 \right) \\
&\leq \psi_{\max} \left(\frac{1}{T} \mathbb{M}_{\hat{F}^{(\ell-1)}} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} \right) \frac{1}{T} \left\| \hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0 \right\|^2 \| \lambda_i^0 \|^2 \\
&= \psi_{\max} \left(\frac{1}{T} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} \mathbf{X}_{*,J_i} \right) \frac{1}{T} \left\| \hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0 \right\|^2 \| \lambda_i^0 \|^2 \\
&\leq \bar{c} \frac{1}{T} \left\| \hat{F}^{(\ell-1)} \tilde{H}^{-1} - F^0 \right\|^2 = O_P((\log N)^2 T^{-1} K_a + N^{-1})
\end{aligned}
$$

and

$$
A_{2i}^2 = \left| \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbb{M}_{\hat{F}^{(\ell-1)}} \mathbf{U}_{*,i} \right|^2 \leq 2 \left| \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbf{U}_{*,i} \right|^2 + 2 \left| \frac{1}{T} \mathbf{X}'_{*,J_i} \hat{F}^{(\ell-1)} \frac{1}{T} \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \right|^2.
$$

It is standard to show that $\left| \frac{1}{T} \mathbf{X}'_{*,J_i} \mathbf{U}_{*,i} \right| \leq k_i^{1/2} O_P(T^{-1/2} \log N)$ uniform in i. In addition,

$$
\begin{aligned}
\left| \frac{1}{T} \mathbf{X}'_{*,J_i} \hat{F}^{(\ell-1)} \frac{1}{T} \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \right|^2 &= \mathrm{tr} \left(\frac{1}{T^2} \hat{F}^{(\ell-1)\prime} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \hat{F}^{(\ell-1)} \frac{1}{T^2} \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \mathbf{U}'_{*,i} \hat{F}^{(\ell-1)} \right) \\
&\leq \psi_{\max} \left(\frac{1}{T^2} \hat{F}^{(\ell-1)\prime} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \hat{F}^{(\ell-1)} \right) \frac{1}{T^2} \left| \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \right|^2 \\
&\leq \psi_{\max} \left(\frac{1}{T} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \right) \frac{1}{T^2} \left| \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \right|^2 \\
&= O_P(\gamma_1 \sqrt{K_a} + \gamma_2)^2) \quad \text{uniformly in } i,
\end{aligned}
$$

where the last equality follows from the fact $\psi_{\max} \left(\frac{1}{T} \mathbf{X}_{*,J_i} \mathbf{X}'_{*,J_i} \right) \leq \bar{c}$ w.p.a.1 and $\max_i \frac{1}{T} \left| \hat{F}^{(\ell-1)\prime} \mathbf{U}_{*,i} \right| = O_P(\gamma_1 \sqrt{K_a} + \gamma_2)$ by similar arguments as used to obtain event (9) in the proof of Theorem 3.3. Then uniformly in $i \in [N]$, we have $A_{2i}^2 \leq k_i O_P(T^{-1/2} \log N) + O_P((\gamma_1 \sqrt{K_a} + \gamma_2)^2)$ and

$$
\begin{aligned}
\left| \hat{B}_{J_i,i}^{(\ell)} - B_{J_i,i}^0 \right|^2 &\leq O_P((\log N)^2 T^{-1} K_a + N^{-1}) + k_i O_P(T^{-1} (\log N)^2) + O_P((\gamma_1 \sqrt{K_a} + \gamma_2)^2) \\
&= k_i O_P(T^{-1} (\log N)^2) + O_P((\gamma_1 \sqrt{K_a} + \gamma_2)^2).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{\|\mathbf{X}(\hat{B}^{(\ell)} - B^0)\|_{\mathrm{F}}^2}{NT} &= \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\mathbf{X}(\hat{B}_{*,i}^{(\ell)} - B_{*,i}^0)\right|^2}{T} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}(\hat{B}_{J_i,i}^{(\ell)} - B_{J_i,i}^0)'\mathbf{X}_{*,J_i}'\mathbf{X}_{*,J_i}(\hat{B}_{J_i,i}^{(\ell)} - B_{J_i,i}^0)' \\
&= \max_i\left\|\frac{1}{T}\mathbf{X}_{*,J_i}'\mathbf{X}_{*,J_i}\right\|_{\mathrm{op}}\frac{1}{N}\sum_{i=1}^{N}\left|\hat{B}_{J_i,i}^{(\ell)} - B_{J_i,i}^0\right|^2 \\
&= O_P(1)\left[K_a O_P(T^{-1}(\log N)^2) + O_P(\gamma_1\sqrt{K_a} + \gamma_2)^2)\right] = O_P((\gamma_1\sqrt{K_a} + \gamma_2)^2).
\end{aligned}
$$

Then the result in (ii) follows.

(iii) Note that $\mathbf{Y} - \mathbf{X}\hat{B}^{(\ell)} - F^0\Lambda^{0\prime} = \mathbf{U} - \mathbf{X}(\hat{B}^{(\ell)} - B^0)$. By the result in (ii) and Assumption A.3(i), the operator norm of

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\left\|\mathbf{U} - \mathbf{X}(\hat{B}^{(\ell)} - B^0)\right\|_{\mathrm{op}} &\leq \frac{1}{\sqrt{NT}}\|\mathbf{U}\|_{\mathrm{op}} + \frac{1}{\sqrt{NT}}\left\|\mathbf{X}(\hat{B}^{(\ell)} - B^0)\right\|_{\mathrm{op}} \\
&\leq O_P(\gamma_2) + O_P(\gamma_1\sqrt{K_a} + \gamma_2) = O_P(\gamma_1\sqrt{K_a} + \gamma_2).
\end{aligned}
$$

One can apply analyses similar to proof of Theorem 3.2 to obtain the desired result. ∎

Proof of Theorem 3.5: Let $\hat{\Sigma} = \mathbf{X}'\mathbb{M}_{\hat{F}}\mathbf{X}/T$. From the proof of Theorem 3.4, we have that

$$
\hat{\Sigma}_{J_i,J_i}(\hat{B}_{J_i,i} - B_{J_i,i}^0) = \frac{1}{T}\mathbf{X}_{*,J_i}'\mathbb{M}_{\hat{F}}F^0\lambda_i^0 + \frac{1}{T}\mathbf{X}_{*,J_i}'\mathbb{M}_{\hat{F}}\mathbf{U}_{*,i} - \gamma_4 W^{(1,i)}\mathrm{sgn}(B_{J_i,i}^0). \tag{A.11}
$$

Noting that the columns of $\hat{F}/\sqrt{T}$ are the first $\hat{R}$ eigenvectors of $\frac{1}{NT}(\mathbf{Y} - \mathbf{X}\hat{B})(\mathbf{Y} - \mathbf{X}\hat{B})'$, we have

$$
\hat{F}V_{NT} = \frac{1}{NT}\left(\mathbf{Y} - \mathbf{X}\hat{B}\right)\left(\mathbf{Y} - \mathbf{X}\hat{B}\right)'\hat{F} = \frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\hat{B}_{J_i,i}\right)\left(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\hat{B}_{J_i,i}\right)'\hat{F},
$$

where V_{NT} is a diagonal matrix that consists of the $\hat{R}$ largest eigenvalues of the matrix $T \times T$ matrix $(NT)^{-1}(\mathbf{Y} - \mathbf{X}\hat{B})(\mathbf{Y} - \mathbf{X}\hat{B})'$, arranged in descending order along its diagonal line.

By Theorem 3.3 and Assumption A.5(i), $\max_{k \in J_i} w_{ki} = 0$ w.p.a.1, which implies that $\gamma_4 W^{(1,i)}\mathrm{sgn}(B_{J_i,i}^0) = o_p(T^{-1/2})$. Then we can follow the analysis of oracle least squares estimator to establish the asymptotic distribution of $\hat{B}_{J_i,i}$. Specifically, by arguments as used in the proof of Proposition B.1 in the online supplement, we have

$$
S_i(\hat{B}_{J_i,i} - B_{J_i,i}^0) = S_i(\frac{1}{T}\mathbf{X}_{*,J_i}'\mathbb{M}_{F^0}\mathbf{X}_{*,J_i})^{-1}\frac{1}{T}\mathbf{X}_{*,J_i}'\mathbb{M}_{F^0}\mathbf{U}_{*,i} + o_P(T^{-1/2}).
$$

By arguments as used in the proof of Lemma A.2, we can readily show that $\left\|\frac{1}{T}\mathbf{X}_{*,J_i}'\mathbb{M}_{F^0}\mathbf{X}_{*,J_i} - \Sigma_{J_i,J_i}\right\|_{\mathrm{F}}$ $= O_P\left(K_J T^{-1/2}\log N\right)$ and $|\frac{1}{\sqrt{T}}\mathbf{X}_{*,J_i}'\mathbb{P}_{F^0}\mathbf{U}_{*,i} - \frac{1}{\sqrt{T}}(F^0\Sigma_F^{-1}[\Sigma_{FX}]_{J_i,*})'\mathbf{U}_{*,i}| = O_P(K_J^{1/2}T^{-1/2}\log N)$,

where $[\Sigma_{FX}]_{J_i,*} = \frac{1}{T}E\left[F^{0\prime}\mathbf{X}_{*,J_i}\right]$ is a $R^0 \times k_i$ matrix. It follows that

$$\sqrt{T}S_i(\hat{B}_{J_i,i} - B^0_{J_i,i}) = \frac{1}{\sqrt{T}}S_i(\Sigma_{J_i,J_i})^{-1}(\mathbf{X}_{*,J_i} - F^0\Sigma_F^{-1}[\Sigma_{FX}]_{J_i,*})'\mathbf{U}_{*,i} + o_P(1)$$

$$\equiv T^{-1/2}\sum_{t=1}^{T}z^*_{it}u_{it} + o_P(1),$$

where $z^*_{it} = S_i(\Sigma_{J_i,J_i})^{-1}z^0_{it}$ and z^0_{it} denotes the tth column of the $k_i \times T$ matrix $(\mathbf{X}_{*,J_i} - F^0\Sigma_F^{-1}[\Sigma_{FX}]_{J_i,*})'$. Under Assumption A.1, $\{z^*_{it}u_{it}, t \geq 1\}$ is a martingale difference sequence (m.d.s.) and we can readily verifying the conditions of the martingale central limit theorem by straightforward moment calculations and obtain $\sqrt{T}S_i(\check{B}_{J_i,i} - B^0_{J_i,i}) \xrightarrow{d} N(0, \sigma_i^2 S_i(\Sigma_{J_i,J_i})^{-1}S_i')$, where $\sigma_i^2 = E\left(u_{it}^2\right)$. ∎

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica, 81: 1203-1227.

Alizadeh, S., Brandt, M.W. and Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047-1091.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135-171.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica, 77, 1229-1279.

Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, Z. and Saranadasa, H. 1996. Effect of high dimension: by an example of a two sample problem. Statistica Sinica 6, 311–329.

Barigozzi, M. and Brownlees, C., 2019. Nets: Network estimation for time series. Journal of Applied Econometrics 34, 347-364.

Barigozzi, M. and Hallin, M., 2017. A network analysis of the volatility of high dimensional financial series. Journal of the Royal Statistical Society: Series C 66, 581-605.

Basu, S. and Michailidis, G., 2015. Regularized estimation in sparse high-dimensional time series models. The Annals of Statistics 43, 1535-1567.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80, 2369-2429.

Belloni, A., Chen, M., and Padilla, O. H. M., 2019. High dimensional latent panel quantile regression with an application to asset pricing. arXiv preprint arXiv:1912.02151 .

Bernanke, B.S., Boivin, J. and Eliasz, P., 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. Quarterly Journal of Economics 120, 387-422.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B., 2009. Simultaneous analysis of Lasso and Dantzig selector. Annals of Statistics, 37: 1705-1732.

Billio, M., Getmansky, M., Lo, A.W. and Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics 104, 535-559.

Candes, E. and Tao, T. 2007. The Dantzig selector: Statistical estimation when p is much larger than n. The annals of Statistics, 35: 2313-2351.

Caner, M. and Kock, A.B., 2018. Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative LASSO. Journal of Econometrics 203, 143-168.

Chen, S. and Qin, Y. 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Annals of Statistics, 38, 808-835.

Cheng, X. and Hansen, B.E., 2015. Forecasting with factor-augmented regression: A frequentist model averaging approach. Journal of Econometrics, 186, 280-293.

Chernozhukov, V., Hansen, C., Liao, Y. and Zhu, Y., 2018. Inference For Heterogeneous Effects Using Low-Rank Estimations. arXiv preprint arXiv:1812.08089.

Chudik, A. and Pesaran, M.H., 2011. Infinite-dimensional VARs and factor models. Journal of Econometrics 163, 4-22.

Chudik, A., Pesaran, M.H. and Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. Econometrics Journal 14, C45–C90.

Demirer, M., Diebold, F.X., Liu, L. and Yilmaz, K., 2018. Estimating global bank network connectedness. Journal of Applied Econometrics 33, 1-15.

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 Yilmaz, K., 2015. Financial and macroeconomic connectedness: a network approach to measurement and monitoring. Oxford University Press.

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.

Fan, J. and Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. Annals of Statistics 32, 928-961.

Fan, J., Gong, W., and Z. Zhu Z., 2019. Generalized high-dimensional trace regression via nuclear norm regularization. Journal of Econometrics 212, 177-202.

Feng, J., 2019. Regularized quantile regression with interactive fixed effects. arXiv preprint arXiv:1911.00166.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. Review of Economics and statistics 82, 540-554.

Garman, M.B. and Klass, M.J., 1980. On the estimation of security price volatilities from historical data. Journal of Business 53, 67-78.

Geweke, J., 1977. The dynamic factor analysis of economic time series. Latent Variables in Socio-Economic Models.

Giannone, D., Reichlin, L. and Sala, L., 2004. Monetary policy in real time. NBER Macroeconomics Annual 19, 161-200.

Guo, S., Wang, Y. and Yao, Q., 2016. High-dimensional and banded vector autoregressions. Biometrika 103, 889-903.

Hallin, M. and Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603-617.

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.

Haufe, S., Müller, K.R., Nolte, G. and Krämer, N., 2010, February. Sparse causal discovery in multivariate time series. In I. Guyon, D. Janzing, and B. Schölkopf (eds), Proceedings of Workshop on Causality: Objectives and Assessment, pp.97-106.

Hautsch, N., Schaumburg, J. and Schienle, M., 2014. Financial network systemic risk contributions. Review of Finance 19, 685-738.

Huang, J., Ma, S. and Zhang, C.H., 2008. Adaptive LASSO for sparse high-dimensional regression models. Statistica Sinica 1603-1618.

Hurn, S. Martin, V. L., Phillips, P. C. B., and Yu, J. 2020. Financial Econometric Modelling. Oxford University Press.

Kock, A.B. and Callot, L., 2015. Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics 186, 325-344.

Koltchinskii, V., Lounici, K., and Tsybakov, A. B., 2011. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. The Annals of Statistics 39, 2302-2329.

Lam, C. and Fan, J., 2008. Profile-kernel likelihood inference with diverging number of parameters. Annals of Statistics 36, 2232-2260.

Leeper, E.M., Sims, C.A., Zha, T., Hall, R.E. and Bernanke, B.S., 1996. What does monetary policy do? Brookings Papers on Economic Activity 1996, 1-78.

Lu, X. and Su, L., 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. Journal of Econometrics 190, 148-175.

Ludvigson, S., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. Journal of Financial Economics 83,171-222.

Lükepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer Science & Business Media, New York.

Ma, S., Lan, W., Su, L., and Tsai, C-L., 2019. Testing alphas in conditional time-varying factor models with high dimensional assets. Journal of Business & Economic Statistics 38, 214-227.

Ma, S., Su, L., and Zhang, Y., 2020. Detecting latent communities in network formation model. Working Paper, Singapore Management University.

Mann, H. B. and Wald, A. 1943. On the statistical treatment of linear stochastic difference equations. Econometrica, Journal of the Econometric Society, 173-220.

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á., and Zilberman, E., 2019. Forecasting Inflation in a data-rich environment: the benefits of machine learning methods. Journal of Business & Economic Statistics, 1-22.

Moon, H.R. and Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33, 158–195.

Moon, H.R. and Weidner, M., 2019. Nuclear norm regularized estimation of panel regression models. arXiv:1810.10987.

Negahban, S. and Wainwright, M. J. , 2011. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. The Annals of Statistics 39, 1069-1097.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B., 2012. A unified framework for high dimensional analysis of m-estimators with decomposable regularizers. Statistical Science 27, 538-557.

Onatski, A. 2010. Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics, 92: 1004-1016.

Qian, J. and Su, L., 2016. Shrinkage estimation of regression models with multiple structural changes. Econometric Theory 32, 1376-1433.

Rohde, A. and Tsybakov, A. B., 2011. Estimation of high-dimensional low-rank matrices. The Annals of Statistics 39, 887-930.

Sims, C. A., 1980. Macroeconomics and reality. Econometrica, 48: 1-48.

Sims, C. A., 1992. Interpreting the macroeconomic time series facts: The effects of monetary policy. European Economic Review 36, 975-1000.

Sims, C. A., 1993. A nine-variable probabilistic macroeconomic forecasting model. In Business cycles, indicators and forecasting pp. 179-212. University of Chicago Press.

Smeekes, S. and Wijler, E., 2018. Macroeconomic forecasting using penalized regression methods. International Journal of Forecasting 34, 408-430.

Stock, J.H. and Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293-335.

Stock, J.H. and Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147-162.

Stock, J.H. and Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis, No. w11467. National Bureau of Economic Research.

Su, L., and Wang, X., 2017. On time-varying factor models: Estimation and testing. Journal of Econometrics 198, 84-101.

Tibshirani, R. 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B, 58: 267-288.

Uematsu, Y. and Tanaka, S., 2019. High-dimensional macroeconomic forecasting and variable selection via penalized regression. Econometrics Journal 22, 34-56.

Vershynin, R., 2018. High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, Cambridge.

Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences 102, 14150-14154.

Wu, W.B. and Wu, Y.N., 2016. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. Electronic Journal of Statistics 10, 352-379.

Yu, Y., Wang, T. and Samworth, R.J., 2014. A useful variant of the Davis–Kahan theorem for statisticians. Biometrika 102, 315-323.

Zhao, P. and Yu, B., 2006. On model selection consistency of LASSO. Journal of Machine Learning Research 7, 2541-2563.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American statistical association, 101: 1418-1429.
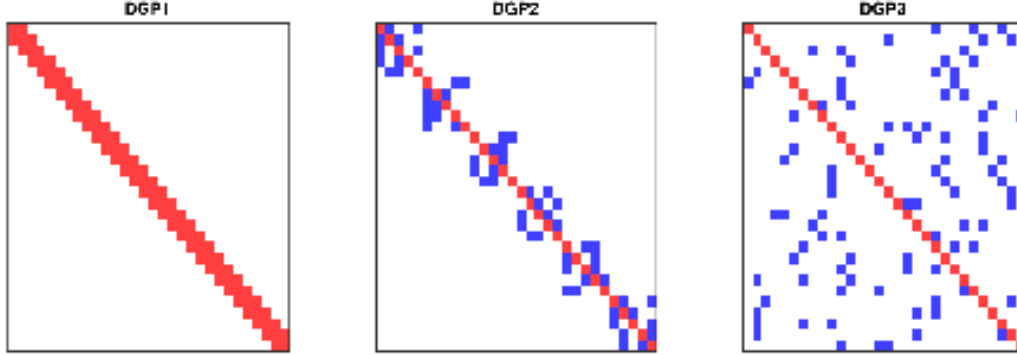
Figure 1: Structure of the transition matrices in the simulations

Table 1: Model Selection Accuracy

			Number of factors		Step 1		Step 2		Step 3	
DGP	N	T	UER	OER	TPR	FPR	TPR	FPR	TPR	FPR
1	30	100	0.0%	0.0%	97.4%	19.3%	98.8%	18.5%	93.7%	8.0%
	30	200	0.0%	0.0%	99.6%	19.1%	99.9%	18.1%	99.4%	5.8%
	30	400	0.0%	0.0%	99.9%	21.8%	100.0%	19.5%	99.9%	4.9%
	60	100	0.0%	0.0%	96.8%	12.7%	98.2%	12.2%	90.5%	5.1%
	60	200	0.0%	0.0%	99.9%	12.2%	100.0%	11.7%	99.1%	2.6%
	60	400	0.0%	0.0%	100.0%	11.9%	100.0%	11.1%	99.9%	1.7%
2	30	100	0.0%	0.0%	86.2%	21.8%	83.9%	18.9%	94.0%	15.7%
	30	200	0.0%	0.0%	95.3%	28.0%	93.7%	24.8%	99.4%	12.8%
	30	400	0.0%	0.0%	99.2%	37.0%	98.7%	33.3%	99.9%	8.2%
	60	100	0.0%	0.0%	76.7%	10.3%	76.5%	9.4%	90.6%	10.7%
	60	200	0.0%	0.0%	88.9%	12.5%	89.7%	12.0%	99.2%	8.9%
	60	400	0.0%	0.0%	96.4%	17.7%	95.8%	16.7%	100.0%	5.5%
3	30	100	0.0%	0.0%	93.2%	24.9%	92.3%	22.0%	96.5%	17.4%
	30	200	0.0%	0.0%	98.1%	31.4%	97.6%	27.6%	99.6%	11.7%
	30	400	0.0%	0.0%	99.5%	38.4%	99.3%	34.4%	99.7%	7.3%
	60	100	0.0%	0.0%	88.1%	12.8%	88.4%	11.8%	95.9%	11.8%
	60	200	0.0%	0.0%	96.1%	15.6%	95.5%	13.9%	99.8%	9.4%
	60	400	0.0%	0.0%	98.9%	19.5%	98.6%	17.9%	100.0%	4.5%

Note: We report the under/over-estimation rate (UER and OER) of the number of factors in the UER and OER columns, respectively. The TPR (true positive rate) columns report the average shares of relevant variables included. The FPR (false positive rate) columns report the average shares of irrelevant variables included.

Table 2: Root mean squared errors of the feasible and oracle transition matrix estimators

			All entries				Nonzero entries			
DGP	N	T	Oracle	Step 1	Step 2	Step 3	Oracle	Step 1	Step 2	Step 3
1	30	100	0.019	0.063	0.059	0.050	0.062	0.145	0.132	0.117
	30	200	0.014	0.055	0.051	0.033	0.044	0.118	0.106	0.066
	30	400	0.010	0.052	0.049	0.029	0.033	0.100	0.092	0.047
	60	100	0.013	0.044	0.041	0.038	0.061	0.150	0.138	0.131
	60	200	0.010	0.035	0.032	0.021	0.043	0.108	0.098	0.066
	60	400	0.007	0.033	0.031	0.016	0.032	0.089	0.080	0.041
2	30	100	0.018	0.065	0.065	0.057	0.056	0.177	0.184	0.154
	30	200	0.012	0.055	0.055	0.038	0.039	0.142	0.150	0.103
	30	400	0.009	0.047	0.047	0.027	0.028	0.110	0.119	0.070
	60	100	0.012	0.050	0.049	0.044	0.054	0.204	0.205	0.179
	60	200	0.008	0.042	0.041	0.028	0.038	0.170	0.168	0.114
	60	400	0.006	0.035	0.035	0.019	0.027	0.138	0.143	0.081
3	30	100	0.019	0.065	0.064	0.055	0.051	0.150	0.155	0.127
	30	200	0.013	0.053	0.053	0.035	0.035	0.117	0.123	0.082
	30	400	0.009	0.047	0.047	0.027	0.025	0.095	0.100	0.058
	60	100	0.013	0.050	0.049	0.042	0.049	0.173	0.173	0.146
	60	200	0.009	0.039	0.040	0.024	0.034	0.135	0.140	0.085
	60	400	0.006	0.033	0.033	0.015	0.024	0.109	0.113	0.056

Note: We report the root mean squared errors (RMSEs) of the feasible and oracle transition matrix estimators.
Columns 4-7 report the RMSEs of all entries, and Columns 8-11 report the RMSEs of non-zero entries.

Table 3: Results of misspecified estiamtes

			LASSO				Conservative LASSO			
DGP	N	T	TPR	FPR	RMSE_a	RMSE_b	TPR	FPR	RMSE_a	RMSE_b
1	30	100	78.7%	34.9%	0.115	0.208	78.4%	45.2%	0.178	0.227
	30	200	88.9%	37.7%	0.094	0.178	88.1%	43.3%	0.129	0.173
	30	400	95.3%	45.0%	0.083	0.150	94.5%	43.0%	0.103	0.134
	60	100	71.0%	22.6%	0.086	0.216	72.8%	39.5%	0.161	0.240
	60	200	86.7%	25.7%	0.070	0.179	87.0%	38.9%	0.114	0.175
	60	400	94.9%	30.2%	0.058	0.148	95.3%	37.9%	0.083	0.128
2	30	100	86.2%	59.6%	0.150	0.202	81.9%	54.8%	0.211	0.233
	30	200	95.0%	61.5%	0.107	0.152	91.7%	51.4%	0.139	0.159
	30	400	98.9%	66.3%	0.080	0.113	97.7%	50.5%	0.098	0.110
	60	100	77.0%	46.6%	0.135	0.218	74.1%	48.9%	0.222	0.263
	60	200	91.6%	51.9%	0.100	0.165	86.8%	44.6%	0.143	0.175
	60	400	98.3%	56.1%	0.072	0.120	96.7%	44.4%	0.097	0.116
3	30	100	89.2%	59.2%	0.139	0.186	85.7%	55.9%	0.196	0.215
	30	200	96.2%	61.4%	0.102	0.141	94.0%	54.3%	0.133	0.148
	30	400	99.1%	67.1%	0.079	0.107	98.3%	53.2%	0.096	0.106
	60	100	82.0%	46.1%	0.126	0.203	79.8%	50.6%	0.208	0.247
	60	200	94.0%	51.7%	0.093	0.151	90.5%	46.6%	0.135	0.164
	60	400	98.8%	55.5%	0.068	0.110	97.6%	45.0%	0.091	0.109

Note: We report the true positive rate (TPR), false positive rate (FPR), root mean squared errors of all entries
(RMSE_a) and nonzero entries (RMSE_b) of misspecified estimates. We consider the LASSO estimator as in Kock and
Callot (2015) and a conservative LASSO estimator. The LASSO estimator was used to construct weights for the
conservative LASSO.

Table 4: Funds information

category	ticker	fund name	category	ticker	fund name
Energy	XLE	Energy Select Sector SPDR Fund	Natu	XLB	Materials Select Sector SPDR Fund
	XOP	Spdr S&P Oil & Gas Explo & Prod Etf		XME	SPDR S&P Metals & Mining ETF
	IYE	iShares U.S. Energy ETF	Tech	XLK	Technology Select Sector SPDR Fund
	OIH	VanEck Vectors Oil Services ETF		SMH	VanEck Vectors Semiconductor ETF
Financial	XLF	Financial Select Sector SPDR Fund	Heal	XLV	Health Care Select Sector SPDR Fund
	KBE	SPDR S&P Bank ETF		IBB	iShares Nasdaq Biotechnology ETF
	KRE	SPDR S&P Regional Banking ETF	Def	XLP	Consumer Staples Select Sector SPDR Fund
Cyc	XLY	Cons. Disc. Select Sector SPDR Fund	Util	XLU	Utilities Select Sector SPDR Fund
	XHB	Spdr S&P Homebuilders Etf	Indu	XLI	Industrial Select Sector SPDR Fund
	ITB	iShares U.S. Home Construction ETF	EPM	GDX	VanEck Vectors Gold Miners ETF
	XRT	Spdr S&P Retail Etf			
Rea	IYR	iShares U.S. Real Estate ETF			
	VNQ	Vanguard Real Estate Index Fund ETF			

Note: Cyc, Rea, Natu, Tech, Heal, Def, Util, Indu and EMP stand for consumer cyclical, real estate, natural resource, technology, health care, consumer defensive, utilities, industrials and equity precious metals, respectively.

Table 5: Descriptive statistics

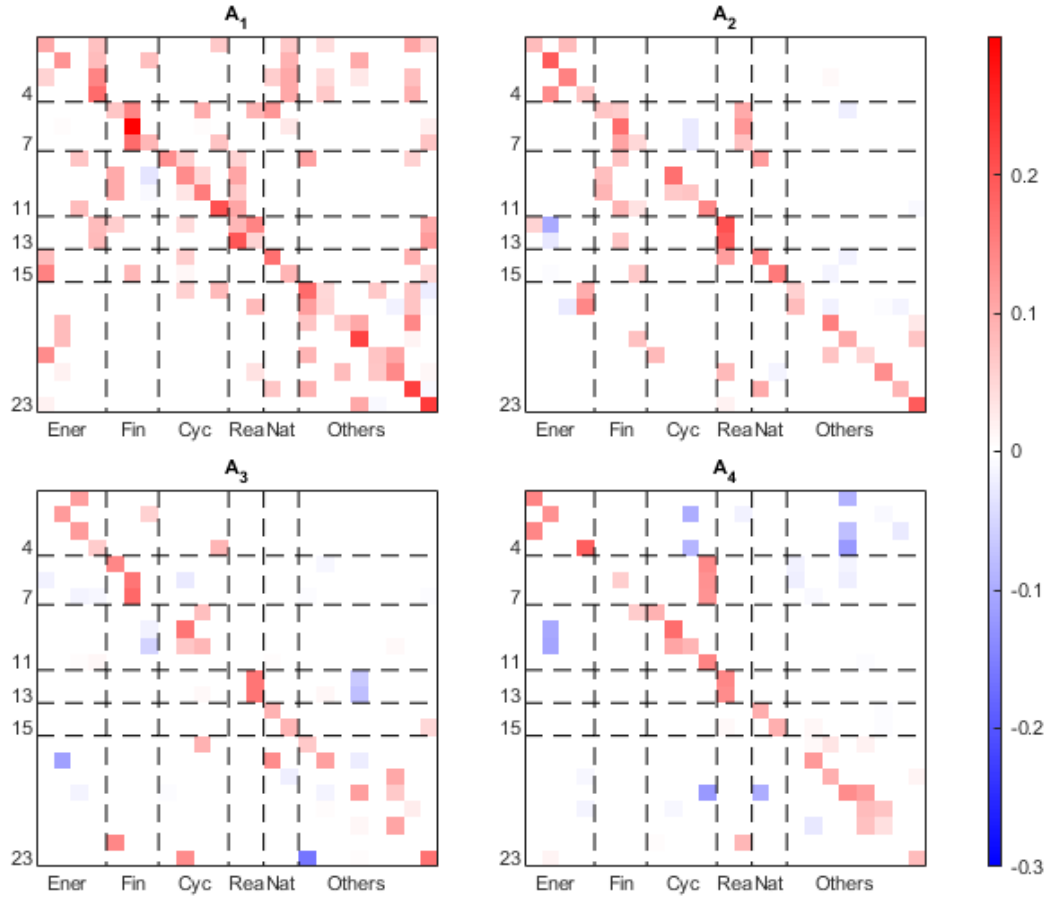| TICKER	XLE	XOP	IYE	OIH	XLF	KBE	KRE	XLY
mean	0.00136	0.00246	0.00141	0.00220	0.00157	0.00194	0.00184	0.00082
median	0.00063	0.00130	0.00059	0.00128	0.00041	0.00059	0.00066	0.00029
max	0.06034	0.06290	0.11527	0.05856	0.05743	0.04793	0.09748	0.03063
min	0.00004	0.00005	0.00004	0.00008	0.00001	0.00002	0.00002	0.00001
std	0.00369	0.00472	0.00549	0.00418	0.00463	0.00484	0.00539	0.00214
skewness	10.954	7.604	15.469	8.159	7.645	5.823	11.530	8.869
kurtosis	151.595	77.386	291.137	88.226	77.152	44.720	175.439	102.667
TICKER	XHB	ITB	XRT	IYR	VNQ	XLB	XME	XLK
mean	0.00218	0.00251	0.00115	0.00137	0.00146	0.00098	0.00264	0.00071
median	0.00079	0.00102	0.00056	0.00039	0.00041	0.00047	0.00133	0.00031
max	0.05071	0.04660	0.03094	0.04847	0.04831	0.02948	0.05631	0.03112
min	0.00007	0.00001	0.00001	0.00003	0.00004	0.00004	0.00014	0.00002
std	0.00431	0.00473	0.00231	0.00377	0.00403	0.00205	0.00510	0.00187
skewness	5.305	4.936	7.783	6.789	6.958	8.059	6.912	9.814
kurtosis	41.414	33.799	83.839	61.695	64.487	90.224	62.231	128.784
TICKER	SMH	XLV	IBB	XLP	XLU	XLI	GDX	
mean	0.00111	0.00054	0.00105	0.00036	0.00062	0.00075	0.00263	
median	0.00069	0.00025	0.00058	0.00016	0.00030	0.00036	0.00154	
max	0.02010	0.02865	0.03488	0.02197	0.03903	0.02108	0.07009	
min	0.00004	0.00002	0.00003	0.00001	0.00003	0.00001	0.00010	
std	0.00153	0.00162	0.00207	0.00111	0.00193	0.00156	0.00439	
skewness	5.713	11.898	9.968	13.670	14.053	7.405	8.300	
kurtosis	52.259	176.016	135.878	237.109	250.309	76.935	102.080	

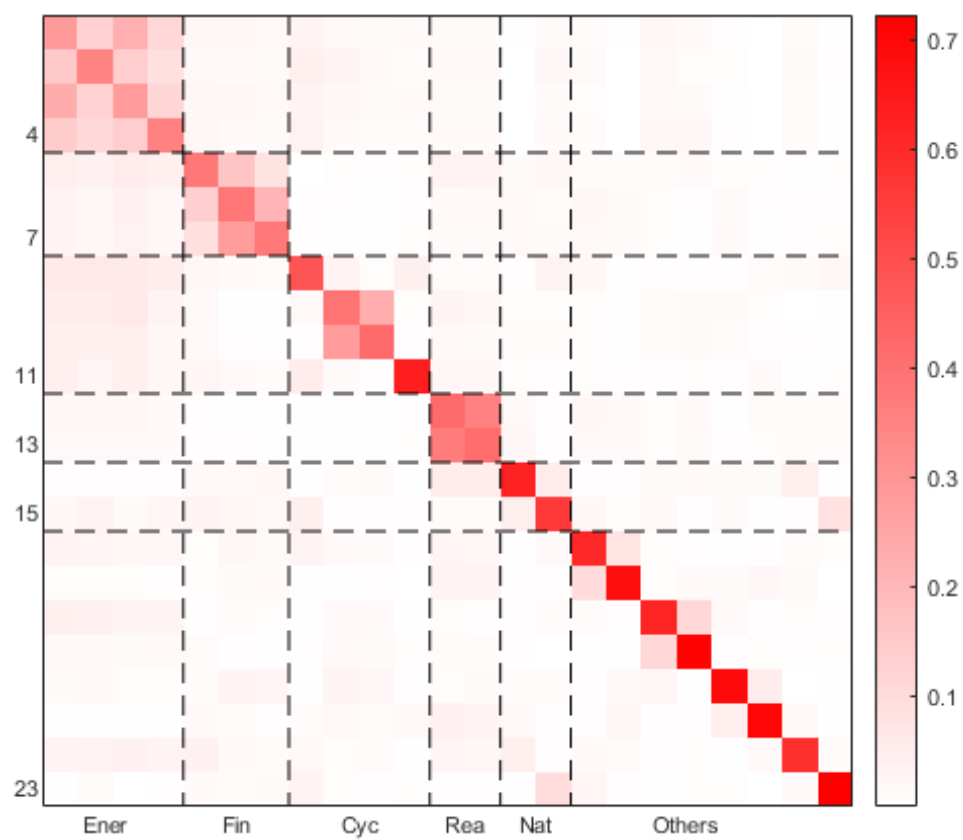Figure 2: Heat map of the transition matrices A_k's

Figure 3: Heat map of $\tilde{D}^{12}$

45

Table 6: Connectedness measures across funds

			Connectedness measures by estimates of VAR with CFs model					
TICKER	XLE	XOP	IYE	OIH	XLF	KBE	KRE	XLY
ν_i	64.9%	59.1%	65.4%	58.0%	65.2%	56.8%	56.6%	72.0%
FROM	71.4%	65.4%	71.7%	64.3%	61.7%	61.3%	62.3%	51.8%
TO_i	106.8%	86.0%	103.9%	71.5%	57.8%	72.6%	51.4%	37.3%
TICKER	XHB	ITB	XRT	IYR	VNQ	XLB	XME	XLK
ν_i	53.6%	49.5%	60.1%	50.7%	49.7%	67.2%	56.9%	70.5%
FROM_i	60.5%	58.3%	36.5%	57.9%	58.6%	37.5%	44.1%	39.0%
TO_i	56.3%	41.7%	19.0%	79.7%	74.4%	26.3%	37.2%	37.3%
TICKER	SMH	XLV	IBB	XLP	XLU	XLI	GDX	average
ν_i	54.8%	64.3%	50.7%	61.3%	50.6%	67.7%	31.0%	$\bar{\nu}$ =56.1%
FROM_i	31.9%	38.3%	28.8%	30.7%	29.7%	40.9%	27.7%	$\bar{d}^{12}$ =49.8%
TO_i	23.3%	34.1%	33.0%	21.2%	19.6%	20.7%	19.1%	
			Connectedness measures by estimates of pure VAR model					
TICKER	XLE	XOP	IYE	OIH	XLF	KBE	KRE	XLY
FROM_i	89.3%	87.1%	89.4%	87.0%	89.6%	86.8%	87.6%	90.9%
TO_i	105.0%	79.5%	103.0%	77.7%	112.9%	97.0%	89.1%	110.5%
TICKER	XHB	ITB	XRT	IYR	VNQ	XLB	XME	XLK
FROM_i	87.3%	86.3%	88.8%	85.7%	86.2%	90.1%	88.8%	89.8%
TO_i	95.8%	80.8%	79.1%	94.0%	89.6%	105.6%	80.1%	103.8%
TICKER	SMH	XLV	IBB	XLP	XLU	XLI	GDX	average
FROM_i	87.6%	88.1%	83.8%	88.4%	85.7%	89.8%	76.5%	$\bar{d}^{12}$ =87.40%
TO_i	74.8%	81.2%	60.8%	80.0%	60.0%	104.3%	45.8%	

Note: Cyc, Rea, Natu, Tech, Heal, Def, Util, Indu and EMP stand for consumer cyclical, real estate, natural resource, technology, health care, consumer defensive, utilities, industrials and equity precious metals, respectively.

Online Supplement for
"High dimensional VAR with common factors"

Ke Miao[a], Peter C.B. Phillips[b] and Liangjun Su[c]

[a] School of Economics, Fudan University

[b] Yale University, University of Auckland,

University of Southampton, & Singapore Management University

[c] School of Economics and Management, Tsinghua University

This supplement has three parts. Section B contains the asymptotic analysis of the oracle least squares estimator and the proofs of some technical lemmas. Section C provides some technical results that are used in the proofs. Section D provides some discussion of Assumption A.1(vi).

B   Supplementary Proof

B.1   Asymptotic analysis of the oracle least squares estimator

In this subsection, we study the asymptotic properties of the oracle least squares estimator that is obtained with information of J_i for $i \in [N] \equiv \{1, ..., N\}$. Specifically, the oracle least squares

estimator can be written as:

$$
\begin{aligned}
(\check{B}, \check{F}) & \equiv \underset{(B,F)\in\mathbb{B}^*\times\mathbb{L}}{\operatorname{argmin}} \operatorname{tr}\left[(\mathbf{Y} - \mathbf{X}B)'\mathbb{M}_F(\mathbf{Y} - \mathbf{X}B)\right] \\
& = \underset{(B,F)\in\mathbb{B}^*\times\mathbb{L}}{\operatorname{argmin}} \sum_{i=1}^{N}\left[(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}B_{J_i,i})'\mathbb{M}_F\left(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}B_{J_i,i}\right)\right],
\end{aligned}
$$

where $\mathbb{B}^* \equiv \{B \in \mathbb{R}^{Np\times N}|B_{J_i^c,i} = 0 \text{ for } i \in [N]\}$ and $\mathbb{L} \equiv \{F \in \mathbb{R}^{T\times R^0}|F'F/T = I_{R^0}\}$. For each i, we define a selector matrix L_i such that $\mathbf{X}L_i = \mathbf{X}_{*,J_i}$ and $L_i'B_{*,i} = B_{J_i,i}$. Recall that $k_i = |J_i|$ denotes the cardinality of J_i.

We do not have a closed-form solution to the above minimization problem. Similar to equations (11)-(12) of Bai (2009), we have the relationship:

$$
\begin{aligned}
\check{B}_{J_i,i} & = (\mathbf{X}_{*,J_i}'\mathbb{M}_{\check{F}}\mathbf{X}_{*,J_i})^{-1}\mathbf{X}_{*,J_i}'\mathbb{M}_{\check{F}}\mathbf{Y}_{*,i}, & \text{(B.1)} \\
\check{F}\check{V}_{NT} & = \frac{1}{NT}\left(\mathbf{Y} - \mathbf{X}\check{B}\right)\left(\mathbf{Y} - \mathbf{X}\check{B}\right)'\check{F} & \text{(B.2)} \\
& = \frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\check{B}_{J_i,i}\right)\left(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\check{B}_{J_i,i}\right)'\check{F},
\end{aligned}
$$

where $\check{V}_{NT}$ is a diagonal matrix that consists of the R^0 largest eigenvalues of the matrix $(NT)^{-1}$ $\times\sum_{i=1}^{N}(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\check{B}_{J_i,i})(\mathbf{Y}_{*,i} - \mathbf{X}_{*,J_i}\check{B}_{J_i,i})'$, arranged in descending order along its diagonal line. We can follow the lead of Bai (2009) to expand equations (B.1)-(B.2).

Proposition B.1 *Suppose Assumptions A.1 and A.3-A.5 hold. Let S_i denote an $L \times |J_i|$ selection matrix such that $\|S_i\|_F$ is finite and L is a fixed integer. Then*

$$
S_i\left(\check{B}_{J_i,i} - B_{J_i,i}^0\right) = S_i(\mathbf{X}_{*,J_i}'\mathbb{M}_{F^0}\mathbf{X}_{*,J_i})^{-1}\mathbf{X}_{*,J_i}'\mathbb{M}_{F^0}\mathbf{U}_{*,i} + o_P(T^{-1/2}).
$$

Proof of Proposition B.1: Insert the identity $\mathbf{Y} = \mathbf{X}B^0 + F^0\Lambda^{0\prime} + \mathbf{U}$ into equation (B.2), we have

$$
\check{F}\check{V}_{NT} - F^0\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}\check{F}}{T} = (I_1 + ... + I_8)\check{F}, \qquad \text{(B.3)}
$$

where

$$
\begin{array}{lll}
I_1 \equiv \mathbf{U}\Lambda^0 F^{0\prime}/(NT), & I_2 \equiv F^0\Lambda^{0\prime}\mathbf{U}'/(NT), & I_3 \equiv \mathbf{U}\mathbf{U}'/(NT), \\
I_4 \equiv \mathbf{X}(\check{B} - B^0)(\check{B} - B^0)'\mathbf{X}'/(NT), & I_5 \equiv -\mathbf{X}(\check{B} - B^0)\Lambda^0 F^{0\prime}/(NT), & I_6 \equiv -F^0\Lambda^{0\prime}(\check{B} - B^0)'\mathbf{X}'/(NT), \\
I_7 \equiv -\mathbf{X}(\check{B} - B^0)\mathbf{U}'/(NT), & I_8 \equiv -\mathbf{U}(\check{B} - B^0)'\mathbf{X}'/(NT).
\end{array}
$$

We can easily show that $(NT)^{-1}\|I_\ell\|_F = o_P(1)$ for $\ell = 1, ..., 8$. Premultiplying both sides of equation

(B.3) by $\check{F}'/T$, we can obtain that

$$\left\|\check{V}_{NT} - \frac{\check{F}'F^0}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}\check{F}}{T}\right\|_F = o_P(1).$$

Given that $\frac{F^{0\prime}\check{F}}{T}$ is asymptotically nonsingular, we have that V_{NT} is invertible asymptotically. Specifically, one can show that rth diagonal element of $\check{V}_{NT}$ converges to the rth singular value of $\Sigma_F\Sigma_\Lambda$. Hence we can write $(\check{F} - F^0\check{H})/\sqrt{T} = I_1^* + ... + I_8^*$, where $\check{H} \equiv (\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}\check{F}/T)\check{V}_{NT}^{-1}$ and $I_\ell^* \equiv I_\ell\check{F}\check{V}_{NT}^{-1}/\sqrt{T}$ for $\ell = 1, ..., 8$. One can also show that $\check{H} - \tilde{H} \to 0$. Noting that

$$\begin{aligned}
||\mathbf{X}(\check{B} - B^0)/\sqrt{NT}||_F^2 &= \frac{1}{N}\sum_{i=1}^N (\check{B}_{J_i,i} - B_{J_i,i}^0)'\frac{\mathbf{X}'_{*,J_i}\mathbf{X}_{*,J_i}}{T}(\check{B}_{J_i,i} - B_{J_i,i}^0) \\
&\leq \bar{c}\frac{1}{N}\sum_{i=1}^N |\check{B}_{J_i,i} - B_{J_i,i}^0|^2 = \bar{c}\frac{1}{N}||\check{B} - B^0||_F^2,
\end{aligned}$$

we have

$$||\mathbf{X}(\check{B} - B^0)/\sqrt{NT}||_F = O_P(d_{NT}) \tag{B.4}$$

where $d_{NT} \equiv N^{-1/2}||\check{B} - B^0||_F$.

Let $\check{Q}_1 = \text{bdiag}\{T^{-1}\mathbf{X}'_{*,J_1}\mathbb{M}_{F^0}\mathbf{X}_{*,J_1}, ..., T^{-1}\mathbf{X}'_{*,J_N}\mathbb{M}_{F^0}\mathbf{X}_{*,J_N}\}$ be a block diagonal matrix with the ith diagonal block given by $T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}$. Let $a_{ij}^0 \equiv \lambda_i^{0\prime}(\Lambda^{0\prime}\Lambda^0/N)^{-1}\lambda_j^0$. Let $\check{Q}_2$ be a $(\sum_{i=1}^N k_i) \times (\sum_{i=1}^N k_i)$ block partitioned matrix with the (i,j)th block given by $(NT)^{-1}a_{1ij}^0\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_j}$ for $i, j \in [N]$. That is,

$$\check{Q}_2 = \begin{bmatrix}
(NT)^{-1}a_{11}^0\mathbf{X}'_{*,J_1}\mathbb{M}_{F^0}\mathbf{X}_{*,J_1} & \cdots & (NT)^{-1}a_{1N}^0\mathbf{X}'_{*,J_1}\mathbb{M}_{F^0}\mathbf{X}_{*,J_N} \\
(NT)^{-1}a_{21}^0\mathbf{X}'_{*,J_2}\mathbb{M}_{F^0}\mathbf{X}_{*,J_1} & \cdots & (NT)^{-1}a_{2N}^0\mathbf{X}'_{*,J_2}\mathbb{M}_{F^0}\mathbf{X}_{*,J_N} \\
\vdots & \ddots & \vdots \\
(NT)^{-1}a_{N1}^0\mathbf{X}'_{*,J_N}\mathbb{M}_{F^0}\mathbf{X}_{*,J_1} & \cdots & (NT)^{-1}a_{NN}^0\mathbf{X}'_{*,J_N}\mathbb{M}_{F^0}\mathbf{X}_{*,J_N}
\end{bmatrix}. \tag{B.5}$$

Let $\check{\mathbf{U}} \equiv [T^{-1}(\mathbf{X}'_{*,J_1}\mathbb{M}_{F^0}\mathbf{U}_{*,1})', ..., T^{-1}(\mathbf{X}'_{*,J_N}\mathbb{M}_{F^0}\mathbf{U}_{*,N})']'$, which is a $\sum_{i=1}^N k_i \times 1$ vector.

To continue the proof, we need the following four lemmas whose proofs are given at the end of next subsection.

Lemma B.2 *Suppose that Assumptions A.1 and A.3-A.5 hold. Let $\delta_{NT} = \sqrt{N} \wedge \sqrt{T}$. Then*
(i) $||I_\ell^*||_F = O_P(\delta_{NT}^{-1})$ *for* $\ell = 1, 2, 3$, *and*
(ii) $||I_\ell^*||_F = O_P(d_{NT})$ *for* $\ell = 4, 5, ..., 8$,
(iii) $T^{-1/2}\left\|\check{F} - F^0\check{H}\right\| = O_P(\delta_{NT}^{-1} + d_{NT})$.

Lemma B.3 *Suppose that Assumptions A.1 and A.3-A.5 hold. Then*
(i) $T^{-1}F^{0\prime}(\check{F} - F^0\check{H}) = O_P(d_{NT} + \delta_{NT}^{-2})$;
(ii) $\check{H}\check{H}' - (F^{0\prime}F^0/T)^{-1} = O_P(d_{NT} + \delta_{NT}^{-2})$;

3

(iii) $T^{-1}\mathbf{U}'_{*,i}(\check{F} - F^0\check{H}) = O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})$;

(iv) $T^{-1}\mathbf{X}'_{*,i}(\check{F} - F^0\check{H}) = O_P(d_{NT} + \delta_{NT}^{-2})$.

Lemma B.4 *Suppose that Assumptions A.1 and A.3-A.5 hold. Let S_i be an arbitrary $L \times k_i$ non-random matrix such that $\|S_i\|_F \leq C < \infty$ and L is a fixed integer. Then*

(i) $\|\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0}\|_F = O_P(d_{NT} + \delta_{NT}^{-2})$;

(ii) $T^{-1}S_i\mathbf{X}'_{*,J_i}(\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0})\mathbf{U}_{*,i} = O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})$;

(iii) $T^{-1}S_i\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(\check{F}-F^0\check{H})\check{H}^{-1}\lambda_i^0 = -\frac{1}{NT}S_i\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}(\check{B}-B^0)\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0+O_P(\delta_{NT}^{-2}+\delta_{NT}^{-1}d_{NT})$.

Lemma B.5 *Suppose that Assumptions A.1 and A.3-A.5 hold. Then*

(i) $\left|\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}\right|_\infty = O_P(K_J K_a^{1/2}[T^{-1}(\log N)^2 + N^{-1}])$;

(ii) $N^{1/2}|\Gamma\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}| = O_P(K_J K_a[T^{-1}(\log N)^2 + N^{-1}]) = o_P(T^{-1/2})$ *for any conformable square matrix Γ with $\|\Gamma\|_{op} = O_P(1)$;*

(iii) $\psi_{\max}(\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}) < 1$ *with probability approaching one (w.p.a.1).*

Note that we allow k_i to be divergent. For a k_i-vector A_i, we introduce the weighted norm $\|\cdot\|_{S_i}$ such that $\|A_i\|_{S_i} = |S_i A_i|$, where the number of rows S_i is given by L, a fixed integer, and $\|S_i\|_F$ is bounded above by a constant. We denote $A_i = o_{P*}(c_{NT})$ if $\|A_i\|_{S_i} = o_P(c_{NT})$ for any S_i with bounded Frobenius norm. Define O_{P*} analogously.

By the identity $\mathbf{Y}_{*,i} = \mathbf{X}_{*,J_i}B_{J_i,i}^0 + F^0\lambda_i^0 + \mathbf{U}_{*,i}$ and (B.1), we obtain that

$$T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}\mathbf{X}_{*,J_i}(\check{B}_{J_i,i} - B_{J_i,i}^0) = T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}(F^0\lambda_i^0 + \mathbf{U}_{*,i}).$$

For $T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}\mathbf{U}_{*,i}$, we have

$$\begin{aligned}
T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}\mathbf{U}_{*,i} &= T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i} - T^{-1}\mathbf{X}'_{*,J_i}(\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0})\mathbf{U}_{*,i} \\
&= T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i} + O_{P*}(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT}),
\end{aligned}$$

where the second equality holds by Lemma B.4(ii). For $T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}F^0\lambda_i^0$, we have

$$\begin{aligned}
T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}F^0\lambda_i^0 &= -T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{\check{F}}(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 \\
&= -T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 - T^{-1}\mathbf{X}'_{*,J_i}(\mathbb{P}_{F^0} - \mathbb{P}_{\check{F}})(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 \\
&= \frac{1}{NT}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}(\check{B} - B^0)\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0 + O_{P*}(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT}),
\end{aligned}$$

where the last equality follows from Lemmas B.4(i) and (iii) and B.2(iii). It follows that for each $i \in [N]$, we have

$$T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}(\check{B}_{J_i,i} - B_{J_i,i}^0) = \frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i} + \frac{1}{NT}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}(\check{B}-B^0)\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0 + \check{R}_i,$$

$$\tag{B.6}$$

4

where $\check{R}_i = O_{P*}(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})$ and its exact form is given by

$$
\begin{aligned}
\check{R}_i &= -T^{-1}\mathbf{X}'_{*,J_i}(\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0})\mathbf{U}_{*,i} - T^{-1}\mathbf{X}'_{*,J_i}(\mathbb{P}_{F^0} - \mathbb{P}_{\check{F}})(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 \\
&\quad -T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 - \frac{1}{NT}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}(\check{B} - B^0)\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0.
\end{aligned}
$$

Let $\check{\boldsymbol{\beta}} = (\check{B}'_{J_1,1}, ..., \check{B}'_{J_N,N})'$ and $\boldsymbol{\beta}^0 = (B^{0\prime}_{J_1,1}, ..., B^0_{J_N,N})$. Then (B.6) can be written as follows:

$$(\check{Q}_1 - \check{Q}_2)(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = \check{\mathbf{U}} + \check{\mathbf{R}}, \tag{B.7}$$

where $\check{\mathbf{R}} = (\check{R}'_1, ..., \check{R}'_N)$ and $\check{\mathbf{U}} = (\check{U}'_1, ..., \check{U}'_N)$ with $\check{U}_i = T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i}$. Note that the minimum eigenvalue of $\check{Q}_1$ is bounded below by some constant w.p.a.1; see Lemma A.5(i). Following the proof of Lemmas B.4(ii)-(iii) and using Lemma B.2(iii), we can also show that

$$\left|\check{\mathbf{R}}\right|_\infty = O_{P*}((\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})\log N) \text{ and } \frac{1}{\sqrt{N}}\left|\check{\mathbf{R}}\right| = O_{P*}(K_a^{1/2}(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})). \tag{B.8}$$

Rewriting (B.7), one can obtain

$$\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \check{Q}_1^{-1}\check{\mathbf{U}} + \check{Q}_1^{-1}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \check{Q}_1^{-1}\check{\mathbf{R}}. \tag{B.9}$$

Iterating (B.9) $\ell \geq 2$ times, we obtain

$$
\begin{aligned}
&\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \\
=~& \check{Q}_1^{-1}\check{\mathbf{U}} + \check{Q}_1^{-1}\check{\mathbf{R}} + \check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}(\check{\mathbf{U}} + \check{\mathbf{R}}) \\
=~& \check{Q}_1^{-1}\check{\mathbf{U}} + \check{Q}_1^{-1}\check{\mathbf{R}} + \check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]\check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]\check{Q}_1^{-1/2}(\check{\mathbf{U}} + \check{\mathbf{R}}) \\
=~& \check{Q}_1^{-1}\check{\mathbf{U}} + \check{Q}_1^{-1}\check{\mathbf{R}} + \check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^2\check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \check{Q}_1^{-1/2}\sum_{l=1}^{2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l\check{Q}_1^{-1/2}(\check{\mathbf{U}} + \check{\mathbf{R}}) \\
=~& ... \\
=~& \check{Q}_1^{-1}\check{\mathbf{U}} + \check{Q}_1^{-1}\check{\mathbf{R}} + \check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^\ell\check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\
&+ \check{Q}_1^{-1/2}\sum_{l=1}^{\ell}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l\check{Q}_1^{-1/2}\check{\mathbf{U}} + \check{Q}_1^{-1/2}\sum_{l=1}^{\ell}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l\check{Q}_1^{-1/2}\check{\mathbf{R}} \\
\equiv~& \check{Q}_1^{-1}\check{\mathbf{U}} + \check{\mathbf{R}}_1 + \check{\mathbf{R}}_2 + \check{\mathbf{R}}_3 + \check{\mathbf{R}}_4, \tag{B.10}
\end{aligned}
$$

where we suppress the dependence of $\check{\mathbf{R}}_2$, $\check{\mathbf{R}}_3$ and $\check{\mathbf{R}}_4$ on ℓ. Define a $k_i \times \sum_{j=1}^{N} k_j$ selection matrix $\mathbb{S}_i$

such that $\check{B}_{J_i,i} - B^0_{J_i,i} = \mathbb{S}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$. Then

$$
\begin{aligned}
\check{B}_{J_i,i} - B^0_{J_i,i} &= \mathbb{S}_i\check{Q}_1^{-1}\check{\mathbf{U}} + \mathbb{S}_i\check{Q}_1^{-1}\check{\mathbf{R}} + \mathbb{S}_i\check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^\ell \check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\
&\quad + \mathbb{S}_i\check{Q}_1^{-1/2}\sum_{l=1}^{\ell}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l \check{Q}_1^{-1/2}\check{\mathbf{U}} + \mathbb{S}_i\check{Q}_1^{-1/2}\sum_{l=1}^{\ell}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l \check{Q}_1^{-1/2}\check{\mathbf{R}} \\
&\equiv \mathbb{S}_i\check{Q}_1^{-1}\check{\mathbf{U}} + \check{R}_{1i} + \check{R}_{2i} + \check{R}_{3i} + \check{R}_{4i}. \quad\quad\quad (B.11)
\end{aligned}
$$

Note that $\check{\mathbf{R}}_l = (\check{R}'_{l1}, ..., \check{R}'_{lN})'$ and $\mathbb{S}_i\check{\mathbf{R}}_l = \check{R}_{li}$ for $l = 1, 2, 3, 4$. Let $\chi_{ij} \equiv T^{-1}X'_{*,J_i}\mathbb{M}_{F^0}X'_{*,J_j}$ for $i, j \in [N]$.

We first study $\check{R}_{1i}$. Noting that $|a'b| \leq |a|_1 |b|_\infty$ for any two conformable vectors a and b, we have

$$
\begin{aligned}
\left|S_i\check{R}_{1i}\right|^2 &= \sum_{l=1}^{L}\left|[S_i]_{l,*}\mathbb{S}_i\check{Q}_1^{-1}\check{\mathbf{R}}\right|^2 \leq \sum_{l=1}^{L}\left|[S_i]_{l,*}\mathbb{S}_i\check{Q}_1^{-1}\right|_1^2 \left|\check{\mathbf{R}}\right|_\infty^2 \\
&\leq \sum_{l=1}^{L}\left|[S_i]_{l,*}\chi_{ii}^{-1}\right|_1^2 \left|\check{\mathbf{R}}\right|_\infty^2 \leq k_i\sum_{l=1}^{L}\left|[S_i]_{l,*}\chi_{ii}^{-1}\right|_2^2 \left|\check{\mathbf{R}}\right|_\infty^2 \\
&= O_P(K_J)\left|\check{\mathbf{R}}\right|_\infty^2 = O_P(K_J[(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})\log N]^2), \quad\quad (B.12)
\end{aligned}
$$

and

$$
\begin{aligned}
N^{-1}\left|\check{\mathbf{R}}_1\right|^2 &= N^{-1}\sum_{i=1}^{N}\left|\check{R}_{1i}\right|^2 = N^{-1}\sum_{i=1}^{N}\left|\mathbb{S}_i\check{Q}_1^{-1}\check{R}_i\right|^2 \leq \max_{i\in[N]}\left\|\mathbb{S}_i\check{Q}_1^{-1}\right\|_F^2 N^{-1}\sum_{i=1}^{N}\left|\check{R}_i\right|^2 \\
&= K_J O_P(N^{-1}|\check{\mathbf{R}}|^2) = K_J O_{P^*}(K_a(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})^2) = o_P\left(T^{-1} + d_{NT}^2\right). \text{(B.13)}
\end{aligned}
$$

Next, we study $\check{R}_{2i}$. By Lemma B.5(iii), $\left\|[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^\ell\right\|_{\text{op}} = \left\|\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}\right\|_{\text{op}}^\ell \to 0$ at the exponential rate as $\ell \to \infty$. This ensures that

$$
\begin{aligned}
\max_i\left|S_i\check{R}_{2i}\right| &= \left|S_i\mathbb{S}_i\check{Q}_1^{-1/2}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^\ell \check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\right| \\
&\leq \left\|S_i\mathbb{S}_i\check{Q}_1^{-1/2}\right\|_{\text{op}}\left\|\check{Q}_1^{-1/2}\check{Q}_2(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\right\|\left\|\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}\right\|_{\text{op}}^\ell = o_P(T^{-1/2}), \text{ (B.14)} \\
N^{-1}\left|\check{\mathbf{R}}_2\right|^2 &= N^{-1}\sum_{i=1}^{N}\left|\check{R}_{2i}\right|^2 = o_P(T^{-1}), \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (B.15)
\end{aligned}
$$

for sufficiently large ℓ.

To study $\check{R}_{3i}$, let $\Gamma_{0\ell} = \check{Q}_1^{-1/2}\sum_{l=0}^{\ell-1}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l\check{Q}_1^{1/2}$. Then

$$
\check{R}_{3i} = \mathbb{S}_i\check{Q}_1^{-1/2}\sum_{l=0}^{\ell-1}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l\check{Q}_1^{1/2}\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}} = \mathbb{S}_i\Gamma_{0\ell}\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}.
$$

6

By Lemma B.5(iii), $\|\Gamma_{0\ell}\|_{\mathrm{op}} \leq \left\|\check{Q}_1^{-1/2}\right\|_{\mathrm{op}} \sum_{l=0}^{\ell-1} \left\|\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}\right\|_{\mathrm{op}}^{l} \left\|\check{Q}_1^{1/2}\right\|_{\mathrm{op}} = O_P(1)$ for each ℓ. Then by Lemma B.5(i)-(ii),

$$\left|S_i\check{R}_{3i}\right| = O_P(K_J K_a^{1/2}[T^{-1}(\log N)^2 + (NT)^{-1/2}]) = o_P(K_a^{-1/2}T^{-1/2}) \text{ and} \quad (\text{B.16})$$

$$N^{-1}\left|\check{\mathbf{R}}_3\right|^2 = N^{-1}\sum_{i=1}^{N}\left|\check{R}_{3i}\right|^2 = o_P\left(T^{-1}\right). \quad (\text{B.17})$$

To study R_{4i}, let $\Gamma_\ell = \sum_{l=1}^{\ell}[\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}]^l$. Then $R_{4i} = \mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}\check{\mathbf{R}}$. Note that $S_i\mathbb{S}_i\check{Q}_1^{-1/2}$ $\times\Gamma_\ell\check{Q}_1^{-1/2}$ has low dimension of rows such that

$$\left\|S_i\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_{1\ell}\check{Q}_1^{-1/2}\right\|_{\mathrm{F}}^2 = \mathrm{tr}\left(S_i\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1}\Gamma_\ell\check{Q}_1^{-1/2}\mathbb{S}_i'C_i'\right)$$

$$\leq \|\Gamma_\ell\|_{\mathrm{op}}^2\left\|\check{Q}_1^{-1}\right\|_{\mathrm{op}}^2\|\mathbb{S}_i\|_{\mathrm{op}}^2\|S_i\|_{\mathrm{F}}^2 = O_P(1) \text{ uniformly in } i \in [N],$$

where we use the fact that $\|\Gamma_\ell\|_{\mathrm{op}} \leq \sum_{l=1}^{\ell}\left\|\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}\right\|_{\mathrm{op}}^l = O_P(1)$ by Lemma B.5(iii), $\left\|\check{Q}_1^{-1}\right\|_{\mathrm{op}} = [\psi_{\min}(\check{Q}_1)]^{-1} = O_P(1)$, $\|\mathbb{S}_i\|_{\mathrm{op}} = 1$, and $\max_{i\in[N]}\|S_i\|_{\mathrm{F}} = O(1)$ by assumption. Write $\Gamma_\ell = \{\Gamma_{\ell,ij}\}$ as a block partitioned matrix with $\Gamma_{\ell,ij}$ being a $k_i \times k_j$ matrix. Then

$$\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2} = \chi_{ii}^{-1/2}(\Gamma_{\ell,i1}\chi_{11}^{-1/2}, \Gamma_{\ell,i2}\chi_{22}^{-1/2}, ..., \Gamma_{\ell,iN}\chi_{NN}^{-1/2}).$$

Our assumptions ensure that Γ_ℓ is absolutely column summable, which implies that the absolute sum of each row of $\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}$ can be bounded by a constant multiplied by $\max_{i\in[N]}\left\|\chi_{ii}^{-1}\right\|_{\mathrm{op}}$ Consequently,

$$|S_iR_{4i}|^2 = \sum_{l=1}^{L}\left|[S_i]_{l,*}\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}\check{\mathbf{R}}\right|^2 \leq \sum_{l=1}^{L}\left|[S_i]_{l,*}\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}\right|_1^2\left|\check{\mathbf{R}}\right|_{\max}^2$$

$$\leq O_P(\max_{i\in[N]}\left\|\chi_{ii}^{-1}\right\|_{\mathrm{op}})\left|\check{\mathbf{R}}\right|_{\max}^2 = O_P((\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})\log N) \quad (\text{B.18})$$

and

$$N^{-1}\left|\check{\mathbf{R}}_4\right|^2 = N^{-1}\sum_{i=1}^{N}\left|\check{R}_{4i}\right|^2 = N^{-1}\sum_{i=1}^{N}\left|\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}\check{R}_i\right|^2$$

$$\leq \max_{i\in[N]}\left\|\mathbb{S}_i\check{Q}_1^{-1/2}\Gamma_\ell\check{Q}_1^{-1/2}\right\|_{\mathrm{F}}^2 N^{-1}\sum_{i=1}^{N}\left|\check{R}_i\right|^2$$

$$= K_J O_P(N^{-1}\left|\check{\mathbf{R}}\right|^2) = K_J O_{P^*}(K_a(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})^2) = o_P\left(T^{-1} + d_{NT}^2\right). (\text{B.19})$$

In sum, we have shown that

$$N^{-1}\sum_{l=1}^{4}\left|\check{\mathbf{R}}_l\right|^2 = o_P\left(T^{-1} + d_{NT}^2\right) \text{ and } \left|S_i\sum_{l=1}^{4}\check{R}_{li}\right| = o_P(T^{-1/2}) + O_P(K_J^{1/2}\delta_{NT}^{-1}d_{NT}\log N). \quad \text{(B.20)}$$

In addition,

$$\begin{aligned}
N^{-1}|\check{Q}_1^{-1}\check{\mathbf{U}}|^2 &= \frac{1}{N}\sum_{i=1}^{N}|\chi_{ii}^{-1}\check{U}_i|^2 = \frac{1}{N}\sum_{i=1}^{N}\left|\chi_{ii}^{-1}\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i}\right|^2 \\
&\leq \max_i\left\|\chi_{ii}^{-1}\right\|_{\text{op}}^2\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i}\right|^2 = O_P\left(K_aT^{-1}\right) \quad \text{(B.21)}
\end{aligned}$$

Combining the results in (B.10), (B.13), (B.15), (B.17), (B.19), and (B.21), we have

$$N^{-1}\left|\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right|^2 = O_P\left(K_aT^{-1}\right) + o_P\left(d_{NT}^2\right) = O_P\left(K_aT^{-1}\right) + o_P\left(N^{-1}\left|\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right|^2\right).$$

It follows that $d_{NT} \equiv N^{-1/2}\|\check{B} - B^0\|_{\text{F}} = N^{-1/2}\left|\check{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right| = O_P((K_a/T)^{1/2})$.

In addition, by (B.11) and the results in (B.12), (B.14), (B.16), and (B.18), we have

$$\begin{aligned}
S_i(\check{B}_{J_i,i} - B^0_{J_i,i}) &= S_i\mathbb{S}_i\check{Q}_1^{-1}\check{\mathbf{U}} + O_P(K_J^{1/2}\delta_{NT}^{-1}d_{NT}\log N) + o_P(T^{-1/2}) \\
&= S_i(\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i})^{-1}\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i} + o_P(T^{-1/2}),
\end{aligned}$$

where the second equality holds by the fact that $K_J^{1/2}\delta_{NT}^{-1}d_{NT}\log N = K_J^{1/2}\delta_{NT}^{-1}O_P((K_a/T)^{1/2})\log N$ $= o_P(T^{-1/2})$ as $K_J^{1/2}K_a^{1/2}(\log N)\delta_{NT}^{-1} = o(1)$. This completes the proof of the proposition. ∎

B.2 Proof of the technical lemmas

Proof of Lemma A.1. The proof follows from that of Lemma C.2 in Chernozhukov et al. (2018). ∎

Proof of Lemma A.2. (i) By direct calculation, we have that

$$\frac{1}{T}\left\|\mathbf{U}'\mathbf{X}\right\|_{\max} = \max_{1\leq l\leq p}\max_{1\leq i\leq N}\max_{1\leq j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}y_{i,t-l}u_{jt}\right|.$$

By Lemma C.1 in the next section, we have that for some constants C_1, C_2 and c_1

$$\begin{aligned}
P\left(\left|\sum_{t=1}^{T}y_{i,t-l}u_{jt}\right| > \frac{T\gamma_1}{2}\right) &\leq 2^{q/2}C_1\frac{T}{(T\gamma_1)^{q/2}} + C_2\exp\left(-\frac{C_3(T\gamma_1)^2}{4T}\right) \\
&= 2^{q/2}C_1\frac{T^{1-q/4}}{(c_1\log N)^{q/2}} + C_2\exp\left(-\frac{C_3(c_1\log N)^2}{4}\right) \\
&= C_1'T^{1-q/4}(\log N)^{-q/2} + C_2'N^{-\underline{c}\log N},
\end{aligned}$$

8

where the first equality holds by inserting $\gamma_1 = c_1 T^{-1/2} \log N$ and the second equality holds by redefining the absolute constants $\underline{c}$, C_1' and C_2'. Then by the union bound, we have

$$P\left(\frac{1}{T}\left\|\mathbf{U}'\mathbf{X}\right\|_{\max} > \frac{\gamma_1}{2}\right) \leq p\sum_{i,j}\max_{1\leq l\leq p}P\left(\left|\sum_{t=1}^{T}y_{i,t-l}u_{jt}\right| > \frac{T\gamma_1}{2}\right)$$

$$\leq pC_1'N^2T^{1-q/4}(\log N)^{-q/2} + pC_2'N^{2-\underline{c}\log N}.$$

Letting $C = p(C_1' \vee C_2')$, we have proved the desired result in (i).

(ii) Noting that $\mathbb{P}_{F^0} = F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$, we have

$$\frac{1}{T}\left\|\mathbf{U}'\mathbb{P}_{F^0}\mathbf{X}\right\|_{\max} = \frac{1}{T}\|\mathbf{U}'F^0(F^{0\prime}F^0)^{-1}F^{0\prime}\mathbf{X}\|_{\max}$$

$$\leq \left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|_{\mathrm{op}}\max_{1\leq i\leq Np}\left\|\frac{F^{0\prime}\mathbf{X}_{*,i}}{T}\right\| \cdot \max_{1\leq i\leq N}\left\|\frac{F^{0\prime}\mathbf{U}_{*,i}}{T}\right\|.$$

As in the proof of (i), we can show that

$$P\left(\max_{1\leq r\leq R^0}\max_{1\leq i\leq N}\left|\sum_{t=1}^{T}f_{r,t}^0 u_{it}\right| > T\gamma_1\right) \leq C_1 NT^{1-q/4}(\log N)^{-q/2} + C_2 N^{1-\underline{c}\log N}.$$

By Lemma A.3 below and choosing $\bar{c} = \frac{1}{2}\left[\psi_{\min}\left(\Sigma_F\right)\right]^{-1}$, we can readily show that

$$P\left(\left\|(F^{0\prime}F^0/T)^{-1}\right\|_{\mathrm{op}} > \bar{c}\right) \leq C_1 NT^{1-q/4}(\log N)^{-q/2} + C_2 N^{1-\underline{c}\log N}.$$

Similarly, $\max_{1\leq i\leq Np}\left\|F^{0\prime}\mathbf{X}_{*,i}/T\right\|$ is bounded by some constant $\bar{c}$ with probability larger than $1 - [C_1 NT^{1-q/4}(\log N)^{-q/2} + C_2 N^{1-\underline{c}\log N}]$. Consequently, we have proved (ii). ∎

Proof of Lemma A.3. By Lemma C.1 in the next Section, we have

$$P\left(\left|\sum_{t=1}^{T}f_{t,r}^0 f_{t,l}^0 - E(f_{t,r}^0 f_{t,l}^0)\right| \geq xT^{1/2}\right) \leq C_1 x^{-q/2}T^{1-q/4} + C_2\exp(-C_3 x^2)$$

for some absolute constants C_1, C_2, and C_3. Applying the union bound yields the desired result. ∎

Proof of Lemma A.4. (i) By the similar arguments as used in the proof of Theorem 3.2, we can establish the result.

(ii) By the proof of Proposition 2.1, $E(y_{it}^2)$ is bounded uniformly in (i,t). By direct calculations,

we have that

$$(\max_{1\leq j\leq pN}|\mathbf{X}_{*,j}|/\sqrt{T})^2 \quad = \quad \max_{1\leq i\leq N}\max_{1\leq l\leq p}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-l}^2$$

$$\leq \quad \max_{1\leq i\leq N}\max_{1\leq l\leq p}\frac{1}{T}\left|\sum_{t=1}^{T}[y_{i,t-l}^2 - E(y_{i,t-l}^2)]\right| + \bar{c},$$

for some constant $\bar{c} > 0$. By Lemma C.1 in the next section, we have

$$P\left(\left|\sum_{t=1}^{T}[y_{i,t-l}^2 - E(y_{i,t-l}^2)]\right| > \bar{c}T\right) \quad \leq \quad 2^{q/2}C_1\frac{T}{(\bar{c}T)^{q/2}} + C_2\exp\left(-\frac{C_3(\bar{c}T)^2}{4T}\right)$$

$$= \quad C_1'T^{1-q/2} + C_2'\exp\left(-C_3T\right).$$

Applying the union bound delivers the first result. The second result can be shown analogously.

(iii) The proof is similar to that of (ii) and omitted.

(iv) To proceed, we operate conditional on $||T^{-1}\mathbf{X}'F^0 - \Sigma_{XF}||_{\max} \leq \bar{c}T^{-1/2}$, $||\tilde{F} - F^0\tilde{H}||_{\mathrm{F}}/\sqrt{T} \leq \bar{c}(\gamma_1\sqrt{K_a} \vee \gamma_2)$, $\max_{1\leq j\leq pN}|\mathbf{X}_{*,j}|/\sqrt{T} \leq \bar{c}$, and $T^{-1}||F^{0\prime}\mathbf{X}||_1 \leq \bar{c}$. One can easily show that these joint events hold with probability at least $1 - \bar{c}'(N^2T^{1-q/4}(\log N)^{-q/2} + N^{(2-\underline{c}\log N)\vee(-1)})$. By the triangle inequality, we have

$$||\tilde{\Sigma} - \Sigma||_{\max} \leq ||T^{-1}\mathbf{X}'\mathbf{X} - \Sigma_X||_{\max} + ||T^{-2}\mathbf{X}'\tilde{F}\tilde{F}'\mathbf{X} - \Sigma_{XF}\Sigma_F^{-1}\Sigma_{XF}'||_{\max}.$$

For the first term on the right hand side (RHS) of the last equation, we can apply similar arguments as used in the proof of part (ii) to establish that $||T^{-1}\mathbf{X}'\mathbf{X} - \Sigma_X||_{\max} \leq \bar{c}$ with probability larger than $1 - \bar{c}N^2(T^{1-q/2}+\exp(-T))$. For the second term, we have

$$||T^{-2}\mathbf{X}'\tilde{F}\tilde{F}'\mathbf{X} - \Sigma_{XF}\Sigma_F^{-1}\Sigma_{XF}'||_{\max} \quad \leq \quad 2||T^{-2}\mathbf{X}'(\tilde{F} - F^0\tilde{H})\tilde{F}'\mathbf{X}||_{\max}$$

$$+||T^{-2}\mathbf{X}'(\tilde{F} - F^0\tilde{H})(\tilde{F} - F^0\tilde{H})'\mathbf{X}||_{\max}$$

$$+||T^{-2}\mathbf{X}'F^0\tilde{H}\tilde{H}'F^{0\prime}\mathbf{X} - \Sigma_{XF}\Sigma_F^{-1}\Sigma_{XF}'||_{\max}.$$

Noting that $\left\|\tilde{F}\right\|_{\mathrm{F}}/\sqrt{T} = R^0$, we have $||T^{-2}\mathbf{X}'(\tilde{F} - F^0\tilde{H})\tilde{F}'\mathbf{X}||_{\max} \leq R^0(\max_{1\leq j\leq pN}|\mathbf{X}_{*,j}|/\sqrt{T})^2$ $\cdot||\tilde{F} - F^0\tilde{H}||_{\mathrm{F}}/\sqrt{T} \leq \bar{c}\sqrt{\log N}(\gamma_1\sqrt{K_a} \vee \gamma_2)$ with probability at least $1 - \bar{c}'(N^2T^{1-q/4}(\log N)^{-q/2} + N^{(2-\underline{c}\log N)\vee(-1)})$. From the proof of Theorem 3.2, $(F^0\tilde{H})'F^0\tilde{H}/T = I_{R^0}$. This implies that $\tilde{H}\tilde{H}' = $

$(F^{0\prime}F^0/T)^{-1}$. Then

$$||T^{-2}\mathbf{X}'F^0\tilde{H}\tilde{H}'F^{0\prime}\mathbf{X}-\Sigma_{XF}\Sigma_F^{-1}\Sigma_{XF}'||_{\max}$$

$$= ||T^{-2}\mathbf{X}'F^0(F^{0\prime}F^0/T)^{-1}F^{0\prime}\mathbf{X}-\Sigma_{XF}\Sigma_F^{-1}\Sigma_{XF}||_{\max}$$

$$\leq ||(T^{-1}\mathbf{X}'F^0-\Sigma_{XF})(F^{0\prime}F^0/T)^{-1}T^{-1}F^{0\prime}\mathbf{X}||_{\max} + ||\Sigma_{XF}[(F^{0\prime}F^0/T)^{-1} - \Sigma_F^{-1}]T^{-1}F^{0\prime}\mathbf{X}||_{\max}$$

$$+||\Sigma_{XF}\Sigma_F^{-1}(T^{-1}F^{0\prime}\mathbf{X}-\Sigma_{XF}')||_{\max}.$$

For $||(T^{-1}\mathbf{X}'F^0-\Sigma_{XF})(F^{0\prime}F^0/T)^{-1}T^{-1}F^{0\prime}\mathbf{X}||_{\max}$, we have with probability at least $1-\vec{c}'(N^2T^{1-q/4}$
$\times(\log N)^{-q/2} + N^{(2-\underline{c}\log N)\vee(-1)})$,

$$||(T^{-1}\mathbf{X}'F^0-\Sigma_{XF})(F^{0\prime}F^0/T)^{-1}T^{-1}F^{0\prime}\mathbf{X}||_{\max} \leq ||T^{-1}\mathbf{X}'F^0-\Sigma_{XF}||_{\max} \cdot ||(F^{0\prime}F^0/T)^{-1}T^{-1}F^{0\prime}\mathbf{X}||_1$$

$$\leq \bar{c}||T^{-1}\mathbf{X}'F^0-\Sigma_{XF}||_{\max} \leq \bar{c}T^{-1/2}.$$

The other two terms can be bounded similarly.

(v) This result can be proved by arguments as used in the proof of Lemma A.2 (i). With the bound $||\tilde{\Sigma} - \Sigma||_{\max} \leq \gamma_3$, the proof is similar to Lemma 10.1 in van de Geer and Bühlmann (2009). Let $v \in \mathbb{R}^{Np}$ such that $|v_{J^c}|_1 \leq 3|v_J|_1$, and $|J| \leq K_J$. One has

$$|v'\tilde{\Sigma}v - v'\Sigma v| = |v'(\tilde{\Sigma} - \Sigma)v| \leq |v|_1 \cdot |(\tilde{\Sigma} - \Sigma)v|_\infty$$

$$\leq ||\tilde{\Sigma} - \Sigma||_{\max}|v|_1^2 \leq \gamma_3|v|_1^2$$

$$\leq 16\gamma_3|v_J|_1^2 \leq 16K_J\gamma_3 \cdot |v_J|_2^2.$$

After some rearrangement, we have

$$\frac{v'\tilde{\Sigma}v}{|v_J|_2^2} \geq \frac{v'\Sigma v}{|v_J|_2^2} - 16K_J\gamma_3 \geq \frac{v'\Sigma v}{|v|_2^2} - 16K_J\gamma_3$$

$$\geq \psi_{\min}(\Sigma) - 16K_J\gamma_3 \geq \psi_{\min}(\Sigma)/2.$$

It follows that the restricted eigenvalue condition is satisfied with $\kappa_{\tilde{\Sigma}}(K_J) \leq \psi_{\min}(\Sigma)/2$. ∎

Proof of Lemma A.5. Let $\hat{\Sigma} \equiv T^{-1}\mathbf{X}'\mathbb{M}_{F^0}\mathbf{X}$. Then we denote the two types of submatrices of $\hat{\Sigma}$ as $\hat{\Sigma}_{J_i,J_i} \equiv T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}$ and $\hat{\Sigma}_{J_i^c,J_i} = T^{-1}\mathbf{X}'_{*,J_i^c}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}$ for $i \in [N]$. Then we have that

$$\max_{i\in[N]} ||\hat{\Sigma}_{J_i,J_i} - \tilde{\Sigma}_{J_i,J_i}||_{\mathrm{op}} = \max_{i\in[N]} T^{-1}||\mathbf{X}'_{*,J_i}(\mathbb{P}_{F^0} - \mathbb{P}_{\tilde{F}})\mathbf{X}_{*,J_i}||_{\mathrm{op}}$$

$$\leq K_J \cdot T^{-1}||\mathbf{X}'(\mathbb{P}_{F^0} - \mathbb{P}_{\tilde{F}})\mathbf{X}||_{\max}$$

$$\leq K_J \cdot ||\mathbb{P}_{F^0} - \mathbb{P}_{\tilde{F}}||_{\mathrm{op}} \cdot \max_{1\leq j\leq Np} T^{-1}|\mathbf{X}_{*,j}|^2 = o_P(1).$$

Similarly, we have that $\max_{i\in[N]} ||\hat{\Sigma}_{J_i^c,J_i} - \tilde{\Sigma}_{J_i^c,J_i}||_{\max} = o_P(1)$. To prove the lemma, it suffices to

establish that (a) $\min_{i\in[N]} \psi_{\min}(\hat{\Sigma}_{J_i,J_i}) > \underline{c}$ and (b) $\max_{i\in[N]} \max_{j\in J_i^c} ||T^{-1}\mathbf{X}'_{*,j}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}||_{\max} < \bar{c}$ for some constants $\underline{c}$ and $\bar{c}$ w.p.a.1.

(a) Recall the decomposition in Section 2.2, and let $\mathbf{X} = \mathbf{X}^{(u)} + \mathbf{X}^{(f)}$, where the tth rows of $\mathbf{X}^{(u)}$ and $\mathbf{X}^{(f)}$ are $X_t^{(u)\prime}$ and $X_t^{(f)\prime}$ respectively. To establish $\min_{i\in[N]} \psi_{\min}(\hat{\Sigma}_{J_i,J_i}) > \underline{c}$, we decompose $\hat{\Sigma}_{J_i,J_i}$ to

$$\hat{\Sigma}_{J_i,J_i} = \hat{\Sigma}^{(u)}_{J_i,J_i} + \hat{\Sigma}^{(f)}_{J_i,J_i} + T^{-1}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}^{(f)}_{*,J_i} + T^{-1}\mathbf{X}^{(f)\prime}_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}^{(u)}_{*,J_i} \equiv S_{1i} + S_{2i} + S_{3i} + S_{4i},$$

where $\hat{\Sigma}^{(u)}_{J_i,J_i} \equiv T^{-1}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}^{(u)}_{*,J_i}$ and $\hat{\Sigma}^{(f)}_{J_i,J_i} \equiv T^{-1}\mathbf{X}^{(f)\prime}_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}^{(f)}_{*,J_i}$.

First, we consider S_{1i}. One can decompose $\hat{\Sigma}^{(u)}_{J_i,J_i}$ as

$$\hat{\Sigma}^{(u)}_{J_i,J_i} = \frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbf{X}^{(u)}_{*,J_i} - \frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbb{P}_{F^0}\mathbf{X}^{(u)}_{*,J_i}.$$

It follows that

$$\min_{i\in[N]} \psi_{\min}(\hat{\Sigma}^{(u)}_{J_i,J_i}) \geq \min_{i\in[N]} \psi_{\min}(\Sigma^{(u)}_{J_i,J_i}) - \max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbf{X}^{(u)}_{*,J_i} - \Sigma^{(u)}_{J_i,J_i}||_{op} - \max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbb{P}_{F^0}\mathbf{X}^{(u)}_{*,J_i}||_{op}.$$

Uniformly across i, one has $\psi_{\min}(\Sigma^{(u)}_{J_i,J_i}) \geq \psi_{\min}(\Sigma^{(u)}_X)$, where $\psi_{\min}(\Sigma^{(u)}_X)$ is bounded below by Proposition 2.1. By inequality

$$\max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbf{X}^{(u)}_{*,J_i} - \Sigma^{(u)}_{J_i,J_i}||_{op} \leq K_J \max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbf{X}^{(u)}_{*,J_i} - \Sigma^{(u)}_{J_i,J_i}||_{\max}$$

$$\leq K_J \cdot ||\frac{1}{T}\mathbf{X}^{(u)\prime}\mathbf{X}^{(u)} - \Sigma^{(u)}_X||_{\max} = O_P(K_J T^{-1/2}\log N) = o_P(1).$$

Therefore, we have established that $\max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbf{X}^{(u)}_{*,J_i} - \Sigma^{(u)}_{J_i,J_i}||_{op} = o_P(1)$. Similarly, we can show that $\max_{i\in[N]}||\frac{1}{T}\mathbf{X}^{(u)\prime}_{*,J_i}\mathbb{P}_{F^0}\mathbf{X}^{(u)}_{*,J_i}||_{op} = o_P(1)$. By similar arguments, we can show that $S_{3i} + S_{4i}$ are uniformly $o_P(1)$. Hence

$$\min_{i\in[N]} \psi_{\min}(\hat{\Sigma}_{J_i,J_i}) \geq \min_{i\in[N]} \psi_{\min}(S_{1i} + S_{2i}) + o_P(1) \geq \min_{i\in[N]} \psi_{\min}(S_{1i}) + o_P(1),$$

where the second inequality is due to the fact that S_{2i} is positive semi-definite for all i.

(b) The proof is analogous to that of (a) and thus omitted. ∎

Proof of Lemma B.2. The proof is analogous to that of Proposition A.1 in Bai (2009). The major difference lies in the fact that the parameter of interest B^0 is a large dimensional sparse matrix of

dimensions $Np \times N$. Take I_1^* and $\|I_4^*\|_F$ as an example. For I_1^*, we have

$$
\begin{aligned}
\|I_1^*\|_F &= \frac{1}{NT\sqrt{T}} \|\mathbf{U}\Lambda^0 F^{0\prime} \check{F} V_{NT}^{-1}\|_F \\
&\leq \left\| \frac{\mathbf{U}}{\sqrt{NT}} \right\|_{\text{op}} \left\| \frac{\Lambda^0}{\sqrt{N}} \right\|_F \left\| \frac{F^{0\prime}\check{F}}{T} V_{NT}^{-1} \right\|_F = O_p(N^{-1/2} + T^{-1/2}).
\end{aligned}
$$

For I_5^*, we have,

$$
\begin{aligned}
\|I_5^*\|_F &= \frac{1}{NT\sqrt{T}} \|\mathbf{X}(\check{B} - B^0)\Lambda^0 F^{0\prime} \check{F} V_{NT}^{-1}\|_F \\
&\leq \left\| \frac{\mathbf{X}(\check{B} - B^0)}{\sqrt{NT}} \right\|_F \left\| \frac{\Lambda^0}{\sqrt{N}} \right\|_F \left\| \frac{F^{0\prime}\check{F}}{T} V_{NT}^{-1} \right\|_F = O_P(d_{NT})
\end{aligned}
$$

where the last equality follows from (B.4) and we recall that $d_{NT} = N^{-1/2} \|\check{B} - B^0\|_F$. Similarly, we can analyze the other terms to obtain the desired results. ∎

Proof of Lemma B.3. (i) Recall the decomposition $\check{F} - F^0 H = \sqrt{T}(I_1^* + ... + I_8^*)$ in the proof of Proposition B.1. One can write $F^{0\prime}(\check{F} - F^0\check{H})/T = F^{0\prime}(J_1^* + ... + J_8^*)/\sqrt{T}$. For $F^{0\prime} J_1/\sqrt{T}$, we have

$$
\frac{1}{\sqrt{T}}\|F^{0\prime} I_1^*\|_F = \frac{1}{NT^2} \left\| F^{0\prime}\mathbf{U}\Lambda^0 F^{0\prime}\check{F}\check{V}_{NT}^{-1} \right\|_F \leq \frac{1}{\sqrt{NT}} \left\| \frac{F^{0\prime}\mathbf{U}\Lambda^0}{\sqrt{NT}} \right\|_F \left\| \frac{F^{0\prime}\check{F}}{T}\check{V}_{NT}^{-1} \right\|_F = O_P((NT)^{-1/2}),
$$

as we can readily show that $(NT)^{-1/2} \|F^{0\prime}\mathbf{U}\Lambda^0\|_F = O_P(1)$ under Assumptions A.1 and A.4. For $F^{0\prime} I_2^*/\sqrt{T}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|F^{0\prime} I_2^*\|_F &= \frac{1}{NT^2} \left\| F^{0\prime} F^0 \Lambda^{0\prime}\mathbf{U}'\check{F}\check{V}_{NT}^{-1} \right\|_F \leq \frac{1}{\sqrt{NT}} \left\| \frac{F^{0\prime} F^0}{T} \right\|_F \left\| \frac{\Lambda^{0\prime}\mathbf{U}'\check{F}}{\sqrt{NT}} \right\|_F \|\check{V}_{NT}^{-1}\|_F \\
&= O_P((NT)^{-1/2}) \left\| \frac{\Lambda^{0\prime}\mathbf{U}'\check{F}}{\sqrt{NT}} \right\|_F.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\left\| \frac{\Lambda^{0\prime}\mathbf{U}'\check{F}}{\sqrt{NT}} \right\|_F &\leq \left\| \frac{\Lambda^{0\prime}\mathbf{U}'F^0\check{H}}{\sqrt{NT}} + \frac{\Lambda^{0\prime}\mathbf{U}'}{\sqrt{NT}}(\check{F} - F^0\check{H}) \right\|_F \\
&\leq \left\| \frac{\Lambda^{0\prime}\mathbf{U}'F^0}{\sqrt{NT}} \right\|_F \|\check{H}\|_F + \sqrt{NT} \left\| \frac{\Lambda^0}{\sqrt{N}} \right\|_F \left\| \frac{\mathbf{U}}{\sqrt{NT}} \right\|_{\text{op}} \left\| \frac{1}{\sqrt{T}}(\check{F} - F^0\check{H}) \right\|_F \\
&= O_P(1 + (NT)^{1/2}\delta_{NT}^{-1}(\delta_{NT}^{-1} + d_{NT})),
\end{aligned}
$$

where we use the fact that $(NT)^{-1/2} \|\mathbf{U}\|_{\text{op}} = \delta_{NT}^{-1}$ by Assumption A.3 and $T^{-1/2} \|\check{F} - F^0\check{H}\|_F = O_P(\delta_{NT}^{-1} + d_{NT})$ by Lemma B.2. Then $\frac{1}{\sqrt{T}}\|F^{0\prime} I_2^*\|_F = O_P((NT)^{-1/2} + \delta_{NT}^{-1}(\delta_{NT}^{-1} + d_{NT}))$. For

13

$F^{0\prime} I_3^* / \sqrt{T}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \| F^{0\prime} I_3^* \|_{\mathrm{F}} &= \frac{1}{NT^2} \left\| F^{0\prime} \mathbf{U} \mathbf{U}' \check{F} V_{NT}^{-1} \right\|_{\mathrm{F}} \le \frac{1}{\sqrt{T}} \left\| \frac{F^{0\prime} \mathbf{U}}{\sqrt{NT}} \right\|_{\mathrm{F}} \left\| \frac{\mathbf{U}}{\sqrt{NT}} \right\|_{\mathrm{op}} \left\| \frac{\check{F}}{\sqrt{T}} \check{V}_{NT}^{-1} \right\|_{\mathrm{F}} \\
&= T^{-1/2} O_P(\delta_{NT}^{-1}) = O_P((NT)^{-1/2} + T^{-1}).
\end{aligned}
$$

For other terms, we can easily show that they are of the order $O_P(d_{NT})$. Then the conclusion in (i) follows.

(ii) Noting that $\check{F}'\check{F}/T = I_{R^0}$ and using $\check{F} = (\check{F} - F^0 \check{H}) + F^0 \check{H}$, we have

$$
I_{R^0} = \frac{1}{T} \check{F}'(\check{F} - F^0 \check{H}) + \frac{1}{T} (\check{F} - F^0 \check{H})' \check{F} + \frac{1}{T} (\check{F} - F^0 \check{H})'(\check{F} - F^0 \check{H}) + \frac{1}{T} \check{H}' F^{0\prime} F^0 \check{H}.
$$

It follows that $\frac{1}{T} \check{H}' F^{0\prime} F^0 \check{H} = I_{R^0} + O_P((NT)^{-1/2} + T^{-1} + d_{NT})$ by Lemmas B.2(iii) and B.3(i). Then

$$
\frac{1}{T} F^{0\prime} F^0 = (\check{H}')^{-1} \check{H}^{-1} + O_P((NT)^{-1/2} + T^{-1} + d_{NT})
$$

and the desired result follows.

(iii) As in part (i), we decompose $\mathbf{U}'_{*,i}(\check{F} - F^0 \check{H})/T = \mathbf{U}'_{*,i}(I_1^* + ... + I_8^*)/\sqrt{T}$. For $\mathbf{U}'_{*,i} I_1^* / \sqrt{T}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \| \mathbf{U}'_{*,i} I_1^* \|_{\mathrm{F}} &= \frac{1}{NT^2} \left\| \mathbf{U}'_{*,i} \mathbf{U} \Lambda^0 F^{0\prime} \check{F} \check{V}_{NT}^{-1} \right\|_{\mathrm{F}} \le \frac{1}{\sqrt{NT}} \left\| \frac{\mathbf{U}'_{*,i} \mathbf{U} \Lambda^0}{\sqrt{NT}} \right\|_{\mathrm{F}} \left\| \frac{F^{0\prime} \check{F}}{T} \check{V}_{NT}^{-1} \right\|_{\mathrm{F}} \\
&= O_P((NT)^{-1/2} + T^{-1}).
\end{aligned}
$$

For $\mathbf{U}'_{*,i} J_2 / \sqrt{T}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \| \mathbf{U}'_{*,i} J_2 \|_{\mathrm{F}} &= \frac{1}{NT^2} \left\| \mathbf{U}'_{*,i} F^0 \Lambda^{0\prime} \mathbf{U}' \check{F} \check{V}_{NT}^{-1} \right\|_{\mathrm{F}} \le \frac{1}{T\sqrt{N}} \left\| \frac{\mathbf{U}'_{*,i} F^0}{\sqrt{T}} \right\|_{\mathrm{F}} \left\| \frac{\Lambda^{0\prime} \mathbf{U}' \check{F}}{\sqrt{NT}} \right\|_{\mathrm{F}} \| \check{V}_{NT}^{-1} \|_{\mathrm{F}} \\
&= \frac{1}{T\sqrt{N}} O_P(1 + (NT)^{1/2} \delta_{NT}^{-1} (\delta_{NT}^{-1} + d_{NT})) = O_P(N^{-1/2} T^{-1} + T^{-1} d_{NT}).
\end{aligned}
$$

For the other terms, one can use similar analyses to show that they are $O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1} d_{NT})$.

(iv) The proof is similar to that of (iii) and thus omitted. ∎

Proof of Lemma B.4. (i) One can decompose $\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0}$ as follows:

$$
\begin{aligned}
\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0} &= \frac{1}{T}(\check{F} - F^0 \check{H})(\check{F} - F^0 \check{H})' + \frac{1}{T} F^0 \check{H}(\check{F} - F^0 \check{H})' + \frac{1}{T}(\check{F} - F^0 \check{H}) \check{H}' F^{0\prime} \\
&\quad + \frac{F^0}{\sqrt{T}} [\check{H} \check{H}' - (\frac{1}{T} F^{0\prime} F^0)^{-1}] \frac{F^{0\prime}}{\sqrt{T}} \\
&\equiv p_1 + p_2 + p_3 + p_4.
\end{aligned}
$$

14

Then the result follows from Lemmas B.2(iii) and B.3(i)-(ii).

(ii) By the decomposition in (i), we have

$$\frac{1}{T}\mathbf{X}'_{*,J_i}(\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0})\mathbf{U}_{*,i} = \sum_{l=1}^4 T^{-1}\mathbf{X}'_{*,J_i}p_l\mathbf{U}_{*,i} \equiv \sum_{l=1}^4 \tilde{p}_{li}.$$

It is easy to apply Lemma B.3(ii)-(iv) to obtain

$$\|\tilde{p}_{1i}\|_{S_i} \leq \frac{1}{T}\left\|S_i\mathbf{X}'_{*,J_i}(\check{F} - F^0\check{H})\right\|_{\mathrm{F}} \frac{1}{T}\left\|(\check{F} - F^0\check{H})'\mathbf{U}_{*,i}\right\|_{\mathrm{F}} = O_P\left(d_{NT}^2 + \delta_{NT}^{-2}\right),$$

$$\|\tilde{p}_{2i}\|_{S_i} \leq \frac{1}{T}\left\|S_i\mathbf{X}'_{*,J_i}F^0\check{H}\right\|_{\mathrm{F}} \frac{1}{T}\left\|(\check{F} - F^0\check{H})'\mathbf{U}_{*,i}\right\|_{\mathrm{F}} = O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT}),$$

$$\|\tilde{p}_{3i}\|_{S_i} \leq \frac{1}{T}\left\|S_i\mathbf{X}'_{*,J_i}(\check{F} - F^0\check{H})\right\|_{\mathrm{F}} \frac{1}{T}\left\|\check{H}'F^{0\prime}\mathbf{U}_{*,i}\right\|_{\mathrm{F}} = T^{-1/2}O_P(d_{NT} + \delta_{NT}^{-2}),$$

$$\|\tilde{p}_{4i}\|_{S_i} \leq \frac{1}{T}\left\|S_i\mathbf{X}'_{*,J_i}F^0\right\|_{\mathrm{F}} \left\|[\check{H}\check{H}' - (F^{0\prime}F^0/T)^{-1}]\right\|_{\mathrm{F}} \frac{1}{T}\left\|F^{0\prime}\mathbf{U}_{*,i}\right\|_{\mathrm{F}} = T^{-1/2}O_P(d_{NT} + \delta_{NT}^{-2}).$$

Then $\left\|\frac{1}{T}\mathbf{X}'_{*,J_i}(\mathbb{P}_{\check{F}} - \mathbb{P}_{F^0})\mathbf{U}_{*,i}\right\|_{S_i} = O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})$.

(iii) Plugging equation (B.3) into $T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0$, one can obtain that

$$T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(\check{F} - F^0\check{H})\check{H}^{-1}\lambda_i^0 = T^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}(I_1^* + ... + I_8^*)\check{F}(\frac{1}{T}F^{0\prime}\check{F})^{-1}(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0$$

$$\equiv \tilde{I}_{1i} + ... + \tilde{I}_{8i}, \text{ say.}$$

For $\tilde{I}_{1i}$, we have

$$\left\|\tilde{I}_{1i}\right\|_{S_i} = \left\|\frac{1}{NT}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0\right\|_{S_i}$$

$$\leq \frac{1}{NT}\left\|S_i\mathbf{X}'_{*,J_i}\mathbf{U}\Lambda^0\right\|_{\mathrm{F}} \left\|(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0\right\|_{\mathrm{F}}$$

$$+ \frac{1}{\sqrt{NT}}\left\|\frac{1}{T}S_i\mathbf{X}'_{*,J_i}F^0\right\|_{\mathrm{F}} \left\|(\frac{1}{T}F^{0\prime}F^0)^{-1}\right\|_{\mathrm{F}} \left\|\frac{1}{\sqrt{NT}}F^{0\prime}\mathbf{U}\Lambda^0\right\|_{\mathrm{F}} \left\|(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0\right\|_{\mathrm{F}}$$

$$= O_P((NT)^{-1/2}).$$

By the identity $\mathbb{M}_{F^0}F^0 = 0$, we have $\tilde{I}_{2i} = \tilde{I}_{6i} = 0$. It is easy to show that $\left\|\tilde{I}_{3i}\right\|_{S_i} = O_P(\delta_{NT}^{-2} + \delta_{NT}^{-1}d_{NT})$ for $l = 3, 4, 7, 8$. For $\tilde{I}_{5i}$, one have that

$$\tilde{I}_{5i} = -\frac{1}{NT}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}(\check{B} - B^0)\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0.$$

and $\left\|\tilde{I}_{5i}\right\|_{S_i} \leq \left\|S_i\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}\right\|_{\mathrm{F}} \left\|\frac{1}{\sqrt{N}}(\check{B} - B^0)\right\|_{\mathrm{F}} \left\|\frac{1}{\sqrt{N}}\Lambda^0(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0\right\|_{\mathrm{F}} = O_P(d_{NT})$. This implies that $\tilde{I}_{5i}$ is a dominant term in the expansion. Combining the above results yields the desired conclusion. ■

Proof of Lemma B.5. (i) Let $\chi_{ij} \equiv T^{-1}X'_{*,J_i}\mathbb{M}_{F^0}X'_{*,J_j}$. Then $\check{Q}_1 = \text{bdiag}(\chi_{11}, ..., \chi_{NN})$ and

$\check{Q}_2 = \{N^{-1}a_{ij}^0\chi_{ij}\}$ are $NK_a \times NK_a$ matrices, where bdiag($\cdot$) signifies a block diagonal matrix and recall that $K_a = N^{-1}\sum_{i=1}^{N}k_i$. Let $s_1 = [k_1] \equiv \{1, 2, ..., k_1\}$, and $s_{j+1} = \{\sum_{i=1}^{j}k_i + 1, ..., \sum_{i=1}^{j+1}k_i\}$ for $j = 2, ..., N$. Note that

$$\left|\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}\right|_{\infty} = \max_{j\in[N]}\left|\chi_{jj}^{-1}[\check{Q}_2]_{s_j,*}(\check{Q}_1)^{-1}\check{\mathbf{U}}\right|_{\infty}$$

and

$$\chi_{jj}^{-1}[\check{Q}_2]_{s_j,*}(\check{Q}_1)^{-1}\check{\mathbf{U}}$$

$$= \chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}a_{ji}^0\chi_{ji}\chi_{ii}^{-1}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{U}_{*,i}$$

$$= \chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{ji}^0\chi_{ji}\chi_{ii}^{-1}\mathbf{X}_{t,J_i}u_{it} - \chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{ji}^0\chi_{ji}\chi_{ii}^{-1}\frac{1}{T}\mathbf{X}'_{*,J_i}F^0(\frac{1}{T}F^{0\prime}F^0)^{-1}f_t^0u_{it}.$$

Let $e_{i,l}$ denote the lth column of I_{k_i}. Consider the first term on the RHS of the last displayed equation. Note that

$$\chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{ji}^0\chi_{ji}\chi_{ii}^{-1}\mathbf{X}_{t,J_i}u_{it}$$

$$= \chi_{jj}^{-1}\frac{1}{N}\sum_{i=1}^{N}a_{ji}^0\{\chi_{ji}\chi_{ii}^{-1} - E(\chi_{ji})[E(\chi_{ii})]^{-1}\}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{t,J_i}u_{it} + \chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{ji}^0E(\chi_{ji})[E(\chi_{ii})]^{-1}\mathbf{X}_{t,J_i}u_{it}$$

$$= \chi_{jj}^{-1}\frac{1}{N}\sum_{i=1}^{N}a_{ji}^0[\chi_{ji} - E(\chi_{ji})]\chi_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{t,J_i}u_{it}$$

$$+\chi_{jj}^{-1}\frac{1}{N}\sum_{i=1}^{N}a_{ji}^0E(\chi_{jj})[E(\chi_{ii})]^{-1}[E(\chi_{ii}) - \chi_{ii}]\chi_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{t,J_i}u_{it}$$

$$+\chi_{jj}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}a_{ji}^0E(\chi_{ji})[E(\chi_{ii})]^{-1}\mathbf{X}_{t,J_i}u_{it}$$

$$\equiv A_{1j} + A_{2j} + A_{3j},$$

where the second equality follows because

$$\chi_{ji}\chi_{ii}^{-1} - E(\chi_{ji})[E(\chi_{ii})]^{-1} = [\chi_{ji} - E(\chi_{ji})]\chi_{ii}^{-1} + E(\chi_{ji})\{\chi_{ii}^{-1} - [E(\chi_{ii})]^{-1}\}$$

$$= [\chi_{ji} - E(\chi_{ji})]\chi_{ii}^{-1} + E(\chi_{jj})[E(\chi_{ii})]^{-1}[E(\chi_{ii}) - \chi_{ii}]\chi_{ii}^{-1}.$$

Similarly to Lemma A.2, we can show that

$$\sup_{i,j}\left\|\chi_{ij} - E(\chi_{ij})\right\|_{\max} = O_P(T^{-1/2}\log N),$$

16

where elements of $E(\chi_{ij})$ are uniformly bounded and $E(\chi_{ii})$ has minimum eigenvalue bounded away from zero. Noting that $|a'Bb| \leq |a|_1 |b|_1 \|B\|_{\max}$ whenever vectors a and b and matrix B are conformable, we have

$$
\begin{aligned}
\max_{j,l} \left| e'_{j,l} A_{1j} \right| &= \max_j \left| e'_{j,l} \chi_{jj}^{-1} \frac{1}{N} \sum_{i=1}^{N} a_{ji}^0 [\chi_{ji} - E(\chi_{ji})] \chi_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{t,J_i} u_{it} \right| \\
&\leq \max_{j,i} \|\chi_{ji} - E(\chi_{ji})\|_{\max} |a_{ji}^0| \frac{1}{N} \sum_{i=1}^{N} \left| e'_{j,l} \chi_{jj}^{-1} \right|_1 \left| \chi_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{t,J_i} u_{it} \right|_1 \\
&= O_P(T^{-1/2} \log N) O(K_J^{1/2}) O_P(K_J^{1/2} K_a^{1/2} T^{-1/2}) = O_P(K_J K_a^{1/2} T^{-1} (\log N)^2),
\end{aligned}
$$

where we use the fact that $\left| e'_{j,l} \chi_{jj}^{-1} \right|_1 \leq \sqrt{k_j} \left| e'_{j,l} \chi_{jj}^{-1} \right|_2 \leq \sqrt{k_j} [\psi_{\min}(\chi_{jj})]^{-1}$ and

$$
\left| \chi_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{t,J_i} u_{it} \right|_1 \leq \sqrt{k_j} \left| \chi_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{t,J_i} u_{it} \right|_2 = \sqrt{k_j k_i} O_P(T^{-1/2} \log N).
$$

By the same token, $\max_j \left| e'_{j,l} A_{2j} \right| = O_P(K_J K_a^{1/2} T^{-1} (\log N)^2)$. In addition, we can show that

$$
\begin{aligned}
\max_{j,l} \left| e'_{j,l} A_{3j} \right| &= \max_{j,l} \left| e'_{j,l} \chi_{jj}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{ji}^0 E(\chi_{ji}) [E(\chi_{ii})]^{-1} \mathbf{X}_{t,J_i} u_{it} \right| \\
&\leq \max_{j,l} \left\| e'_{j,l} \chi_{jj}^{-1} \right\|_{\max} \max_j \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{ji}^0 E(\chi_{ji}) [E(\chi_{ii})]^{-1} \mathbf{X}_{t,J_i} u_{it} \right|_1 \\
&= O_P(K_J K_a^{1/2} (NT)^{-1/2} \log N).
\end{aligned}
$$

It follows that $\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{ij}^0 \chi_{jj}^{-1} \chi_{ji} \mathbf{X}_{t,J_i} u_{it} \right\|_{S_j} = O_P(K_J K_a^{1/2} [T^{-1} (\log N)^2 + (NT)^{-1/2} \log N]) = O_P(K_J K_a^{1/2} [T^{-1} (\log N)^2 + N^{-1}])$.

Similarly, we can show that

$$
\max_j \left| \chi_{jj}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{ji}^0 \chi_{ji} \chi_{ii}^{-1} \frac{1}{T} \mathbf{X}'_{*,J_i} F^0 \left(\frac{1}{T} F^{0\prime} F^0 \right)^{-1} f_t^0 u_{it} \right|_\infty = O_P(K_J K_a^{1/2} [T^{-1} (\log N)^2 + N^{-1}]).
$$

Then $\left| \check{Q}_1^{-1} \check{Q}_2 \check{Q}_1^{-1} \check{\mathbf{U}} \right|_\infty = O_P(K_J K_a^{1/2} [T^{-1} (\log N)^2 + N^{-1}])$.

(ii) The proofs follows from that of (i) closely. Noting that

$$
\frac{1}{\sqrt{N}} \left| \Gamma \check{Q}_1^{-1} \check{Q}_2 \check{Q}_1^{-1} \check{\mathbf{U}} \right| \leq \|\Gamma\|_{\mathrm{op}} \frac{1}{\sqrt{N}} \left| \check{Q}_1^{-1} \check{Q}_2 \check{Q}_1^{-1} \check{\mathbf{U}} \right|,
$$

it suffices to show that $\frac{1}{\sqrt{N}} \left| \check{Q}_1^{-1} \check{Q}_2 \check{Q}_1^{-1} \check{\mathbf{U}} \right| = O_P(K_J K_a [T^{-1} (\log N)^2 + N^{-1}]) = o_P(T^{-1/2})$. The result

follows from (i) under Assumption A.5(ii). To see this, notice that

$$\frac{1}{N}\left|\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}\right|^2 = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbb{S}_i\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}\right|^2 \le \frac{1}{N}\sum_{i=1}^{N}k_i\left|\check{Q}_1^{-1}\check{Q}_2\check{Q}_1^{-1}\check{\mathbf{U}}\right|_\infty^2$$

$$= K_a O_P(K_J^2 K_a[T^{-1}(\log N)^2 + N^{-1}]^2) = o_P(T^{-1}),$$

where $\mathbb{S}_i$ is defined in the proof of Proposition B.1. A more complicated argument can relax the restriction on K_J and K_a slightly, but we do not pursue it for brevity.

(iii) Note that $\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}$ has the (i,j)th block given by $\frac{a_{ij}^0}{N}W_{ij}$, where

$$W_{ij} \equiv \left(\frac{1}{T}\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}\mathbf{X}_{*,J_i}\right)^{-1/2}\frac{\mathbf{X}'_{*,J_i}\mathbb{M}_{F^0}}{\sqrt{T}}\frac{\mathbb{M}_{F^0}\mathbf{X}_{*,J_j}}{\sqrt{T}}\left(\frac{1}{T}\mathbf{X}'_{*,J_j}\mathbb{M}_{F^0}\mathbf{X}_{*,J_j}\right)^{-1/2}.$$

Obviously, we have that $\psi_{\max}(W_{ij}) \le 1$. In addition, it is easy to see that the inequality does not bind for all pairs of (i,j)'s w.p.a.1. For any $V \in \mathbb{R}^{\Sigma_{i=1}^N k_i}$ and $|V| = 1$, we can decompose it to $V = (V_1',...,V_N')'$, where $V_i \in \mathbb{R}^{k_i}$. Let $\tilde{W} \in \mathbb{R}^{N \times N}$, with $\tilde{W}_{ij} = V_i'W_{ij}V_j$. Then we have

$$V'\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}V = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{a_{ij}^0}{N}V_i'W_{ij}V_j = \mathrm{tr}\left(P_{\Lambda^0}\tilde{W}\right) \le ||P_{\Lambda^0}||_{\mathrm{op}}||\tilde{W}||_*$$

$$= \mathrm{tr}(\tilde{W}) = \sum_{i=1}^{N}V_i'W_{ii}V_i \le \sum_{i=1}^{N}|V_i|^2 = |V|^2 = 1.$$

The equality holds in all places only if the columns of $\tilde{W}$ are linear combinations of Λ^0. We can show that the inequality does not bind w.p.a.1. That is, $\psi_{\max}(\check{Q}_1^{-1/2}\check{Q}_2\check{Q}_1^{-1/2}) < 1$ w.a.p.1. This completes the proof. ∎

C   Some Technical Lemmas

In this section we introduce the Nagaev inequality established by Wu and Wu (2016) and then prove some additional technical lemmas used in the proofs in Section B.

C.1   Nagaev inequality for time series

In Theorem C.1 below, we aim to bound the partial sum of the form $S_n = \sum_{i=1}^{n}a_ie_i$, where $a_i \in \mathbb{R}$ are nonrandom, the scalar process $\{e_i\}$ has the form $e_i = g(...,\varepsilon_{i-1},\varepsilon_i)$, where ε_i is independently and identically distributed (i.i.d.) random variables, and $g(\cdot)$ is a measurable function. Letting $\mathcal{F}_i \equiv (...,\varepsilon_{i-1},\varepsilon_i)$, we write $e_i = g(\mathcal{F}_i)$. Then a coupled process e_i^* can be defined as $e_i^* = g(\mathcal{F}_i^*)$, where $\mathcal{F}_i^* = (...,\varepsilon_{-1},\varepsilon_0^*,\varepsilon_1,...,\varepsilon_{i-1},\varepsilon_i)$ and ε_0^* is an independent copy of ε_0. Recall that $|||\cdot|||_q \equiv (E|\cdot|_q^q)^{1/q} < \infty$. Assuming that $|||e_i|||_q < \infty$ for some $q \ge 1$, we define the functional dependence

measure:

$$\delta_{i,q}(e.) \equiv |||e_i - e_i^*|||_q = |||g(\mathcal{F}_i) - g(\mathcal{F}_i^*)|||_q,$$

where $e_i^* = g(\mathcal{F}_i^*)$. The measure $\delta_{i,q}(e.)$ reflects the effect of shock ε_0 on e_i. Accordingly, we assume the cumulative effect of ε_0 on $\{e_i\}_{i \geq m}$ to be summable:

$$\Delta_{m,q}(e.) \equiv \sum_{i=m}^{\infty} \delta_{i,q}(e.) < \infty.$$

As in Wu and Wu (2016) we define the *dependence-adjusted norm* (DAN):

$$||e.||_{q,\alpha} = \sup_{m \geq 0}(m+1)^{\alpha}\Delta_{m,q}(e.).$$

With these definitions, we can present the following Nagaev inequality for time series as a simplified version of Theorem 2 of Wu and Wu (2016).

Theorem C.1 *Let $a = (a_1, ..., a_n)'$ and $|a|_q = (\sum_{i=1}^{n}|a_i|^q)^{1/q}$. Suppose that $\sum_{i=1}^{n}a_i^2 = n$, $E(e_i) = 0$, and $||e.||_{q,\alpha} < \infty$ for some $q > 2$ and $\alpha > 1$. Then for all $x > 0$,*

$$P(|S_n| > x) \leq C_1\frac{|a|_q^q||e.||_{q,\alpha}^q}{x^q} + C_2 exp\left(-\frac{C_3 x^2}{n||e.||_{2,\alpha}^2}\right),$$

where C_1, C_2, C_3 are constants that only depend of q and α.

C.2 Some additional lemmas

Following the decomposition (2.5) in Section 2, we have $y_{it} = y_{it}^{(f)} + y_{it}^{(u)}$, where

$$y_{it}^{(u)} = \sum_{j=0}^{\infty}\alpha_{iN}^{(u)}(j)u_{t-j} = \sum_{j=0}^{\infty}\alpha_{iN}^{(u)}(j)C^{(u)}\epsilon_{t-j}^{(u)} \equiv \sum_{j=0}^{\infty}C_j^{(i,u)}\epsilon_{t-j}^{(u)},$$

$$y_{it}^{(f)} = E(y_{it}) + \sum_{j=0}^{\infty}\alpha_{iN}^{(f)}(j)f_{t-j}^0 \equiv E(y_{it}) + \sum_{j=0}^{\infty}C_j^{(i,f)}\epsilon_{t-j}^{(f)}, \text{ where } C_j^{(i,f)} \equiv \sum_{k=0}^{j}\alpha_{iN}^{(f)}(j-k)C_k^{(f)}.$$

Let $f_{r,t}^0$ be the rth entry of f_t^0. Lemma C.1 below establishes the DAN's for time series $f_{r,\cdot}^0$, $y_{i,\cdot}^{(f)}$, $y_{i\cdot}^{(u)}$ and $y_{i\cdot}$.

Lemma C.1. Suppose that Assumption A.1 holds and $q > 1$. There is a constant $\bar{c} < \infty$ such that the following statements hold:

(i) $||f_{r,\cdot}^0||_{q,\alpha} < \bar{c}|||\epsilon_0^{(f)}|||_q$ for $r = 1, ..., R^0$;

(ii) $\max_{1 \leq i \leq N}||y_{i,\cdot}^{(f)}||_{q,\alpha} \leq \bar{c}R^0|||\epsilon_0^{(f)}|||_q$;

(iii) $\max_{1 \leq i \leq N}||y_{i,\cdot}^{(u)}||_{q,\alpha} \leq \bar{c}\sqrt{||\Sigma_u||_{op}}|||\epsilon_{1,0}^{(u)}|||_q$;

(iv) $\max_{1 \leq i \leq N}||y_{i,\cdot}||_{q,\alpha} < \bar{c}$.

Proof of Lemma C.1. (i) Let $p = q/(q-1)$ where $q > 1$. By the Hölder inequality, we have $\delta_{t,q}(f_{r\cdot}^0) \leq 2|[C_t^{(f)}]_{r,*}|_p \cdot |||\epsilon_0^{(f)}|||_q \leq 2(R^0)^{1/p}|[C_t^{(f)}]_{r,*}|_\infty \cdot |||\epsilon_0^{(f)}|||_q$. It follows that

$$\Delta_{m,q}(f_\cdot) \leq 2|||\epsilon_0^{(f)}|||_q \sum_{t=m}^{\infty} |[C_t^{(f)}]_{r,*}|_p \leq \bar{c}|||\epsilon_0^{(f)}|||_q(m+1)^{-\alpha},$$

where the last inequality holds by Assumption A.1(ii). The desired result follows immediately.

(ii) Noting that $y_{it}^{(f)}$ is a linear process, we can directly calculate that

$$\delta_{t,q}(y_{i\cdot}^{(f)}) \leq 2|||C_t^{(i,f)}\epsilon_0^{(f)}|||_q \leq 2|C_t^{(i,f)}|_p \cdot |||\epsilon_0^{(f)}|||_q \text{ and } \Delta_{m,q}(y_{i\cdot}^{(f)}) \leq 2|||\epsilon_0^{(f)}|||_q \sum_{t=m}^{\infty} |C_t^{(i,f)}|_p.$$

It suffices to bound $\sum_{t=m}^{\infty} ||C_t^{(i,f)}||_p$. Noting that $C_k^{(i,f)} = \sum_{j=0}^{k} \alpha_{iN}^{(f)}(j)C_{k-j}^{(f)}$, we have

$$
\begin{aligned}
\sum_{t=m}^{\infty} |C_t^{(i,f)}|_p &\leq (R^0)^{1/p} \sum_{t=m}^{\infty} |C_t^{(i,f)}|_\infty \leq (R^0)^{1/p} \sum_{t=m}^{\infty} \sum_{j=0}^{t} |\alpha_{iN}^{(f)}(j)| \cdot ||C_{t-j}^{(f)}||_\mathrm{F} \\
&\leq \bar{c} \sum_{t=m}^{\infty} \sum_{j=0}^{t} \rho^j ||C_{t-j}^{(f)}||_\mathrm{F} = \bar{c} \sum_{j=0}^{\infty} \rho^j \sum_{t=(m-j)\vee 0}^{\infty} ||C_t^{(f)}||_\mathrm{F} \\
&= \bar{c} \left[\sum_{j=m}^{\infty} \rho^j \sum_{t=0}^{\infty} ||C_t^{(f)}||_\mathrm{F} + \sum_{j=0}^{m-1} \rho^j \sum_{t=(m-j)}^{\infty} ||C_t^{(f)}||_\mathrm{F} \right] \\
&\leq \bar{c} \left[\sum_{t=0}^{\infty} ||C_t^{(f)}||_\mathrm{F} \frac{\rho^m}{1-\rho} + R^0 \sum_{j=0}^{m-1} \rho^j (m-j)^{-\alpha} \right],
\end{aligned}
$$

where the second inequality is by Assumption A.1(vi), and the last inequality is by Assumption A.1(ii). To show $\sup_{m\geq 1}(m+1)^\alpha \Delta_{m,q}(y_{i\cdot}^{(f)}) \leq \bar{c}$ for some $\bar{c} < \infty$, we need to show $\sup_{m\geq 1}(m+1)^\alpha \sum_{j=0}^{m-1} \rho^j(m-j)^{-\alpha} \leq \bar{c}'$ for some $\bar{c}' < \infty$. The last result follows because

$$
\begin{aligned}
(m+1)^\alpha \sum_{j=0}^{m-1} \rho^j(m-j)^{-\alpha} &= (m+1)^\alpha \left(\sum_{j=0}^{\lfloor\sqrt{m}\rfloor} + \sum_{j=\lfloor\sqrt{m}\rfloor+1}^{m-1} \right) \rho^j(m-j)^{-\alpha} \\
&\leq \frac{(m+1)^\alpha}{(1-\rho)(m-\lfloor\sqrt{m}\rfloor)^\alpha} + (m+1)^\alpha \rho^{\lfloor\sqrt{m}\rfloor+1}/(1-\rho) \\
&\to \frac{1}{1-\rho} \text{ as } m \to \infty,
\end{aligned}
$$

where $\lfloor\sqrt{m}\rfloor$ is the largest integer that is not greater than $\sqrt{m}$. It follows that

$$||y_{i,\cdot}^{(f)}||_{q,\alpha} = \sup_{m\geq 0}(m+1)^\alpha \Delta_{m,q}(y_{i,\cdot}^{(f)}) < \bar{c}'|||\epsilon_0^{(f)}|||_q,$$

for some $\bar{c}' < \infty$.

(iii) Note that $y_{it}^{(u)}$ is a linear function of $(..., \epsilon_{t-1}^{(u)}, \epsilon_t^{(u)})$ with $\epsilon_t^{(u)} \in \mathbb{R}^m$. Given $||\alpha_{iN}^{(u)}(j)|| < \bar{c}\rho^j$ by Assumption A.1(vii), we can see that $|\alpha_{iN}^{(u)}(j)C^{(u)}| \leq \bar{c}\rho^j \sqrt{\psi_1(\Sigma_u)}$. Let $[\alpha_{iN}^{(u)}(t)C^{(u)}]_j$ denote the jth component of $\alpha_{iN}^{(u)}(j)C^{(u)}$. Then we can calculate $\delta_{t,q}(y_{i\cdot}^{(u)})$:

$$
\begin{aligned}
\delta_{t,q}(y_{i\cdot}^{(u)}) &\leq 2\left|\left|\left| \alpha_{iN}^{(u)}(t)C^{(u)}\epsilon_0^{(u)} \right|\right|\right|_q \\
&\leq 2\left|\left|\left| \left[\sum_{j=1}^{m} ([\alpha_{iN}^{(u)}(t)C^{(u)}]_j)^2 (\epsilon_{j,0}^{(u)})^2 \right]^{1/2} \right|\right|\right|_q \\
&= 2\left(\left|\left|\left| \sum_{j=1}^{m} ([\alpha_{iN}^{(u)}(t)C^{(u)}]_j)^2 (\epsilon_{j,0}^{(u)})^2 \right|\right|\right|_{q/2} \right)^{1/2} \\
&\leq 2\left(\sum_{j=1}^{m} ([\alpha_{iN}^{(u)}(t)C^{(u)}]_j)^2 |||(\epsilon_{j,0}^{(u)})^2|||_{q/2} \right)^{1/2} \\
&= 2|||\epsilon_{1,0}^{(u)}|||_q |\alpha_{iN}^{(u)}(t)C^{(u)}| \leq 2\rho^t \bar{c}\sqrt{\psi_{\max}(\Sigma_u)}|||\epsilon_{1,0}^{(u)}|||_q,
\end{aligned}
$$

where the second inequality holds by Burkholder inequality (see, e.g., Hall, 1980, p. 23). Then we have

$$
\Delta_{m,q}(y_{i,\cdot}^{(u)}) \leq 2\bar{c}\sqrt{\psi_{\max}(\Sigma_u)}|||\epsilon_{1,0}^{(u)}|||_q \frac{\rho^m}{1-\rho},
$$

and it follows that

$$
||y_{i,\cdot}^{(u)}||_{q,\alpha} \leq \bar{c}'\sqrt{\psi_{\max}(\Sigma_u)}|||\epsilon_{1,0}^{(u)}|||_q < \infty.
$$

(iv) The result follows (ii) and (iii). ∎

The following lemma bounds the DAN of summation of product of two linear processes:

Lemma C.2. Consider two time series $e_t = g(..., \varepsilon_{t-1}, \varepsilon_t)$ and $x_t = h(..., \varepsilon_{t-1}, \varepsilon_t))$. Suppose that $||x.||_{\iota,\alpha_X} < \infty$ and $||e.||_{q,\alpha_e} < \infty$ with $q > 2, \iota > 4$ and $\alpha_x, \alpha_e > 0$. Consider the time series $x.e. = \{x_t e_t\}$. Then

$$
||x.e.||_{\tau,\alpha} \leq 2||x.||_{\iota,\alpha_X}||e.||_{q,\alpha_e}
$$

for $\alpha = \alpha_X \wedge \alpha_e$ and $\tau = q\iota/(q+\iota)$.

Proof of Lemma C.2. We have that

$$
\begin{aligned}
\Delta_{m,\tau}(x.e.) &= \sum_{t=m}^{\infty} \delta_{t,\tau}(x.e.) = \sum_{t=m}^{\infty} |||x_t e_t - x_t^* e_t^*|||_\tau \\
&\leq \sum_{t=m}^{\infty} (|||x_t(e_t - e_t^*)|||_\tau + |||(x_t - x_t^*)e_t^*|||_\tau) \\
&\leq \sum_{t=m}^{\infty} (|||x_t|||_\iota |||e_t - e_t^*|||_q + |||x_t - x_t^*|||_\iota |||e_t^*|||_q) \\
&\leq \max_t |||x_t|||_\iota \Delta_{m,q}(e.) + \max_t |||e_t|||_q \Delta_{m,\iota}(x.).
\end{aligned}
$$

It follows that

$$
||x.e.||_{\tau,\alpha} \leq \max_t |||x_t|||_\iota ||e.||_{q,\alpha_e} + \max_t |||e_t|||_q ||x.||_{\iota,\alpha_X} \leq 2||x.||_{\iota,\alpha_X} ||e.||_{q,\alpha_e},
$$

where we used the fact that $\max_t |||x_t|||_\iota \leq ||x.||_{\iota,\alpha_X}$. ∎

Lemma C.3. Suppose that Assumption A.1 holds. For $z_{1,t}$, $z_{2,t} = 1$, $f_{r,t}^0$, $y_{i,t-l_1}$, $y_{i,t-l_2}^{(u)}$, $y_{i,t-l_3}^{(u)}$ and u_{it}, with $i = 1, ..., N$, $l_1, l_2, l_3 = 1, ..., p$, we have

$$
P\left(\left| \sum_{t=1}^{T} z_{1,t} z_{2,t} - E(z_{1,t} z_{2,t}) \right| \geq x \right) \leq C_1 \frac{T}{x^{q/2}} + C_2 \exp\left(-\frac{C_3 x^2}{T} \right),
$$

where C_1, C_2, and C_3 are constants that do not depend on (N, T) and $(z_{1,t}, z_{2,t})$.

Proof of Lemma C.3. We apply the Nagaev inequality in Theorem C.1 to prove the claim. By Lemma C.1, we have $||y_{i,\cdot}||_{q,\alpha} < \bar{c}$ and $||u_{i,\cdot}||_{q,\alpha} < \bar{c}$ for some constant $\bar{c} < \infty$. By Lemma C.2, we can obtain that $||z_{1,t} z_{2,t} - E(z_{1,t} z_{2,t})||_{q/2,\alpha} \leq 2 ||z_{1,\cdot}||_{q,\alpha} ||z_{2,\cdot}||_{q,\alpha} \leq 2\bar{c}^2$. By Theorem C.1, we have

$$
\begin{aligned}
P\left(\left| \sum_{t=1}^{T} z_{1,t} z_{2,t} - E(z_{1,t} z_{2,t}) \right| \geq x \right) &\leq C_1 \frac{|a|_{q/2}^{q/2} ||z_{1,t} z_{2,t} - E(z_{1,t} z_{2,t})||_{q/2,\alpha}^{q/2}}{x^{q/2}} \\
&\quad + C_2 \exp\left(-\frac{C_3 x^2}{T ||z_{1,t} z_{2,t} - E(z_{1,t} z_{2,t})||_{q/2,\alpha}^2} \right),
\end{aligned}
$$

where $a = \iota_T$ such that $|a|_{q/2}^{q/2} = T$. The desired result is proved. ∎

Lemma C.4. Suppose that Assumption A.1 holds and $N^2 T^{1-q/4}(\log N)^{-q/2} + N^{2-\underline{c} \log N} \to 0$. Then
  (i) There is an absolute constant $\underline{c}'$ such that $P\left[\min_{1 \leq i \leq N}(\psi_{\min}(\hat{\Sigma}_{J_i, J_i}^{(u)})) > \underline{c}' \right] \to 1$;
  (ii) $\left\| \max_{1 \leq i \leq N}(T^{-1} \mathbf{X}_{*,J_i}^{(f)\prime} \mathbb{M}_{F^0} \mathbf{X}_{*,J_i}^{(u)}) \right\|_{\text{op}} \to 0$.

Proof of Lemma C.4. (i) Note that

$$
\hat{\Sigma}_{J_i, J_i}^{(u)} = \frac{1}{T} \mathbf{X}_{*,J_i}^{(u)\prime} \mathbf{X}_{*,J_i}^{(u)} - \frac{1}{T} \mathbf{X}_{*,J_i}^{(u)\prime} \mathbb{P}_{F^0} \mathbf{X}_{*,J_i}^{(u)}.
$$

The first term converges in probability to $\Sigma^{(u)}_{J_j,J_j} \equiv E[(X^{(u)}_t)_{J_i}(X^{(u)}_t)'_{J_i}]$, where $\psi_{\min}(\Sigma^{(u)}_{J_j,J_j}) \geq \psi_{\min}(\Sigma^{(u)}_X)$. The second term converges in probability to zero as $X^{(u)}_t$ is zero mean and independent of F^0. By using the results of Lemmas A.2-A.4, we can establish the uniform result in (i).

(ii) The proof is analogous to that of (i) and thus omitted. ∎

D Discussion on Assumption A.1(vi)

In this section, we first give a discussion on the operator norm of Φ. Then we give a sufficient condition for Assumption A.1(vi).

It is well known that requiring the eigenvalues of Φ to be in the unit circle can ensure the stationarity of the process Y_t. However, this condition does not ensure $||\Phi||_{\mathrm{op}} \leq 1$. For instance, consider $p = 1$ and the following $N \times N$ transition matrix

$$\Phi = A^0_1 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

for which all eigenvalues are zero but $||\Phi||_{\mathrm{op}} = N - 1$. Basu and Michailidis (2005) also show that $||\Phi||_{\mathrm{op}} \geq 1$ as long as $p > 1$.

However, as we require that the spectral radius of Φ is bounded by $\rho < 1$, we know that $\left\|\Phi^j\right\|_{\mathrm{op}} \rightarrow 0$ as $j \rightarrow \infty$. To see this, one can resort to the Jordan canonical form (e.g., page 656 of Lükepohl 2005). As $(\Phi^j)_{[N],[N]}$ is a principal submatrix of Φ, imposing the high level condition $\psi_{\max}[(\Phi^j)_{[N],[N]}] \leq \bar{c}\rho^j$ with a large enough $\bar{c}$ is reasonable. In the special case that $p = 1$ and A^0_1 is a block diagonal matrix with bounded block size, we can easily show that the Assumption A.1(vi) is satisfied.

References

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica, 77, 1229-1279.

Basu, S. and Michailidis, G., 2015. Regularized estimation in sparse high-dimensional time series models. Annals of Statistics 43, 1535-1567.

Chernozhukov, V., Hansen, C., Liao, Y. and Zhu, Y., 2018. Inference For Heterogeneous Effects Using Low-Rank Estimations. arXiv preprint arXiv:1812.08089.

Hall, P. and Heyde, C.C., 2014. Martingale limit theory and its application. Academic press.

Lükepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media, New York.

van de Geer, S.A. and Bühlmann, P., 2009. On the conditions used to prove oracle results for the Lasso. Electronic Journal of Statistics, 3, 1360-1392.

Wu, W.B. and Wu, Y.N., 2016. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. Electronic Journal of Statistics 10, 352-379.