Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(3): 5575–5599.
DOI: 10.3934/math.2026230
Received: 27 December 2025
Revised: 12 February 2026
Accepted: 26 February 2026
Published: 05 March 2026

Research article

A time-varying latent factor model with GARCH noise for high-dimensional covariance matrix estimation

Yuwen Ruan[1], **Xingfa Zhang**[1], **Yujiao Liu**[1], **Yan Wang**[2], **and Tianli Lei**[3,*]

[1] School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China
[2] Business School, Asia Pacific University of Technology and Innovation, Kuala Lumpur 57000, Malaysia
[3] Institute of Applied Mathematics, Shenzhen Polytechnic University, Shenzhen 518051, China

* **Correspondence:** Email: 364407945@qq.com.

Abstract: In this paper, we established a time-varying latent factor model with GARCH (generalized autoregressive conditional heteroskedasticity) noise to study volatilities (conditional covariance matrix) under the high-dimensional framework, when factors are unobservable, factor loadings are time-varying, and the idiosyncratic error term shows heteroskedasticity. A projection method was proposed for more-precise estimation of the conditional covariance matrix. We demonstrated that our model is robust when the dimensionality and sample size are large. Asymptotic theories were developed for the proposed estimation. A simulation study was conducted to evaluate the performance of the proposed model, and a real example was provided to illustrate this approach.

Keywords: time-varying factor model; multivariate GARCH; high-dimensionality; conditional covariance matrix
Mathematics Subject Classification: 62F12, 62G05, 62M10

1. Introduction

Over the past several decades, the study of time series volatilities has been an important and popular topic in both statistics and economics, especially in high-dimensional frameworks, due to the involvement of a large number of assets in the modern financial market and large data dimensionality. However, when dealing with complex data, an unavoidable issue arises, known as the "curse of dimensionality", which leads to overparameterization and identifiability problems. To overcome these difficulties, many researchers have developed various methods to reduce dimensionality in time series analysis. Among them, the factor model is a general approach, and the related theories have been well-developed. See, for example, Connor and Linton [1], Hafner and Preminger [2], and Onatski [3].

These factor models have been widely used in various fields, but their effectiveness is limited for their restrictive assumption that factor loadings are constant over a long time period. In fact, there exist various driving forces such as institutional switching, economic transition, and technological progress that may influence the relationship between factors and observed variables significantly. Hence, some scholars, such as Motta et al. [4], Su and Wang [5], and Chen et al. [6], have explored factor models where factor loadings change over time. Building on these studies, Barigozzi [7] proposed a new time-varying generalized dynamic factor model for high-dimensional locally stationary time series, and Chen et al. [8] developed a quantile regression approach to consistently estimate all quantile-relevant factors and loadings. Chen et al. [9] introduced a screening estimation method for a semiparametric conditional factor model with latent factors. Wang et al. [10] proposed a time-varying panel data model with interactive fixed effects and established the bias-corrected coefficient estimator, as well as the limiting distributions and uniform consistency of the estimated factors and factor loadings. Tu [11] considered a time-varying coefficient quantile predictive regression model with factor-augmented predictors to capture smooth structural changes and incorporate information from high-dimensional data for forecasting.

Besides, the impact of the "curse of dimensionality" on time series volatility analysis and modeling goes beyond this. In the high-dimensional framework, a large amount of noise may be incorrectly incorporated into the model, and its cumulative effect can increase the residual variance, leading to heteroskedasticity, which can severely impact model fitting and the estimation of conditional covariance matrices (so-called volatilities). In other words, it is essential to consider the accumulated noises and the heteroscedasticity exhibited in the data. To this end, some scholars proposed the GARCH model to estimate the covariance matrices, which can be traced back to the multivariate GARCH model proposed by Engle and Kroner [12]. Building on this, Bollerslev [13] introduced the constant conditional correlation GARCH (CCC-GARCH) model, and Engle [14] proposed the dynamic conditional correlation GARCH (DCC-GARCH) model. Further, Guo et al. [15] proposed a dynamic structure and an estimation procedure for the high-dimensional conditional covariance matrix. Wang et al. [16] employed the ADCC-GARCH approach to investigate the dynamic correlation between different financial assets. Li et al. [17] and Liu et al. [18] combined the ideas of the factor model and GARCH model, which performs relatively well when factors are observable and the data has heteroskedastic effects. Laksaci et al. [19] developed a kernel-based nonparametric estimator for multifunctional expectiles in high-frequency functional financial data and established its consistency to improve GARCH-based risk assessment for multi-asset risk management. Alraddadi [20] proposed an improved Markov-switching threshold bilinear-GARCH model to capture the complex behavior of financial time series.

While the aforementioned literature have made significant progress in addressing the time-varying nature of factors and heteroskedasticity separately, how to simultaneously model the smooth temporal variation of factor loadings and the conditional heteroskedasticity of idiosyncratic components within a unified, high-dimensional framework where factors are unobservable remains an under-explored challenge. Hence, we develop an alternative model for estimating the covariance matrix within a high-dimensional framework where both factor loadings and factors are unobservable, factor loadings vary over time, and the idiosyncratic error term exhibits heteroskedasticity. The core idea is to employ a projection technique to construct a time-varying factor model. For the covariance matrix, we use a GARCH model to fit the idiosyncratic components, thereby achieving our objective. Simulation and empirical studies in the subsequent sections demonstrate that the proposed model offers good robustness and superior performance.

This work is related to our prior research, Ruan et al. [21], which proposed a semi-parametric factor-GARCH framework for high-dimensional covariance matrix estimation involving latent factors and heteroskedastic idiosyncratic components. However, the present study introduces a significant advancement by explicitly modeling time-varying factor loadings. While Ruan et al. [21] established a robust approach for handling unobserved factors and volatility clustering, it retained the conventional assumption of constant loadings over time. This assumption may be overly restrictive for capturing evolving market dynamics, which can fundamentally alter the relationship between assets and common risk factors. Consequently, this work distinguishes itself from our prior research by focusing on capturing smooth temporal variation in factor exposures, providing a more precise tool for volatility estimation and portfolio risk management in high-dimensional settings.

The rest of this article is organized as follows. Section 2 introduces our projection technique and the latent factor model. Section 3 provides the analysis for asymptotic properties of the proposed estimators, including the necessary assumptions. Section 4 constructs a portfolio allocation based on the formula for the Markowitz optimal portfolio using the proposed model. Section 5 presents numerical and empirical results. Section 6 concludes the article. All the proofs are given in the Appendix.

Throughout the paper, we use $\mathbf{A}^{\mathcal{T}}$, $\lambda_{\min}(\mathbf{A})$, and $\lambda_{\max}(\mathbf{A})$ to respectively denote the transpose, minimum eigenvalue, and maximum eigenvalue of a matrix $\mathbf{A}$. Let $\mathrm{vec}(\mathbf{A})$ be the column-stacking vector of a matrix $\mathbf{A}$. For a vector $\mathbf{v}$, let $\|\mathbf{v}\|$ denote its Euclidean norm. The notations $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$, and $\|\mathbf{A}\|_{\max}$ represent the Frobenius norm, spectral norm (operator norm), and element-wise norm of a matrix $\mathbf{A}$, defined respectively by

$$\|\mathbf{A}\|_F = \mathrm{tr}^{1/2}(\mathbf{A}^{\mathcal{T}}\mathbf{A}), \quad \|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}^{\mathcal{T}}\mathbf{A}), \quad \text{and} \quad \|\mathbf{A}\|_{\max} = \max_{i,j}|a_{ij}|. \tag{1.1}$$

2. The model and methodology

2.1. Model

The factor model is a statistical model used to explain the relationships among multiple observed variables influenced by some common factors. Assume that the observed data is given by y_{it}, where $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. The definition of a time-varying factor model is provided as follows:

$$y_{it} = \boldsymbol{\lambda}_{it}^{\mathcal{T}}\mathbf{f}_t + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{2.1}$$

where $\mathbf{f}_t = (f_{t1}, f_{t2}, \dots, f_{tR})^{\mathcal{T}}$ is the factor vector, $\boldsymbol{\lambda}_{it} = (\lambda_{it,1}, \lambda_{it,2}, \dots, \lambda_{it,R})^{\mathcal{T}}$ is the factor loading vector corresponding to variable i, and u_{it} is the idiosyncratic component of y_{it}. By stacking (2.1) over i for each t, we have

$$\mathbf{y}_t = \boldsymbol{\Lambda}_t\mathbf{f}_t + \mathbf{u}_t, \tag{2.2}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})^{\mathcal{T}}$ is the data vector, $\boldsymbol{\Lambda}_t = (\boldsymbol{\lambda}_{1t}, \boldsymbol{\lambda}_{2t}, \dots, \boldsymbol{\lambda}_{Nt})^{\mathcal{T}}$ is the $N \times R$ factor loading matrix, and $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})^{\mathcal{T}}$ is the vector of idiosyncratic errors of $\mathbf{y}_t$.

Since we consider the idiosyncratic error term to exhibit heteroskedasticity, based on model (2.2), we assume that each component of $\mathbf{u}_t$ follows a GARCH structure such that

$$E(\mathbf{u}_t|\mathcal{F}_{t-1}) = 0, \quad \mathrm{cov}(\mathbf{u}_t|\mathcal{F}_{t-1}) \equiv \boldsymbol{\Sigma}_u(t) = \mathrm{diag}(\sigma_{1t}^2, \cdots, \sigma_{Nt}^2),$$

where

$$\sigma_{\ell t}^2 = \alpha_{\ell 0} + \sum_{i=1}^{m} \alpha_{\ell i} u_{\ell t-i}^2 + \sum_{j=1}^{s} \beta_{\ell j} \sigma_{\ell t-j}^2, \quad \ell = 1, \cdots, N. \tag{2.3}$$

The true parameters of this GARCH structure are $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_N)$, where $\boldsymbol{\theta}_\ell = (\alpha_{\ell 0}, \cdots, \alpha_{\ell m}, \beta_{\ell 1}, \cdots, \beta_{\ell s})^{\mathcal{T}}$, for $\ell = 1, \cdots, N$.

Our objective is to propose an estimation method that can simultaneously handle time-varying factor loadings, latent factors, and heteroskedastic noise. Hence, for the sake of simplicity of analysis, we simplify the structure of the factor covariance matrix as much as possible. Let $\boldsymbol{\Sigma}_f(t)$ be the conditional covariance matrix of $\mathbf{f}_t$, which does not dependent on $\mathcal{F}_{t-1}$, and the information is set up to time $t-1$. We focus our attention on the conditional covariance matrix of $\mathbf{y}_t$ such that

$$\operatorname{cov}(\mathbf{y}_t | \mathcal{F}_{t-1}) \equiv \boldsymbol{\Sigma}_y(t) = \boldsymbol{\Lambda}_t \boldsymbol{\Sigma}_f(t) \boldsymbol{\Lambda}_t^{\mathcal{T}} + \boldsymbol{\Sigma}_u(t). \tag{2.4}$$

The above model can be called the factor-GARCH model. There are two challenges when using it to estimate the conditional covariance matrix. On the one hand, we need to overcome the curse of dimensionality to obtain accurate estimations of factors and factor loadings. On the other hand, it is necessary to estimate the relevant parameters of the GARCH structure in the residuals before calculating the estimation of $\boldsymbol{\Sigma}_u(t)$. Next, we will introduce methods to address these two issues.

2.2. Projection technique

To address the problem of the curse of dimensionality to establish an accurate factor model, projecting high-dimensional data onto a low-dimensional space spanned by a covariate may be helpful. Let $\mathbf{Y}$ be the matrix of y_{it}. To this end, we refer to the common correlated effects (CCE) proxy factor method proposed by Pesaran [22] to project $\mathbf{Y}$ onto the space spanned by covariate $\mathbf{X}$.

Suppose $\mathbf{X}$ can be divided into M groups with each group containing d components, such that

$$\mathbf{x}_{jt} = \mathbf{D}_{jt} \mathbf{f}_t + \mathbf{v}_{jt}, \tag{2.5}$$

where $j = 1, \ldots, M$ denotes the group, $\mathbf{x}_{jt}$ is a $d \times 1$ vector, $\mathbf{D}_{jt}$ is a $d \times R$ matrix of time-varying loadings, and $\mathbf{v}_{jt}$ is the noise vector, which can be reduced by averaging $\mathbf{x}_{jt}$ over j:

$$\bar{\mathbf{x}}_t = \mathbf{D}_t \mathbf{f}_t + \bar{\mathbf{v}}_t. \tag{2.6}$$

Before using the projection technique, we rescale the factor loadings $\boldsymbol{\lambda}_{it} = \boldsymbol{\lambda}_i(t/T)$ and turn them to constants since

$$\boldsymbol{\lambda}_{it} = \boldsymbol{\lambda}_i\left(\frac{t}{T}\right) \approx \boldsymbol{\lambda}_i\left(\frac{r}{T}\right) = \boldsymbol{\lambda}_{ir}, \quad \text{when} \quad \frac{t}{T} \approx \frac{r}{T}. \tag{2.7}$$

Let $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT})^{\mathcal{T}}$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_T)^{\mathcal{T}}$, $\mathbf{u}_i = (u_{i1}, \cdots, u_{iT})^{\mathcal{T}}$, and $\mathbf{K}^{1/2}(r) = \operatorname{diag}(k_{1r}^{1/2}, \cdots, k_{Tr}^{1/2})$, which is the matrix of the kernel function $k_{tr} = h^{-1}K((t-r)/(Th))$. We multiply (2.1) by $k_{tr}^{1/2}$ and stack it over t for each i, which leads to

$$\mathbf{y}_i^{(r)} \approx \mathbf{F}^{(r)} \boldsymbol{\lambda}_{ir} + \mathbf{u}_i^{(r)} \quad \text{when} \quad \frac{t}{T} \approx \frac{r}{T}, \quad i = 1, \ldots, N, \tag{2.8}$$

where $\mathbf{y}_i^{(r)} = \mathbf{K}^{1/2}(r)\mathbf{y}_i$, $\mathbf{F}^{(r)} = \mathbf{K}^{1/2}(r)\mathbf{F}$, and $\mathbf{u}_i^{(r)} = \mathbf{K}^{1/2}(r)\mathbf{u}_i$. The matrix form of (2.8) can be expressed as:

$$\mathbf{Y}^{(r)} \approx \mathbf{F}^{(r)} \boldsymbol{\Lambda}_r^{\mathcal{T}} + \mathbf{U}^{(r)} \quad \text{when} \quad \frac{t}{T} \approx \frac{r}{T}, \tag{2.9}$$

where $\mathbf{Y}^{(r)} = (\mathbf{y}_1^{(r)}, \mathbf{y}_2^{(r)}, \cdots, \mathbf{y}_N^{(r)})$ is an $T \times N$ matrix, $\mathbf{F}^{(r)}$ is a $T \times R$ matrix, $\mathbf{\Lambda}_r = (\lambda_{1r}, \lambda_{2r} \cdots, \lambda_{Nr})^{\mathcal{T}}$ is a $N \times R$ matrix, and $\mathbf{U}^{(r)} = (\mathbf{u}_1^{(r)}, \mathbf{u}_2^{(r)} \cdots, \mathbf{u}_N^{(r)})$ is a $T \times N$ matrix. The above localization process is normal, which references the local principal component analysis (LPCA) method proposed by Su and Wang [5].

Let $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_T)^{\mathcal{T}}$ and $\bar{\mathbf{V}} = (\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \ldots, \bar{\mathbf{v}}_T)^{\mathcal{T}}$. Similarly, we apply the localization process to the factor loading matrix in model (2.6), such that

$$\bar{\mathbf{X}}^{(r)} \approx \mathbf{F}^{(r)}\mathbf{D}_r^{\mathcal{T}} + \bar{\mathbf{V}}^{(r)} \quad \text{when} \quad \frac{t}{T} \approx \frac{r}{T}, \tag{2.10}$$

where $\bar{\mathbf{X}}^{(r)} = \mathbf{K}^{1/2}(r)\bar{\mathbf{X}}$ and $\bar{\mathbf{V}}^{(r)} = \mathbf{K}^{1/2}(r)\bar{\mathbf{V}}$.

Now we project the localized observable data matrix $\mathbf{Y}^{(r)}$ onto the space spanned by $\bar{\mathbf{X}}^{(r)}$. Let the $T \times T$ matrix

$$\mathbf{P}^{(r)} = \bar{\mathbf{X}}^{(r)} \left(\bar{\mathbf{X}}^{(r)\mathcal{T}}\bar{\mathbf{X}}^{(r)}\right)^{-1} \bar{\mathbf{X}}^{(r)\mathcal{T}} \tag{2.11}$$

be the projection matrix. Premultiplying (2.9) by $(1/\sqrt{TN})\mathbf{P}^{(r)}$, we have

$$\frac{1}{\sqrt{TN}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)} \approx \frac{1}{\sqrt{TN}}\mathbf{P}^{(r)}\mathbf{F}^{(r)}\mathbf{\Lambda}_r^{\mathcal{T}} + \frac{1}{\sqrt{TN}}\mathbf{P}^{(r)}\mathbf{U}^{(r)}. \tag{2.12}$$

To accurately estimate the factors and the factor loading matrix, it is reasonable that the projection error term $\mathbf{P}^{(r)}\mathbf{U}^{(r)}$ is as small as possible. Thus, the estimation problem is transformed into the following criterion:

$$\min_{\mathbf{P}^{(r)}\mathbf{F}^{(r)}, \mathbf{\Lambda}_r} \frac{1}{TN}\text{trace}\left[\left(\mathbf{P}^{(r)}\mathbf{Y}^{(r)} - \mathbf{P}^{(r)}\mathbf{F}^{(r)}\mathbf{\Lambda}_r^{\mathcal{T}}\right)\left(\mathbf{P}^{(r)}\mathbf{Y}^{(r)} - \mathbf{P}^{(r)}\mathbf{F}^{(r)}\mathbf{\Lambda}_r^{\mathcal{T}}\right)^{\mathcal{T}}\right], \tag{2.13}$$

which is equivalent to

$$\max_{\mathbf{\Lambda}_r} \frac{1}{TN}\text{trace}\left(\mathbf{\Lambda}_r^{\mathcal{T}}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}\mathbf{\Lambda}_r\right). \tag{2.14}$$

We now impose the identification restrictions that $T^{-1}\mathbf{F}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{F}^{(r)}$ is a diagonal matrix with distinct entries and $N^{-1}\mathbf{\Lambda}_r^{\mathcal{T}}\mathbf{\Lambda}_r = \mathbf{I}_R$, so (2.14) can be regarded as a standard principal component analysis (PCA) problem, implying that $\hat{\mathbf{\Lambda}}_r$ can be described as the first R principal components of $(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}$.

Once the estimated factor loading matrix $\hat{\mathbf{\Lambda}}_r$ is given, the least squares estimator of the factor is $\hat{\mathbf{F}}^{(r)} = \mathbf{Y}^{(r)}\hat{\mathbf{\Lambda}}_r/N$ by (2.9). However, this is not optimal since $\mathbf{Y}^{(r)}$ is still disturbed by non-negligible noises, which can only be eliminated by projecting the responses onto the space spanned by $\bar{\mathbf{X}}^{(r)}$, leading to the following estimator:

$$\widehat{\mathbf{P}^{(r)}\mathbf{F}^{(r)}} = \frac{1}{N}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}\hat{\mathbf{\Lambda}}_r. \tag{2.15}$$

Then, we recover $\mathbf{F}$ by:

$$\hat{\mathbf{F}} = \left(\sum_{r=1}^{T}\mathbf{K}(r)\right)^{-1}\sum_{r=1}^{T}\mathbf{K}^{1/2}(r)\widehat{\mathbf{P}^{(r)}\mathbf{F}^{(r)}}. \tag{2.16}$$

We now address the problem of estimating the number of factors R when it is unknown. Based on Ahn and Horenstein [23], we define $\lambda_j(\mathbf{A})$ as the jth largest eigenvalue of matrix $\mathbf{A}$, and select R according to

$$\hat{R} = \arg\max_{j<d} \frac{\lambda_j[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]}{\lambda_{j+1}[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]}, \tag{2.17}$$

where d is the dimension of $\mathbf{x}_{jt}$, which should be no smaller than R, otherwise the number of factors will be underestimated.

2.3. Parameter estimation

Define

$$e_{\ell t} \equiv \hat{u}_{\ell t} = y_{\ell t} - \hat{\boldsymbol{\lambda}}_{\ell t}^{\mathcal{T}} \hat{\mathbf{f}}_t, \quad \ell = 1, \cdots, N.$$

Let $u_{\ell t}$ in (2.3) be replaced by $e_{\ell t}$, and a synthetic GARCH model can be acquired as follows:

$$\sigma_{\ell t}^2 = \alpha_{\ell 0} + \sum_{i=1}^{m} \alpha_{\ell i} e_{\ell t-i}^2 + \sum_{j=1}^{s} \beta_{\ell j} \sigma_{\ell t-j}^2, \quad \ell = 1, \cdots, N. \tag{2.18}$$

The QMLE (quasi-maximum likelihood estimation) approach is used to estimate $\boldsymbol{\theta}_\ell$, and the negative quasi log-likelihood function of $\boldsymbol{\theta}_\ell$ is defined as follows:

$$Q_T(\boldsymbol{\theta}_\ell) = \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{e_{\ell t}^2}{\sigma_{\ell t}^2(\boldsymbol{\theta}_\ell)} + \log \sigma_{\ell t}^2(\boldsymbol{\theta}_\ell) \right\}, \tag{2.19}$$

where $\sigma_{\ell t}^2(\boldsymbol{\theta}_\ell)$ can be recursively computed by (2.18) with initial values

$$e_{\ell 0}^2 = \cdots = e_{\ell,1-m}^2 = \sigma_{\ell 0}^2 = \cdots = \sigma_{\ell,1-s}^2 = \alpha_{\ell 0}. \tag{2.20}$$

Note that Francq et al. [24] demonstrated that the choice of the initial values is unimportant for the asymptotic properties of the QMLE. Hence, we minimize (2.19) corresponding to $\boldsymbol{\theta}_\ell$ on a compact set Λ defined in Assumption A8 and obtain the QMLE of $\hat{\boldsymbol{\theta}}_\ell$ as follows:

$$\hat{\boldsymbol{\theta}}_\ell = \arg\min_{\boldsymbol{\theta}_\ell \in \Lambda} Q_T(\boldsymbol{\theta}_\ell), \quad \ell = 1, \cdots, N. \tag{2.21}$$

Once $\boldsymbol{\theta}_\ell s$ are estimated, we substitute them into (2.18) and set the initial values for $\sigma_{\ell t}^2$ and $e_{\ell t}^2$, e.g., in (2.20), an estimator $\hat{\sigma}_{\ell t}^2$ of $\sigma_{\ell t}^2$ can be acquired, namely, an estimator $\hat{\boldsymbol{\Sigma}}_u(t)$ of $\boldsymbol{\Sigma}_u(t)$ is obtained.

In order to estimate $\boldsymbol{\Sigma}_f(t)$, we use the sample covariance matrix estimator proposed by Fan et al. [25] such that

$$\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) \equiv \widehat{\text{cov}}(\hat{\mathbf{f}}_t) = (T-1)^{-1} \hat{\mathbf{F}}^{\mathcal{T}} \hat{\mathbf{F}} - [T(T-1)]^{-1} \hat{\mathbf{F}}^{\mathcal{T}} \mathbf{1} \mathbf{1}^{\mathcal{T}} \hat{\mathbf{F}}, \tag{2.22}$$

where $\mathbf{1}$ is a T-dimensional column vector of ones. Once the above estimations are given, the conditional covariance matrix of $\mathbf{y}_t$ is as follows:

$$\hat{\boldsymbol{\Sigma}}_y(t) = \hat{\boldsymbol{\Lambda}}_t \hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) \hat{\boldsymbol{\Lambda}}_t^{\mathcal{T}} + \hat{\boldsymbol{\Sigma}}_u(t). \tag{2.23}$$

3. Asymptotic theories

Before the statement of asymptotic properties, some assumptions are necessary.

(A1) (i)$E(\|\mathbf{f}_t\|^4) < \infty$ and $T^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{F} \xrightarrow{P} \boldsymbol{\Sigma}_F$, where $\boldsymbol{\Sigma}_F$ is a diagonal matrix with positive distinct entries.
(ii)There is a constant L, such that $P\left\{\sup_{t\leq T} \|\mathbf{f}_t\| \leq L\right\} = 1 - O\left(\frac{1}{T^{1+\epsilon}}\right)$ with a small constant ϵ.

(A2) For each $i \le N, t \le T$, there is a constant C such that $N^{-1}\mathbf{\Lambda}_t^{\mathcal{T}}\mathbf{\Lambda}_t = \mathbf{I}_R$ and $\|\boldsymbol{\lambda}_{it}\| < C$.

(A3) (i) $\{\mathbf{f}_t\}$ are independent of noises $\{u_{it}\}$ and $\{\mathbf{V}_{jt}\}$.

(ii) $\{u_{it}\}$ and $\{\mathbf{V}_{jt}\}$ can be correlated in the sense that, for $p = 1, \ldots, d$ and $n = 1, \ldots, N$,

$$\left(h^{1/2}/\sqrt{TM}\right)\sum_{j=1}^{M}\sum_{t=1}^{T} k_{tr}u_{nt}V_{jt,p} = O_p(1),$$

where $V_{jt,p}$ is the pth element of $\mathbf{V}_{jt}$.

(A4) (i) For $d \times 1$ noises, $E(\|\mathbf{V}_{jt}\|) = 0$ for all j and t.

(ii) For each t and $p, q = 1, \ldots, d$,

$$(1/M)\sum_{i=1}^{M}\sum_{j=1}^{M}|E(V_{it,p}V_{jt,q})| = O(1).$$

(iii) Dimension d of X_{it} is greater than the number R of factors.

(A5) (i) $\{\mathbf{f}_t, u_{it}, \mathbf{V}_{jt}\}_{t\le T}$ is strictly stationary and a strong α-mixing with mixing coefficients $\alpha(t)$. For noise u_{it}, $E(u_{it}) = 0$ for all i and t.

(ii) There exist $r_1, C_1 > 0$ such that for all $t > 0$,

$$\alpha(t) < \exp(-C_1 t^{r_1}),$$

where $\alpha(t)$ is defined in (i).

(iii) There is $C_2 > 0$ such that

$$\max_{j\le N}\sum_{i=1}^{N}|Eu_{it}u_{jt}| < C_2,$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|Eu_{it}u_{js}| < C_2,$$

$$\max_{i\le N}\frac{1}{NT}\sum_{k=1}^{N}\sum_{m=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\mathrm{cov}(u_{it}u_{kt}, u_{is}u_{ms})| < C_2.$$

(iv) Exponential tail: There exist $r_2, r_3 > 0$ satisfying $r_1^{-1} + r_2^{-1} + r_3^{-1} > 1$ and $b_1, b_2 > 0$, such that for any $s > 0$, $i \le N$, $t \le T$, and $j \le R$,

$$P(|u_{it}| > s) \le \exp\left(-\left(\frac{s}{b_1}\right)^{r_2}\right), \quad P(|f_{tj}| > s) \le \exp\left(-\left(\frac{s}{b_2}\right)^{r_3}\right).$$

(A6) As $T, N, M \to \infty$, $h \to 0$,

$$Th^2 \to \infty, \quad Th^3 \to 0, \quad M/\sqrt{Th} \to 0, \quad Th/N \to 0, \quad Th/N^{1/2} \to \infty.$$

Assumptions A1 and A2 are the identification conditions of factors and factor loadings, which also require that $\mathbf{f}_t$ and $\boldsymbol{\lambda}_{it}$ are bounded for each $i \leq N, t \leq T$. Assumptions A3–A5 are standard, allowing $\{u_{it}\}$ and $\{\mathbf{V}_{jt}\}$ to be weak correlated. Specifically, condition (iii) in Assumption A5 is commonly imposed for high-dimensional factor analysis (see Stock and Watson [26], Bai [27]).

Formally, we have the following theorems:

Theorem 1. *Under Assumptions A1–A6, the estimation of R satisfies*

$$\lim_{N,T\to\infty} P\left(\hat{R} = R\right) = 1.$$

Theorem 2. *Under Assumptions A1–A6, for each $i \leq N, r \leq T$, we have*

$$\|\hat{\boldsymbol{\lambda}}_{ir} - \boldsymbol{\lambda}_{ir}\| = O_p\left(\frac{1}{\sqrt{Th}}\right).$$

Theorem 3. *Under Assumptions A1–A6, for each $t \leq T$, we have*

$$\|\hat{\mathbf{f}}_t - \mathbf{f}_t\| = O_p\left(\frac{1}{\sqrt{Th}}\right).$$

By Theorems 2 and 3, for each $i \leq N$ and $r, t \leq T$, we have

$$\frac{1}{N}\|\hat{\boldsymbol{\Lambda}}_r - \boldsymbol{\Lambda}_r\|_F^2 = \frac{1}{N}\sum_{i=1}^{N}\|\hat{\boldsymbol{\lambda}}_{ir} - \boldsymbol{\lambda}_{ir}\|^2 = O_p\left(\frac{1}{Th}\right)$$

and

$$\frac{1}{T}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2 = \frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 = O_p\left(\frac{1}{Th}\right).$$

Now we impose a natural assumption as follows.

(A7) For each $i \leq N$ and $r, t \leq T$, there exist constants $C_3 > 0$ and a small constant $\epsilon > 0$ defined in Assumption A8 below, such that

$$P\left(\sup_{i\leq N}\|\hat{\boldsymbol{\lambda}}_{ir} - \boldsymbol{\lambda}_{ir}\| > C_3\left(\frac{1}{\sqrt{Th}}\right)\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right)$$

and

$$P\left\{\sup_{t\leq T}\|\hat{\mathbf{f}}_t - \mathbf{f}_t\| > C_3\left(\frac{1}{\sqrt{Th}}\right)\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

The following assumption is standard for GARCH models as in Guo et al. [15].

(A8) (i) For each $\ell = 1, \cdots, N$, let $\{u_{\ell t}, \sigma_{\ell t}^2\}_{t\geq 1}$ be a strictly stationary process with GARCH(m, s) structure, for $\sup_{1\leq\ell\leq N} E\sigma_{\ell 1}^{2\kappa} < \infty$ and $\kappa > 4$.
(ii) Assume $\zeta_{\ell t} = \sigma_{\ell t}^{-1}u_{\ell t}$, for each t and ℓ. Then, for each $\ell = 1, \cdots, N$, $\zeta_{\ell t} \overset{\text{iid}}{\sim} N(0, 1)$ and $\sup_{1\leq\ell\leq N} E(\zeta_{\ell t}^{2\kappa}) \leq \infty$, where $0 < 2\epsilon \leq \kappa/2 - 2$, with a small constant ϵ.
(iii) For each $\ell = 1, \cdots, N$, the true value $\theta_{\ell,0}$ belongs to the interior of Λ, where Λ is a compact set, and $\Lambda \subset (c, +\infty) \times (c, +\infty)^{m+s}$ for a constant $c > 0$.
(iv) Assume $A_{\ell\theta_\ell}(z) = \sum_{i=1}^{m} \alpha_{\ell i} z^i$ and $B_{\ell\theta_\ell}(z) = 1 - \sum_{j=1}^{s} \beta_{\ell j} z^j$, for $\ell = 1, \cdots, N$. If $s > 0$, $A_{\ell\theta_{\ell,0}}(z)$ and $B_{\ell\theta_{\ell,0}}(z)$ have no common roots, $A_{\ell\theta_{\ell,0}}(1) \neq 0$, and $\alpha_{\ell,0m} + \beta_{\ell,0s} \neq 0$.

In addition to common norms, such as the Frobenius norm, there are other norms sought to help us understand the factor structure, and the entropy loss norm introduced by James and Stein [28] is a good choice. Its formula is

$$\left\|\hat{\mathbf{A}} - \mathbf{A}\right\|_{\Sigma} = N^{-1/2}\left\|\mathbf{A}^{-1/2}\left(\hat{\mathbf{A}} - \mathbf{A}\right)\mathbf{A}^{-1/2}\right\|_F .$$

Now we give the asymptotic properties of the estimator for $\mathbf{\Sigma}_y(t)$ in the following theorem, where $r^{-1} = r_1^{-1} + 1.5r_2^{-1}$, and r_1 and r_2 are defined in Assumption A5.

Theorem 4. *Suppose $N^{6/r-1} = o(T)$ as $N \to \infty$. Under Assumptions A1–A8, there exist constants $C_4 > 0$ and a small constant ϵ defined in Assumption A8(ii), such that*

$$P\left\{\left\|\hat{\mathbf{\Sigma}}_y(t) - \mathbf{\Sigma}_y(t)\right\|_{\Sigma}^2 > \frac{C_4 N}{(Th)^2} + \frac{C_4 \log T}{NT}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

4. Portfolio allocation

In this section, a brief introduction on the procedures of constructing the estimated portfolio allocation using our model is given. First, denote $E(\mathbf{y}_t|\mathcal{F}_{t-1})$ as the conditional expectation of $\mathbf{y}_t$, and then we can acquire

$$E(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{\Lambda}_t E(\mathbf{f}_t|\mathcal{F}_{t-1}).$$

Hence,

$$\hat{E}(\mathbf{y}_t|\mathcal{F}_{t-1}) = \hat{\mathbf{\Lambda}}_t \hat{E}(\mathbf{f}_t|\mathcal{F}_{t-1})$$

is utilized as the estimation of $E(\mathbf{y}_t|\mathcal{F}_{t-1})$, where $E(\mathbf{f}_t|\mathcal{F}_{t-1})$ is estimated by applying a VAR(1) model (see Tsay [29]) as follows:

$$\hat{E}(\mathbf{f}_t|\mathcal{F}_{t-1}) = \hat{\boldsymbol{\mu}} + \hat{\mathbf{\Xi}}\mathbf{f}_{t-1}.$$

Based on the mean-variance optimal portfolio, we construct the estimated optimal portfolio allocation similar to Guo et al. [15]: Let $\boldsymbol{\omega}$ denote the allocation vector of N risky assets, to be held between times $t-1$ and t, solve the following minimization problem, and assume the solution as the estimation of $\boldsymbol{\omega}$,

$$\min_{\omega} \ \boldsymbol{\omega}^{\mathcal{T}}\mathbf{\Sigma}_y(t)\boldsymbol{\omega},$$
$$\text{subject to } \boldsymbol{\omega}^{\mathcal{T}}\mathbf{1}_p = 1 \text{ and } \boldsymbol{\omega}^{\mathcal{T}}E(\mathbf{y}_t|\mathcal{F}_{t-1}) = \pi,$$

with π as the target return imposed on the portfolio, and $\mathbf{1}_N$ is an N-dimensional column vector of ones. The existent result gives the explicit estimator of w as follows:

$$\hat{\boldsymbol{\omega}} = \frac{d_3 - d_2\pi}{d_1 d_3 - d_2^2}\hat{\mathbf{\Sigma}}_y^{-1}(t)\mathbf{1}_N + \frac{d_1\pi - d_2}{d_1 d_3 - d_2^2}\hat{\mathbf{\Sigma}}_y^{-1}(t)\hat{E}(\mathbf{y}_t|\mathcal{F}_{t-1}),$$

where

$$d_1 = \mathbf{1}_N^{\mathcal{T}}\hat{\mathbf{\Sigma}}_y^{-1}(t)\mathbf{1}_N, \quad d_2 = \mathbf{1}_N^{\mathcal{T}}\hat{\mathbf{\Sigma}}_y^{-1}(t)\hat{E}(\mathbf{y}_t|\mathcal{F}_{t-1}),$$

$$d_3 = \hat{E}(\mathbf{y}_t|\mathcal{F}_{t-1})^{\mathcal{T}}\hat{\mathbf{\Sigma}}_y^{-1}(t)\hat{E}(\mathbf{y}_t|\mathcal{F}_{t-1}).$$

5. Simulations and empirical study

5.1. Simulations

The data are generated according to

$$\mathbf{y}_t = \mathbf{\Lambda}_t \mathbf{f}_t + \mathbf{u}_t, \quad t = 1, \ldots, T,$$

and

$$\bar{\mathbf{x}}_t = \mathbf{D}_t \mathbf{f}_t + \bar{\mathbf{v}}_t, \quad t = 1, \ldots, T.$$

The factors are generated from

$$\mathbf{f}_t = 0.2\mathbf{f}_{t-1} + \sqrt{1 - 0.2^2}\mathcal{N}(0, \mathbf{\Sigma}_F),$$

where $\mathbf{\Sigma}_F$ is a diagonal matrix with different positive elements. For simplicity, we set the elements of $\mathbf{\Sigma}_F$ as 1 or 2 randomly.

To generate the continuous loadings satisfying $(1/N)\mathbf{\Lambda}_t^{\mathcal{T}} \mathbf{\Lambda}_t = \mathbf{I}_R$, we use a three-step method. The first step is to generate $\mathbf{\Lambda}_1$ such that $(1/N)\mathbf{\Lambda}_1^{\mathcal{T}} \mathbf{\Lambda}_1 = \mathbf{I}_2$. For example, we arbitrarily generate two $N \times 1$ vectors with each entry being independent and identically distributed (i.i.d.) drawn from $\mathcal{N}(-1, 1)$ and then orthogonalize them such that their lengths are $\sqrt{N}$. The second step is to generate a series of orthogonal matrices $\mathbf{H}_t$, $t = 1, \ldots, T$. To this end, we first generate g_{it}, which is an arbitrary function of i and t, and let $\mathbf{g}_t = (g_{1t}, \ldots, g_{Nt})^{\mathcal{T}}$. In particular, $g_{it} = -2 + \text{sign}(\delta_i) \sin(2\pi\delta_i t/T)^{(1/2)}\delta_i + \delta_i$, where $\delta_i \sim \mathcal{N}(0, 1)$. We then normalize $\mathbf{g}_t$, that is, $\mathbf{h}_t = \mathbf{g}_t/\|\mathbf{g}_t\|$. At last, we construct the orthogonal matrices $\mathbf{H}_t = \mathbf{I}_N - 2\mathbf{h}_t\mathbf{h}_t^{\mathcal{T}}$. The third step is to premultiply $\mathbf{\Lambda}_1$ with these matrices $\mathbf{H}_t$ successively.

For the idiosyncratic error vector $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})^{\mathcal{T}}$, let $m = s = 1$ and we assume $\alpha_0 = 0.25$, $\alpha_1 = 0.15, \beta_1 = 0.1$. We generate $\zeta_{\ell t}$ from $\mathcal{N}(0, 1)$, such that $u_{\ell t} = \sigma_{\ell t}\zeta_{\ell t}$ for each $\ell = 1, \cdots, N$ and $t = 1, \cdots, T$.

The loadings $\mathbf{D}_{it}$ are a $3 \times R$ matrix, so $d = 3$. We set the (1, 1)th element of $\mathbf{D}_{it}$ as

$$D_{11,it} = 10/(1 + 0.5 \exp(-t/T)) + \delta_{11,i}[3 \times (t/T) - 2(t/T)^2],$$

the (3, 2)th element as

$$D_{32,it} = -2 + \text{sign}(\delta_{32,i}) \sin(2\pi\delta_{32,i}t/T)^{(1/2)}\delta_{32,i} + \delta_{32,i},$$

and the rest of the elements are $D_{mn,it} = \theta_{mn} + \delta_{mn,i}$, where the subscript mn refers to the (m, n)th element. $\delta_{mn,i}$ are i.i.d. drawn from $\mathcal{N}(0, 1)$, and θ_{mn} are i.i.d. drawn from discrete $U[-2, 3]$. The idiosyncratic error vector $\mathbf{V}_t = [\mathbf{V}_{1t}^\top, \ldots, \mathbf{V}_{Mt}^\top]^\top$ is drawn from

$$\mathbf{V}_t = 0.2\mathbf{V}_{t-1} + \mathcal{N}(0, (1 - 0.2^2)\mathbf{\Sigma}_v),$$

where the (i, j)th element of $\mathbf{\Sigma}_v$ is $0.2^{|i-j|}$.

The kernel function is set as the Epanechnikov kernel, with the normal reference bandwidth

$$h = 2.34/\sqrt{12}T^{-0.2}N^{-0.1}.$$

We first consider the performances of estimating the factor number R. We set $T = 100$ or 200, with N ranging from 49 to 144, and consider the cases with the number of factors $R = 2, 3, 4$, respectively.

To demonstrate the superiority of our projected method, named P-LPCA in this part, two alternative methods, PCA and LPCA, are applied for comparison. While our projected method P-LPCA works on $(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}$, PCA works on the original data matrix $(NT)^{-1}\mathbf{Y}^{\mathcal{T}}\mathbf{Y}$ and LPCA works on $(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{Y}^{(r)}$. In each situation, we reproduce 200 times, and the results are displayed in Figure 1. It can be observed that, compared to other methods, our projection method provides the most accurate estimations, with accuracy improving as the sample size increases.

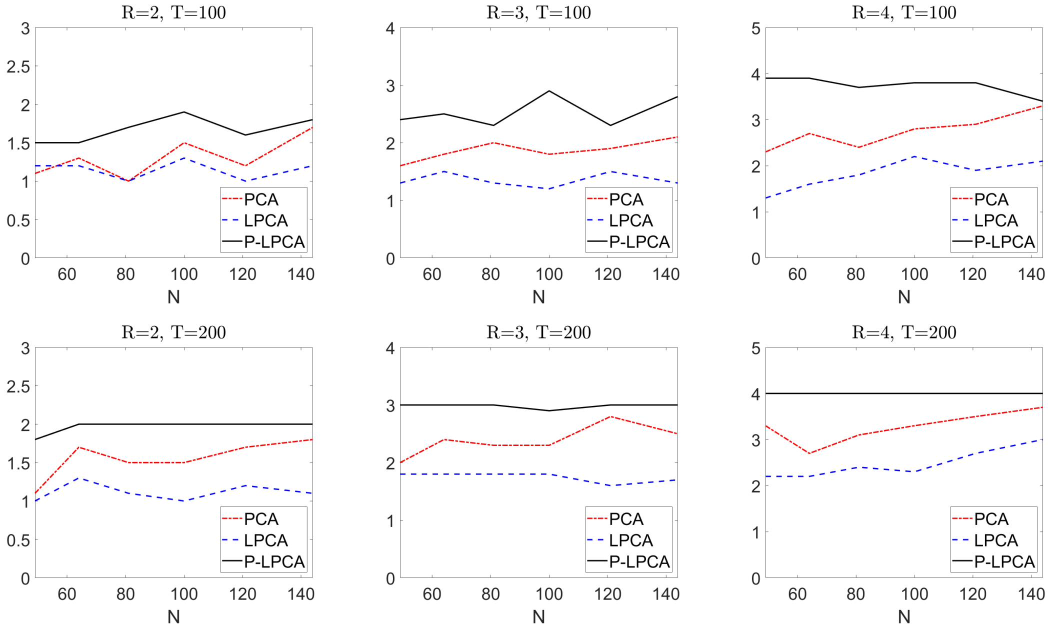

Figure 1. The estimation of factor number by our method (solid black), LPCA (dashed blue), and traditional PCA (dot-dash red).

Then we consider the estimation of factors, factor loading and the conditional covariance matrix. For simplicity, we only consider the setting that $R = 2$, $T = 100$ or 200, with N ranging from 49 to 144. We reproduce 200 times, and in each replication, we compute the estimation error of factor loadings by averaging each $\frac{1}{N}\|\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\|_F^2$ across $r \leq T$, so as to the conditional covariance matrix. To demonstrate the superiority of our proposed method, named PLPCA-GARCH, four alternative methods are applied for comparison. The details of all measures are provided as follows:

(1) PCA: In this method, the columns of $\hat{\mathbf{\Lambda}}$ are the first R eigenvectors of $\mathbf{Y}^{\mathcal{T}}\mathbf{Y}$, and let $\hat{\mathbf{f}}_t = \hat{\mathbf{\Lambda}}\mathbf{y}_t/N$ by the least squares method. Then, we calculate the simple sample covariance matrix of $\hat{\mathbf{u}}_t$ as an estimation of $\mathbf{\Sigma}_u(t)$.

(2) LPCA: In this method, the columns of $\hat{\mathbf{\Lambda}}_t$ are the first R eigenvectors of $(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{Y}^{(r)}$, and let $\hat{\mathbf{f}}_t = \hat{\mathbf{\Lambda}}_t\mathbf{y}_t/N$ by the least squares method. Then, we calculate the simple sample covariance matrix of $\hat{\mathbf{u}}_t$ as an estimation of $\mathbf{\Sigma}_u(t)$.

(3) LPCA-GARCH: Use the same method as LPCA to estimate factors and factor loadings, and then get the estimation of $\mathbf{\Sigma}_u(t)$ by the QMLE approach.

(4) PLPCA: Use our projected method to estimate factors and factor loadings, and then calculate the simple sample covariance matrix of $\hat{\mathbf{u}}_t$ as an estimation of $\mathbf{\Sigma}_u(t)$.

(5) PLPCA-GARCH: Use our projected method to estimate factors and factor loadings, and then get the estimation of $\mathbf{\Sigma}_u(t)$ by the QMLE approach.

In the aforementioned methods, PCA and LPCA consider neither projection nor the GARCH structure. The only difference between these two methods lies in whether the data is locally processed. Besides, PLPCA uses our projection method but similarly does not account for the GARCH effects, while LPCA-GARCH only considers the GARCH structure.

Figures 2 and 3 illustrate different cases by examining the estimation errors of factors and factor loadings. It can be observed that the proposed method demonstrates a significant advantage over traditional PCA and LPCA. The estimation errors of our projected method are smaller, and our estimators exhibit a certain level of robustness, which is consistent with our asymptotic theory.

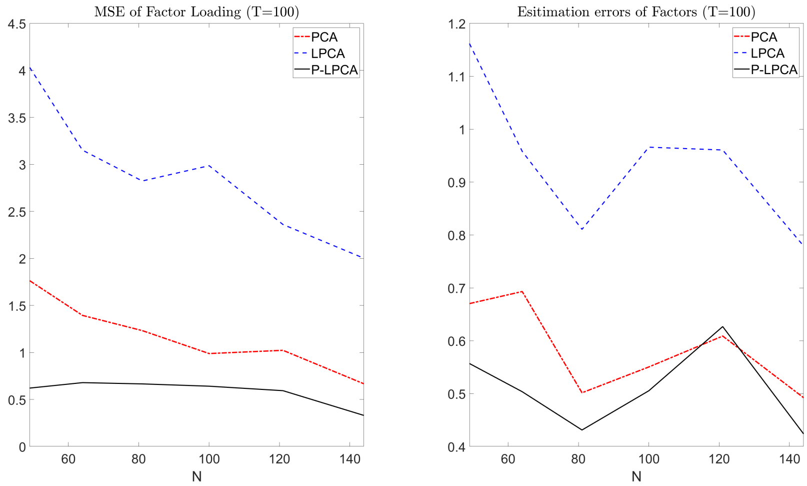

Figure 2. Averaged $\frac{1}{N}\|\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\|_F^2$ and $\frac{1}{T}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2$ ($T = 100$) by our method (solid black), LPCA (dashed blue), and traditional PCA (dot-dash red).

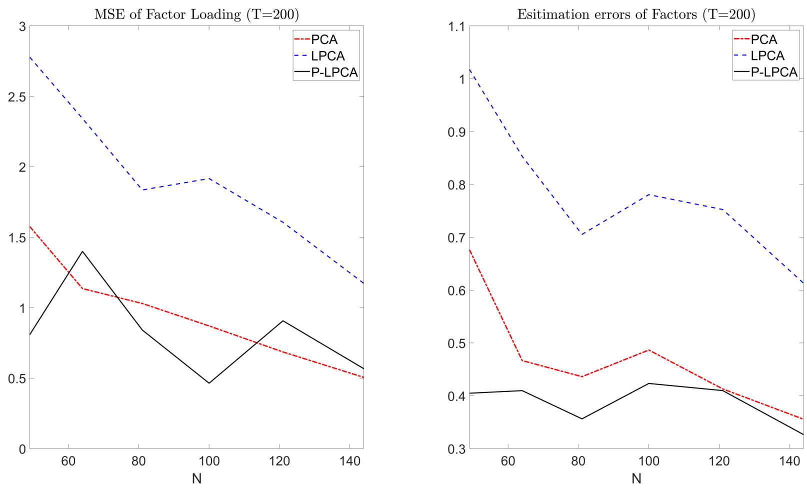

Figure 3. Averaged $\frac{1}{N}\|\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\|_F^2$ and $\frac{1}{T}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2$ ($T = 200$) by our method (solid black), LPCA (dashed blue), and traditional PCA (dot-dash red).

Figures 4–5 and Tables 1–4 show the performance of the estimation of the conditional covariance matrix. When we do not consider the GARCH structure in the idiosyncratic error term, the estimation of the conditional covariance matrix becomes less accurate, because the accumulated heteroskedastic noise is non-negligible in the high-dimensionality framework. Building on this, it is also necessary to consider the precise estimation of factors and factor loadings, as the large errors in the PCA, LPCA, and LPCA-GARCH methods indicate. Our method, however, takes all of the above considerations into account, resulting in a more accurate estimation.

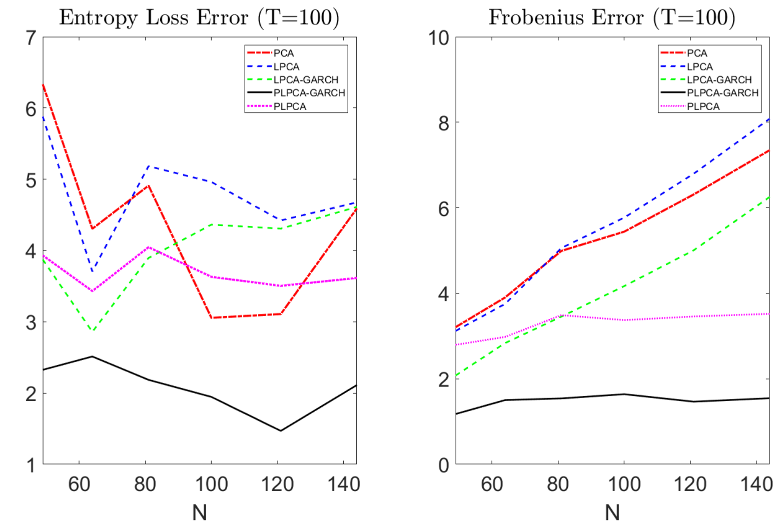

Figure 4. Conditional covariance matrix estimation error ($T = 100$) by our method (solid black), traditional PCA (dot-dash red), LPCA (dashed blue), LPCA-GARCH (dashed green), and PLPCA (dotted purple).

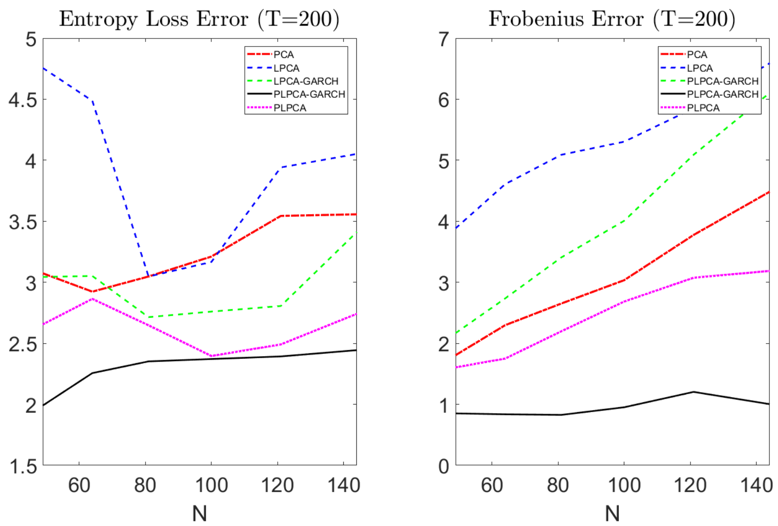

Figure 5. Conditional covariance matrix estimation error ($T = 200$) by our method (solid black), traditional PCA (dot-dash red), LPCA (dashed blue), LPCA-GARCH (dashed green), and PLPCA (dotted purple).

Table 1. Mean estimation errors and standard deviations under $\|\cdot\|_\Sigma$ for $T = 100$.

Dimension	PCA	LPCA	LPCA-GARCH	PLPCA-GARCH	PLPCA
$N = 49$	6.3328	5.8812	3.8732	**2.3257**	3.9319
	(0.5012)	(0.2430)	(0.5344)	(0.0395)	(0.3580)
$N = 64$	4.3053	3.7060	2.8588	**2.5128**	3.4306
	(2.0595)	(2.1280)	(0.4502)	(0.6189)	(0.3152)
$N = 81$	4.9104	5.1837	3.8954	**2.1851**	4.0469
	(2.2894)	(1.9050)	(1.3287)	(1.2762)	(0.0515)
$N = 100$	3.0557	4.9620	4.363	**1.9458**	3.6308
	(0.0705)	(0.1335)	(0.6025)	(0.1213)	(0.6367)
$N = 121$	3.1084	4.4222	4.3071	**1.4685**	3.5036
	(0.2006)	(0.3172)	(0.7145)	(0.0419)	(1.0591)
$N = 144$	4.5903	4.6753	4.6116	**2.1115**	3.6154
	(0.5091)	(0.2589)	(0.6198)	(0.1901)	(0.0119)

Table 2. Mean estimation errors and standard deviations under $\|\cdot\|_F$ for $T = 100$.

Dimension	PCA	LPCA	LPCA-GARCH	PLPCA-GARCH	PLPCA
$N = 49$	3.2066	3.1184	2.0754	**1.1759**	2.7940
	(0.1854)	(0.3162)	(1.3661)	(1.4976)	(0.4045)
$N = 64$	3.9028	3.7518	2.8362	**1.5022**	2.9793
	(0.1065)	(0.9422)	(0.861)	(1.2912)	(0.2052)
$N = 81$	4.9911	5.0577	3.4535	**1.5410**	3.4883
	(0.2579)	(0.7512)	(0.7031)	(0.7395)	(1.0508)
$N = 100$	5.4402	5.7683	4.1677	**1.6396**	3.3723
	(0.2465)	(1.8583)	(1.4635)	(1.6769)	(0.0328)
$N = 121$	6.3100	6.7945	5.0042	**1.4650**	3.4575
	(0.2915)	(1.5958)	(0.6856)	(0.8393)	(1.1470)
$N = 144$	7.3468	8.0816	6.2526	**1.5454**	3.5194
	(0.0549)	(1.1198)	(1.2614)	(0.9536)	(0.7993)

Table 3. Mean estimation errors and standard deviations under $\|\cdot\|_\Sigma$ for $T = 200$.

Dimension	PCA	LPCA	LPCA-GARCH	PLPCA-GARCH	PLPCA
$N = 49$	3.0738	4.7562	3.0438	**1.9901**	2.6551
	(1.1503)	(2.7467)	(0.3522)	(0.2966)	(1.7730)
$N = 64$	2.9232	4.4867	3.0516	**2.2560**	2.8641
	(0.2922)	(0.8904)	(0.5073)	(0.9366)	(1.7086)
$N = 81$	3.0475	3.0470	2.7145	**2.3519**	2.6475
	(0.7082)	(0.7457)	(1.4632)	(0.9853)	(0.2953)
$N = 100$	3.2107	3.1664	2.7599	**2.3712**	2.3962
	(0.8429)	(1.4291)	(0.9497)	(0.4782)	(0.8907)
$N = 121$	3.5438	3.9410	2.8057	**2.3925**	2.4910
	(0.3604)	(0.5834)	(1.3318)	(0.1466)	(0.3828)
$N = 144$	3.5569	4.0518	3.4117	**2.4439**	2.7418
	(0.1753)	(0.4593)	(0.9971)	(1.0227)	(1.8652)

Table 4. Mean estimation errors and standard deviations under $\|\cdot\|_F$ for $T = 200$.

Dimension	PCA	LPCA	LPCA-GARCH	PLPCA-GARCH	PLPCA
$N = 49$	1.8051	3.8840	2.1664	**0.8515**	1.6060
	(0.0423)	(0.1943)	(0.9591)	(0.1891)	(0.3466)
$N = 64$	2.2970	4.6077	2.7353	**0.8369**	1.7496
	(0.0240)	(0.2453)	(1.625)	(0.4635)	(0.0742)
$N = 81$	2.6523	5.0894	3.4074	**0.8272**	2.1971
	(0.0680)	(0.1215)	(0.908)	(0.5783)	(0.0685)
$N = 100$	3.0342	5.3037	4.0078	**0.9518**	2.6843
	(0.3033)	(0.6530)	(1.3139)	(1.0566)	(0.2129)
$N = 121$	3.7755	5.8507	5.0892	**1.2038**	3.0744
	(0.2124)	(0.9977)	(1.5375)	(0.9840)	(0.2696)
$N = 144$	4.4837	6.5906	6.1036	**1.0034**	3.1862
	(0.1836)	(0.6705)	(1.3635)	(0.7858)	(0.4661)

It is of interest to check the impact of the bandwidth h. We examine the model's estimation performance under a fixed setup when the bandwidth h takes different values within a small range, with the results shown in Tables 5 and 6. It can be seen that the estimation error tends to decrease as the bandwidth increases, which is consistent with our asymptotic theory. Besides, the reduction in estimation error is quite limited. This suggests that bandwidth selection does not need to be overly strict. Commonly used bandwidth choices, such as rule-of-thumb bandwidths, plug-in bandwidths, and cross-validation bandwidths, could be sufficient to achieve good estimation performance.

Table 5. Mean estimation errors and standard deviations under $\|\cdot\|_\Sigma$ for different h.

Sample size	Dimension	Bandwidth h	PLPCA-GARCH	PLPCA
		0.05	**2.0378** (0.7261)	3.8849 (0.3732)
		0.1	**2.0196** (0.165)	3.8097 (0.2897)
$T = 100$	$N = 100$	0.15	**2.1191** (0.1579)	3.5989 (0.3639)
		0.2	**1.987** (0.655)	3.6624 (0.5492)
		0.25	**1.9535** (0.5024)	3.5919 (0.3162)
		0.3	**1.9655** (0.1948)	3.5827 (0.2179)

Table 6. Mean estimation errors and standard deviations under $\|\cdot\|_F$ for different h.

Sample size	Dimension	Bandwidth h	PLPCA-GARCH	PLPCA
		0.05	**1.746** (1.0296)	3.4853 (0.4042)
		0.1	**1.7571** (0.6807)	3.4475 (1.2328)
$T = 100$	$N = 100$	0.15	**1.7098** (0.4162)	3.3112 (0.7719)
		0.2	**1.6158** (0.9407)	3.3913 (0.3466)
		0.25	**1.5708** (1.5312)	3.3534 (0.3145)
		0.3	**1.5971** (1.5435)	3.2926 (0.3081)

5.2. Empirical study

Now we apply our model to a real dataset, which is the daily data for the 380 constituent stocks of the SSE 380 Index listed on the Shanghai Stock Exchange (https://www.sse.com.cn/) from 2020 to 2023, including comprehensive daily closing prices and the corresponding industry average returns. Excess returns for each stock on each trading day have been computed.

We first consider the possibility of heteroskedasticity in the data. Referring to Li et al. [17], we use the four statistics in Tsay [29] for testing, and the results are presented in Table 7. It is shown that all the tests suggest the presence of heteroskedasticity in the data.

Table 7. Heteroskedasticity tests.

Method	Test statistic	p-value
$Q(m)$ of squared series (LM test)	2721.172	0.000
$Q_r(m)$ of squared series	3984.948	0.000
Robust test (5% trimming)	2692.520	0.000
Rank-based test	2696.126	0.000

Different methods are used to estimate the conditional covariance matrix of these datasets and calculate the annualized Sharpe ratio. For simplicity, we assume that all possible investment portfolios are fully allocated, short-selling is allowed, and there are no transaction costs. At the end of each trading day, a portfolio $\hat{\omega}$ is formed and held until the end of the next trading day. The calculation method for returns from trading day $t-1$ to t is $R_t(\hat{\omega}) = \hat{\omega}^{\mathcal{T}}\mathbf{y}_t$, where $\hat{\omega}$ is calculated based on $(\mathbf{f}_{t-i}, \mathbf{y}_{t-i})$, $i = 1, \ldots, N$, up to time T. Therefore, the formula for the annualized Sharpe ratio is

$$\mathrm{SR}(\hat{\omega}) = \frac{\overline{R}(\hat{\omega})}{\mathrm{sd}(\hat{\omega})}\sqrt{T},$$

where

$$\overline{R}(\hat{\omega}) = \frac{1}{T}\sum_{t=1}^{T}[R_t(\hat{\omega}) - R_{ft}], \quad \mathrm{sd}(\hat{\omega}) = \left\{\frac{1}{T}\sum_{t=1}^{T}[R_t(\hat{\omega}) - R_{ft} - \overline{R}(\hat{\omega})]^2\right\}^{1/2},$$

where R_{ft} is the risk-free rate on day t. We calculate the annualized Sharpe ratio and the results are shown in Figure 6. It can be seen that our method performs the best, due to the fact that our method can effectively capture the presence of GARCH effects in the data considered.

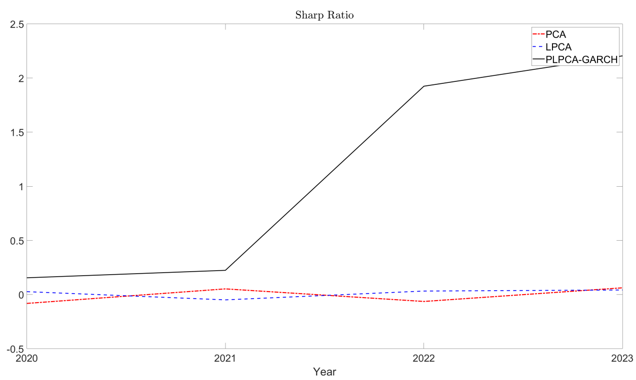

Figure 6. Annualized Sharpe ratios for constituent stocks of SSE 380 calculated by our method (solid black), LPCA (dashed blue), and traditional PCA (dot-dash red).

Next, we employ additional measures to comprehensively assess the performance of various methods. Since rational investors generally prefer assets exhibiting higher skewness and lower kurtosis (see Scott and Horvath [30] and Dittmar [31]) for a greater likelihood of positive extreme returns and a reduced probability of extreme fluctuations, it is essential to take these metrics into consideration. To this end, referring to Zhu et al. [32], we develop three enhanced evaluation frameworks: MV (mean-variance), MVSK (mean-variance with skewness and kurtosis) and SRSK (Sharpe ratio with skewness and kurtosis), which are defined as follows:

$$\mathrm{MV} = \overline{R}(\hat{\omega}) - \lambda_1 * \mathrm{sd}(\hat{\omega}),$$

$$\mathrm{MVSK} = \overline{R}(\hat{\omega}) - \lambda_1 * \mathrm{sd}(\hat{\omega}) + \lambda_2 * \mathrm{Skew}(\hat{\omega}) - \lambda_3 * \mathrm{Kurt}(\hat{\omega}),$$

$$\mathrm{SRSK} = \mathrm{SR}(\hat{\omega}) + \lambda_2 * \mathrm{Skew}(\hat{\omega}) - \lambda_3 * \mathrm{Kurt}(\hat{\omega}),$$

where

$$\mathrm{Skew}(\hat{\omega}) = \frac{\frac{1}{T}\sum_{t=1}^{T}[R_t(\hat{\omega}) - R_{ft} - \overline{R}(\hat{\omega})]^3}{[sd(\hat{\omega})]^3},$$

$$\mathrm{Kurt}(\hat{\omega}) = \frac{\frac{1}{T}\sum_{t=1}^{T}[R_t(\hat{\omega}) - R_{ft} - \overline{R}(\hat{\omega})]^4}{[sd(\hat{\omega})]^4}.$$

In the above measures, λ_1, λ_2, and λ_3 are positive hyperparameters tuned by the grid search within the sets A_1, A_2, and A_3, respectively, to maximize the value of MVSK of each portfolio, where

$$A_1 = \{a \times 10^{-b} \mid a = 1, 2, \ldots, 9 \text{ and } b = -1, 0, \ldots, 3\},$$

$$A_2 = A_3 = \{a \times 10^{-b} : a = 1, 2, \ldots, 9 \text{ and } b = 2, 3, \ldots, 6\}.$$

Since we aim to construct a portfolio with higher expected returns and skewness while maintaining lower risk and kurtosis, larger values of the aforementioned metrics are preferred. The empirical results demonstrate that our proposed method achieves superior portfolio performance. As evidenced in Table 8, our approach consistently outperforms competing methods. While delivering the highest mean returns and effectively controlling downside risk, our method generates significantly greater positive skewness and less kurtosis than the alternatives, increasing the probability of extreme positive returns. Besides, the dominance in both the MVSK and SRSK metrics further confirms that our framework is the preferred choice for investors seeking an optimal portfolio.

Table 8. The performance of three methods.

	Method	2020	2021	2022	2023
MV	PCA	-0.0138	-0.0043	-0.7691	-0.0100
	LPCA	-0.0214	-0.3487	-0.0262	-0.0128
	PLPCA-GARCH	**0.0052**	**0.0016**	**0.0043**	**0.0083**
MVSK	PCA	-0.3853	-0.2684	-0.8260	-0.0591
	LPCA	-0.0730	-0.4913	-0.0330	-0.0982
	PLPCA-GARCH	**0.0017**	**-0.0003**	**0.0031**	**0.0022**
SR	PCA	-0.082	0.0522	-0.064	0.062
	LPCA	0.0265	-0.0493	0.0321	0.042
	PLPCA-GARCH	**0.1546**	**0.2235**	**1.9238**	**2.2039**
SRSK	PCA	-0.4535	-0.2119	-0.1210	0.0129
	LPCA	-0.0251	-0.1919	0.0253	-0.0434
	PLPCA-GARCH	**0.1511**	**0.2216**	**1.9226**	**2.1978**

6. Conclusions

This paper introduces and examines a novel high-dimensional time-varying factor model with GARCH noise, designed to partially address two important issues in financial time series analysis: the smooth evolution of factors and factor loadings, with persistent heteroskedasticity in high-dimensional settings. By the projection technique for estimating continuously varying factor loadings and latent factors, with a GARCH specification to model idiosyncratic volatility, the proposed approach captures dynamic market structures while mitigating the curse of dimensionality. We establish the asymptotic properties of the resulting estimators under suitable regularity conditions. Simulation studies and an empirical application in portfolio allocation demonstrate that the model outperforms static factor benchmarks and provides robust, precise estimates of conditional covariance matrices in evolving high-dimensional environments. These results show the model's practical value for volatility forecasting and financial risk management.

Author contributions

Yuwen Ruan: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing–original draft, Writing–review and editing; Xingfa Zhang: Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Writing–review and editing; Yujiao Liu: Data curation, Formal analysis, Validation; Yan Wang: Software, Supervision; Tianli Lei: Funding acquisition, Supervision, Writing–review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Guangdong Basic and Applied Basic Research Foundation (2022A1515110009), the National Nature Science Foundation of China (12471251), and the Science and Technology Program of Guangzhou of China (2025A03J3083).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. G. Connor, O. Linton, Semiparametric estimation of a characteristic-based factor model of common stock returns, *J. Empir. Finance*, **14** (2007), 694–717. https://doi.org/10.1016/j.jempfin.2006.10.001
2. C. M. Hafner, A. Preminger, Asymptotic theory for a factor GARCH model, *Econ. Theory*, **25** (2009), 336–363. https://doi.org/10.1017/S0266466608090117

3. A. Onatski, Asymptotics of the principal components estimator of large factor models with weakly influential factors, *J. Econ.*, **168** (2012), 244–258. https://doi.org/10.1016/j.jeconom.2012.01.034
4. G. Motta, C. M. Hafner, R. von Sachs, Locally stationary factor models: Identification and nonparametric estimation, *Econ. Theory*, **27** (2011), 1279–1319. https://doi.org/10.1017/S0266466611000053
5. L. Su, X. Wang, On time-varying factor models: Estimation and testing, *J. Econ.*, **198** (2017), 84–101. https://doi.org/10.1016/j.jeconom.2016.12.004
6. M. Chen, X. Tan, J. Wu, Time varying factor models with possibly strongly correlated noises, *J. Appl. Stat.*, **48** (2021), 887–906. https://doi.org/10.1080/02664763.2020.1753024
7. M. Barigozzi, M. Hallin, S. Soccorsi, R. von Sachs, Time-varying general dynamic factor models and the measurement of financial connectedness, *J. Econ.*, **222** (2021), 324–343. https://doi.org/10.1016/j.jeconom.2020.07.004
8. L. Chen, J. J. Dolado, J. Gonzalo, Quantile factor models, *Econometrica*, **89** (2021), 875–910. https://doi.org/10.3982/ECTA15746
9. Q. Chen, N. Roussanov, X. Wang, Semiparametric conditional factor models: Estimation and inference, technical report, *National Bureau of Economic Research*, 2023.
10. X. Wang, S. Jin, Y. Li, J. Qian, L. Su, On time-varying panel data models with time-varying interactive fixed effects, *J. Econ.*, **249** (2025), 105960. https://doi.org/10.1016/j.jeconom.2025.105960
11. Y. Tu, S. Wang, Quantile prediction with factor-augmented regression: Structural instability and model uncertainty, *J. Econ.*, **249** (2025), 105999. https://doi.org/10.1016/j.jeconom.2025.105999
12. R. F. Engle, K. F. Kroner, Multivariate simultaneous generalized arch, *Econ. Theory*, **11** (1995), 122–150. https://doi.org/10.1017/S0266466600009063
13. T. Bollerslev, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model, *Rev. Econom. Stat.*, 1990, 498–505.
14. R. Engle, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *J. Busin. Econ. Stat.*, **20** (2002), 339–350. https://doi.org/10.1198/073500102288618487
15. S. Guo, J. L. Box, W. Zhang, A dynamic structure for high-dimensional covariance matrices and its application in portfolio allocation, *J. Amer. Stat. Assoc.*, **112** (2017), 235–253. https://doi.org/10.1080/01621459.2015.1129969
16. P. Wang, X. Liu, S. Wu, Dynamic linkage between bitcoin and traditional financial assets: A comparative analysis of different time frequencies, *Entropy*, **24** (2022), 1565. https://doi.org/10.3390/e24111565
17. X. L. Li, Y. Li, J. Z. Pan, X. F. Zhang, A factor-garch model for high dimensional volatilities, *Acta Math. Appl. Sin. Engl. Ser.*, **38** (2022), 635–663. https://doi.org/10.1007/s10255-022-1104-6
18. Y. Liu, Y. Li, X. Zhang, Estimation of conditional covariance matrices for a class of high-dimensional varying coefficient factor-generalized autoregressive conditional heteroscedasticity model, *Symmetry*, **16** (2024), 1635. https://doi.org/10.3390/sym16121635

19. A. Laksaci, F. Alshahrani, I. M. Almanjahie, Z. Kaid, Nonparametric multifunctional garch time series data analysis: Application to dynamic forecasting in financial data, *AIMS Math.*, **10** (2025), 26459–26483. https://doi.org/10.3934/math.20251163

20. R. Alraddadi, The markov-switching threshold blgarch model, *AIMS Math.*, **10** (2025), 18838–18860. http://doi.org/10.3934/math.2025842

21. Y. Ruan, X. Zhang, Y. Liu, A semi-parametric factor-garch model for high dimensional covariance matrix estimation, *J. Korean Stat. Soc.*, **54** (2025), 888–919. https://doi.org/10.1007/s42952-025-00324-4

22. M. H. Pesaran, Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica*, **74** (2006), 967–1012. https://doi.org/10.1111/j.1468-0262.2006.00692.x

23. S. C. Ahn, A. R. Horenstein, Eigenvalue ratio test for the number of factors, *Econometrica*, **81** (2013), 1203–1227. https://doi.org/10.3982/ECTA8968

24. C. Francq, J. M. Zakoïan, Estimating multivariate volatility models equation by equation, *J. Royal Stat. Soc. Ser. B: Stat. Methodol.*, **78** (2016), 613–635. https://doi.org/10.1111/rssb.12126

25. J. Fan, Y. Fan, J. Lv, High dimensional covariance matrix estimation using a factor model, *J. Econ.*, **147** (2008), 186–197. https://doi.org/10.1016/j.jeconom.2008.09.017

26. J. H. Stock, M. W. Watson, Forecasting using principal components from a large number of predictors, *J. Amer. Stat. Assoc.*, **97** (2002), 1167–1179. https://doi.org/10.1198/016214502388618960

27. J. Bai, Inferential theory for factor models of large dimensions, *Econometrica*, **71** (2003), 135–171. https://doi.org/10.1111/1468-0262.00392

28. W. James, C. Stein, Estimation with quadratic loss, In: *Breakthroughs in Statistics: Foundations and Basic Theory*, New York: Springer, 1992, 443–460.

29. R. S. Tsay, *Multivariate Time Series Analysis: With R and Financial Applications*, Hoboken: John Wiley & Sons, 2013.

30. R. C. Scott, P. A. Horvath, On the direction of preference for moments of higher order than the variance, *J. Finance*, **35** (1980), 915–919. https://doi.org/10.2307/2327209

31. R. F. Dittmar, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *J. Finance*, **57** (2002), 369–403. https://doi.org/10.1111/1540-6261.00425

32. Z. Zhu, N. Zhang, K. Zhu, Big portfolio selection by graph-based conditional moments method, *J. Empir. Finance*, **78** (2024), 101533. https://doi.org/10.1016/j.jempfin.2024.101533

33. J. Fan, Y. Liao, M. Mincheva, High dimensional covariance matrix estimation in approximate factor models, *Ann. Stat.*, **39** (2011), 3320. https://doi.org/10.1214/11-AOS944

34. X. Li, X. Zhang, Y. Li, High-dimensional conditional covariance matrices estimation using a factor-garch model, *Symmetry*, **14** (2022), 158. https://doi.org/10.3390/sym14010158

Appendix

Before proving Theorem 1, we first need to present some propositions, which follow from the Lemma A.1 of Chen et al. [6].

Proposition 1. *Under Assumptions A1–A6, we have*

$$(i)\quad \frac{1}{T}\bar{\mathbf{X}}^{(r)\mathcal{T}}\bar{\mathbf{X}}^{(r)} = \frac{1}{T}\bar{\mathbf{X}}^{\mathcal{T}}\mathbf{K}(r)\bar{\mathbf{X}} = \mathbf{D}_r\mathbf{\Sigma}_F\mathbf{D}_r^{\mathcal{T}} + O_p\left(\frac{1}{\sqrt{Th}} + \frac{1}{M}\right),$$

$$(ii)\quad \|\bar{\mathbf{X}}^{(r)}\| = O_p(\sqrt{T}),\quad \|(\bar{\mathbf{X}}^{(r)\mathcal{T}}\bar{\mathbf{X}}^{(r)})^{+}\| = O_p\left(\frac{1}{T}\right),$$

$$(iii)\quad \frac{1}{T}\mathbf{F}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{F}^{(r)} = \mathbf{\Sigma}_F + O_p\left(\frac{1}{\sqrt{Th}} + \frac{1}{M}\right),\quad \|\mathbf{F}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\| = O_p(\sqrt{T}),$$

$$(iv)\quad \|\mathbf{X}^{(r)\mathcal{T}}\mathbf{u}_i^{(r)}\|^2 = O_p\left(\frac{T}{h}\right),\quad \|\mathbf{P}^{(r)}\mathbf{U}^{(r)}\|_F^2 = O_p\left(\frac{N}{h}\right),$$

where $\mathbf{A}^{+}$ *denotes the Moore Penrose inverse of* $\mathbf{A}$.

Proposition 2. *Under Assumptions A1–A6, we have*

$$(i)\quad \frac{\lambda_j[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]}{\lambda_{j+1}[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]} = O_p(1),\quad when\quad j < R,$$

$$(ii)\quad \frac{\lambda_j[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]}{\lambda_{j+1}[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]} \to \infty,\quad when\quad j = R.$$

Proof of Theorem 1. Given Propositions 1 and 2, it is not difficult to accomplish the proof if we can show that

$$\frac{\lambda_{R+j}[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]}{\lambda_{R+j+1}[(NT)^{-1}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}]} = O_p(1),\quad \text{when}\quad j < d - R,$$

which follows from Theorem 4.2 in Chen et al. [6]. □

Let $\mathbf{V}_{NT}$ be the diagonal matrix whose diagonal is made up of the largest R eigenvalues of matrix $(NT)^{-1}\mathbf{Y}^{(r)T}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}$. By the definition of eigenvalues, we have

$$\frac{1}{TN}\mathbf{Y}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}\hat{\mathbf{\Lambda}}_r = \hat{\mathbf{\Lambda}}_r\mathbf{V}_{NT}.$$

Let

$$\mathbf{H}^{(r)} = \frac{1}{TN}\mathbf{F}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{F}^{(r)}\mathbf{\Lambda}_r^{\mathcal{T}}\hat{\mathbf{\Lambda}}_r\mathbf{V}_{NT}^{-1}.$$

Substituting (2.9), we have

$$\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\mathbf{H}^{(r)} = \left(\sum_{i=1}^{3}\mathbf{A}_i\right)\mathbf{V}_{NT}^{-1},$$

where

$$\mathbf{A}_1 = \frac{1}{TN}\mathbf{\Lambda}_r\mathbf{F}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{U}^{(r)}\hat{\mathbf{\Lambda}}_r,$$

$$\mathbf{A}_2 = \frac{1}{TN}\mathbf{U}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{U}^{(r)}\hat{\mathbf{\Lambda}}_r,$$

$$\mathbf{A}_3 = \frac{1}{TN}\mathbf{U}^{(r)\mathcal{T}}\mathbf{P}^{(r)}\mathbf{F}^{(r)}\mathbf{\Lambda}_r^{\mathcal{T}}\hat{\mathbf{\Lambda}}_r.$$

Now we give the convergence rates of the above terms, which follow Lemma A.2 in Chen et al. [6].

Proposition 3. *Under Assumptions A1–A6, we have*

$$(i)\, \|\mathbf{A}_1\|_F = O_p(\frac{N}{Th}), \quad (ii)\, \|\mathbf{A}_2\|_F = O_p(\frac{N}{T^2h^2}), \quad (iii)\, \|\mathbf{A}_3\|_F = O_p(\frac{N}{Th}).$$

Given Proposition 3, the convergence rate of $\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\mathbf{H}^{(r)}$ can be straightforwardly calculated such that

$$\frac{1}{N}\|\hat{\mathbf{\Lambda}}_r - \mathbf{\Lambda}_r\mathbf{H}^{(r)}\|_F^2 = O_p\left(\frac{1}{Th}\right).$$

Proofs of Theorems 2 and 3. By Propositions 1–3, the proofs of Theorems 2 and 3 can be similarly summarized to Theorem 4.3 and Theorem 4.4 in Chen et al. [6]. □

Proposition 4. *Under Assumptions A1–A8, there exist constants $C_5 > 0$ and a small constant ϵ defined in Assumption A8 (ii), such that*

$$P\left\{\sup_{t\le T, i\le N} |\hat{u}_{it} - u_{it}| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Proof.

$$\begin{aligned}
&P\left\{\sup_{t\le T, i\le N} |\hat{u}_{it} - u_{it}| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\}\\
&= P\left\{\sup_{t\le T, i\le N} |\boldsymbol{\lambda}_{ir}^{\mathcal{T}}\mathbf{f}_t - \hat{\boldsymbol{\lambda}}_{ir}^{\mathcal{T}}\hat{\mathbf{f}}_t| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\}\\
&= P\left\{\sup_{t\le T, i\le N} |(\boldsymbol{\lambda}_{ir} - \hat{\boldsymbol{\lambda}}_{ir})^{\mathcal{T}}\mathbf{f}_t + \hat{\boldsymbol{\lambda}}_{ir}^{\mathcal{T}}(\mathbf{f}_t - \hat{\mathbf{f}}_t)| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\}\\
&\le P\left\{\sup_{i\le N}\|\hat{\boldsymbol{\lambda}}_{ir} - \boldsymbol{\lambda}_{ir}\|\sup_{t\le T}\|\mathbf{f}_t\| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\}\\
&\quad + P\left\{\sup_{i\le N}\|\hat{\boldsymbol{\lambda}}_{ir}\|\sup_{t\le T}\|\hat{\mathbf{f}}_t - \mathbf{f}_t\| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\}.
\end{aligned}$$

By Assumption A1 and Theorem 2, we have

$$P\left\{\sup_{i\le N}\|\hat{\boldsymbol{\lambda}}_{ir} - \boldsymbol{\lambda}_{ir}\|\sup_{t\le T}\|\mathbf{f}_t\| > C_5\left(\frac{1}{\sqrt{Th}}\right)\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Since the factor loadings estimator $\hat{\mathbf{\Lambda}}_r$ for each $r \le T$ can be described as the first R principal components of $(NT)^{-1}\mathbf{Y}^{(r)\top}\mathbf{P}^{(r)}\mathbf{Y}^{(r)}$ arranged in descending order, we have $N^{-1}\hat{\mathbf{\Lambda}}_r^{\mathcal{T}}\hat{\mathbf{\Lambda}}_r = \mathbf{I}_R$, implying that $\sup_{i\le N}\|\hat{\boldsymbol{\lambda}}_{ir}\|$ is bounded. Hence, by Assumption A7, this lemma follows. □

Proposition 5. *Under Assumptions A1–A8, there exist constants $C_6 > 0$ and a small constant ϵ defined in Assumption A8 (ii), such that*

$$P\left\{\|\hat{\mathbf{\Sigma}}_{\hat{f}}(t) - \mathbf{\Sigma}_f(t)\|_F^2 > \frac{C_6}{Th} + \frac{C_6\log T}{T}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Proof. Define

$$\hat{\boldsymbol{\Sigma}}_f(t) \equiv \hat{\mathrm{cov}}(\mathbf{f}_t) = (T-1)^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{F} - [T(T-1)]^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{1}\mathbf{1}^{\mathcal{T}}\mathbf{F},$$

where $\mathbf{1}$ is a T-dimensional column vector of ones. It is easy to derive that

$$\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) - \boldsymbol{\Sigma}_f(t)\|_F^2 = \|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) - \hat{\boldsymbol{\Sigma}}_f(t) + \hat{\boldsymbol{\Sigma}}_f(t) - \boldsymbol{\Sigma}_f(t)\|_F^2 \le \|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) - \hat{\boldsymbol{\Sigma}}_f(t)\|_F^2 + \|\hat{\boldsymbol{\Sigma}}_f(t) - \boldsymbol{\Sigma}_f(t)\|_F^2.$$

For the first part $\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) - \hat{\boldsymbol{\Sigma}}_f(t)\|_F^2$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t) - \hat{\boldsymbol{\Sigma}}_f(t)\|_F^2 \le\ & \|(T-1)^{-1}\hat{\mathbf{F}}^{\mathcal{T}}\hat{\mathbf{F}} - (T-1)^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{F}\|_F^2 \\
& + \|[T(T-1)]^{-1}\hat{\mathbf{F}}^{\mathcal{T}}\mathbf{1}\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}} - [T(T-1)]^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{1}\mathbf{1}^{\mathcal{T}}\mathbf{F}\|_F^2 \\
=\ & \mathrm{I}_1 + \mathrm{I}_2.
\end{aligned}$$

By Assumption A1, $\|(T-1)^{-1/2}\mathbf{F}\|_F^2 = \mathrm{tr}[(T-1)^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{F}]$ is bounded, leading to the result that $\|(T-1)^{-1/2}\mathbf{F}\|_F^2 = O_p(1)$ and

$$\|(T-1)^{-1/2}\hat{\mathbf{F}}\|_F^2 \le (T-1)^{-1}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2 + \|(T-1)^{-1/2}\mathbf{F}\|_F^2 = O_p(1).$$

Hence,

$$\begin{aligned}
\mathrm{I}_1 &= \|(T-1)^{-1}\hat{\mathbf{F}}^{\mathcal{T}}\hat{\mathbf{F}} - (T-1)^{-1}\mathbf{F}^{\mathcal{T}}\mathbf{F}\|_F^2 \\
&= \|(T-1)^{-1}(\hat{\mathbf{F}} - \mathbf{F})^{\mathcal{T}}\hat{\mathbf{F}} + (T-1)^{-1}\mathbf{F}^{\mathcal{T}}(\hat{\mathbf{F}} - \mathbf{F})\|_F^2 \\
&\le (T-1)^{-1}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2\|(T-1)^{-1/2}\hat{\mathbf{F}}\|_F^2 \\
&\quad + (T-1)^{-1}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2\|(T-1)^{-1/2}\mathbf{F}\|_F^2 \\
&= O_p\left(\frac{1}{Th}\right).
\end{aligned}$$

For I_2, since

$$\|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}\mathbf{F}\|_F^2 \le T^{-1}\|\mathbf{1}^{\mathcal{T}}\|_F^2\|(T-1)^{-1/2}\mathbf{F}\|_F^2 = O_p(1),$$

and

$$\begin{aligned}
\|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}}\|_F^2 &\le \|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}(\hat{\mathbf{F}} - \mathbf{F})\|_F^2 \\
&\quad + \|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}\mathbf{F}\|_F^2 \\
&\le T^{-1}\|\mathbf{1}^{\mathcal{T}}\|_F^2(T-1)^{-1}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2 + O_p(1) \\
&= O_p(1),
\end{aligned}$$

we can calculate I_2 similar to I_1, such that

$$\begin{aligned}
\mathrm{I}_2 &= \|[T(T-1)]^{-1}(\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}})^{\mathcal{T}}\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}} - [T(T-1)]^{-1}(\mathbf{1}^{\mathcal{T}}\mathbf{F})^{\mathcal{T}}\mathbf{1}^{\mathcal{T}}\mathbf{F}\|_F^2 \\
&= \|[T(T-1)]^{-1}(\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}} - \mathbf{1}^{\mathcal{T}}\mathbf{F})^{\mathcal{T}}\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}} \\
&\quad + [T(T-1)]^{-1}(\mathbf{1}^{\mathcal{T}}\mathbf{F})^{\mathcal{T}}(\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}} - \mathbf{1}^{\mathcal{T}}\mathbf{F})\|_F^2 \\
&\le (T-1)^{-1}\|\hat{\mathbf{F}} - \mathbf{F}\|_F^2 T^{-1}\|\mathbf{1}\|_F^2\|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}\hat{\mathbf{F}}\|_F^2
\end{aligned}$$

$$+ (T-1)^{-1}\|\hat{\mathbf{F}}-\mathbf{F}\|_F^2 T^{-1}\|\mathbf{1}^{\mathcal{T}}\|_F^2\|[T(T-1)]^{-1/2}\mathbf{1}^{\mathcal{T}}\mathbf{F}\|_F^2$$
$$= O_p\left(\frac{1}{Th}\right).$$

Note that $\frac{1}{T}\|\hat{\mathbf{F}}-\mathbf{F}\|_F^2 = \frac{1}{T}\sum_{t=1}^T \|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2 \leq \sup_{t\leq T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2$. Hence, by Assumption A7, we can proof the convergence rate of the first part $\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t)-\hat{\boldsymbol{\Sigma}}_f(t)\|_F^2$ such that

$$P\left\{\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t)-\hat{\boldsymbol{\Sigma}}_f(t)\|_F^2 > \frac{C_6}{Th}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Since f_{tj} has an exponential tail for any $t \leq T$ and $j \leq R$, we can proof the convergence rate of the second part $\|\hat{\boldsymbol{\Sigma}}_f(t)-\boldsymbol{\Sigma}_f(t)\|_F^2$ in a similar fashion to Lemma B.2 in Fan et al. [33] such that

$$P\left\{\|\hat{\boldsymbol{\Sigma}}_f(t)-\boldsymbol{\Sigma}_f(t)\|_F^2 > \frac{C_6\log T}{T}\right\} = O\left(\frac{1}{T^2}\right).$$

Hence,

$$P\left\{\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t)-\boldsymbol{\Sigma}_f(t)\|_F^2 > \frac{C_6}{Th}+\frac{C_6\log T}{T}\right\}$$
$$\leq P\left\{\|\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t)-\hat{\boldsymbol{\Sigma}}_f(t)\|_F^2 > \frac{C_6}{Th}\right\} + P\left\{\|\hat{\boldsymbol{\Sigma}}_f(t)-\boldsymbol{\Sigma}_f(t)\|_F^2 > \frac{C_6\log T}{T}\right\}$$
$$= O\left(\frac{1}{T^{1+\epsilon}}\right).$$

□

Proposition 6. *Under Assumptions A1–A8, there exist constants* $C_8 > 0$ *and a small constant* ϵ *defined in Assumption A8 (ii), such that*

$$P\left\{\sup_{\ell\leq p}\|\hat{\boldsymbol{\theta}}_\ell-\boldsymbol{\theta}_\ell\| > C_8\frac{1}{\sqrt{Th}}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Proof. Note that by Proposition 4, $\sup_{t\leq T,i\leq N}|\hat{u}_{it}-u_{it}| = O_p(\frac{1}{\sqrt{Th}})$. The proof of this lemma can be similarly summarized to Theorem 1(IV) in Guo et al. [15]. □

Proof of Theorem 4. By Proposition 6, it can be similarly demonstrated as in Theorem 2 in Li et al. [17] that there exist constants $C_9 > 0$, such that

$$P\left\{\left\|\hat{\boldsymbol{\Sigma}}_u(t)-\boldsymbol{\Sigma}_u(t)\right\|_\Sigma^2 > \frac{C_9}{Th}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Define

$$\mathbf{C}_{Tr} = \hat{\boldsymbol{\Lambda}}_r - \boldsymbol{\Lambda}_r, \quad \mathbf{G}_T = \hat{\boldsymbol{\Sigma}}_{\hat{f}} - \boldsymbol{\Sigma}_f.$$

Note that for each $r \leq T$, $\frac{1}{N}\|\mathbf{C}_{Tr}\|_F^2 = \frac{1}{N}\sum_{i=1}^N\|\hat{\boldsymbol{\lambda}}_{ir}-\boldsymbol{\lambda}_{ir}\|^2 \leq \sup_{i\leq N}\|\hat{\boldsymbol{\lambda}}_{ir}-\boldsymbol{\lambda}_{ir}\|^2$. It is easy to obtain

$$\left\|\hat{\boldsymbol{\Sigma}}_y(t)-\boldsymbol{\Sigma}_y(t)\right\|_\Sigma^2 \leq \left\|\left(\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Sigma}}_{\hat{f}}(t)\hat{\boldsymbol{\Lambda}}^{\mathcal{T}} - \boldsymbol{\Lambda}\boldsymbol{\Sigma}_f(t)\boldsymbol{\Lambda}^{\mathcal{T}}\right) + \left(\hat{\boldsymbol{\Sigma}}_u(t)-\boldsymbol{\Sigma}_u(t)\right)\right\|_\Sigma^2$$

$$\leq 8\left\|\mathbf{C}_{\mathcal{T}}\hat{\mathbf{\Sigma}}_{\hat{f}}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2} + 8\left\|\mathbf{\Lambda}\mathbf{G}_{\mathcal{T}}\mathbf{\Lambda}^{\mathcal{T}}\right\|_{\Sigma}^{2} + 16\left\|\mathbf{\Lambda}\hat{\mathbf{\Sigma}}_{\hat{f}}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2}$$
$$+ 2\left\|\hat{\mathbf{\Sigma}}_{u}(t) - \mathbf{\Sigma}_{u}(t)\right\|_{\Sigma}^{2}.$$

Similar to the proof of Lemma B.3 in Li et al. [34], there exist constants $C_{10} > 0$, such that

$$\left\|\mathbf{C}_{\mathcal{T}}\hat{\mathbf{\Sigma}}_{\hat{f}}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2} \leq O_p(N^{-1})\|\mathbf{C}_{\mathcal{T}r}\|_F^4 = O_p\left(\frac{C_{10}N}{(Th)^2}\right),$$

$$\left\|\mathbf{\Lambda}\mathbf{G}_{\mathcal{T}}\mathbf{\Lambda}^{\mathcal{T}}\right\|_{\Sigma}^{2} \leq O_p(N^{-1})\|\mathbf{G}_{\mathcal{T}}\|_F^2 = O_p\left(\frac{C_{10}}{NTh} + \frac{C_{10}\log T}{NT}\right),$$

$$\left\|\mathbf{\Lambda}\hat{\mathbf{\Sigma}}_{\hat{f}}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2} \leq O_p(N^{-1})\|\mathbf{C}_{\mathcal{T}r}\|_F^2 = O_p\left(\frac{C_{10}}{Th}\right),$$

leading to the result that

$$P\left\{\left\|\mathbf{C}_{\mathcal{T}}\hat{\mathbf{\Sigma}}_{f}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2} > \frac{C_{10}N}{(Th)^2}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right),$$

and

$$P\left\{\left\|\mathbf{\Lambda}\mathbf{G}_{\mathcal{T}}\mathbf{\Lambda}^{\mathcal{T}}\right\|_{\Sigma}^{2} + \left\|\mathbf{\Lambda}\hat{\mathbf{\Sigma}}_{f}(t)\mathbf{C}_{\mathcal{T}}^{\mathcal{T}}\right\|_{\Sigma}^{2} > \frac{C_{10}}{Th} + \frac{C_{10}\log T}{NT}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Hence, there exist constants $C_4 > 0$ and a small constant ϵ defined in Assumption A8 (ii), such that

$$P\left\{\left\|\hat{\mathbf{\Sigma}}_{y}(t) - \mathbf{\Sigma}_{y}(t)\right\|_{\Sigma}^{2} > \frac{C_4N}{(Th)^2} + \frac{C_4\log T}{NT}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

□

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)