

# Low-Rank Structured Nonparametric Prediction of Instantaneous Volatility

Sung Hoon Choi\*      Donggyu Kim†

July 2025

## Abstract

Based on Itô semimartingale models, several studies have proposed methods for forecasting intraday volatility using high-frequency financial data. These approaches typically rely on restrictive parametric assumptions and are often vulnerable to model misspecification. To address this issue, we introduce a novel nonparametric prediction method for the future intraday instantaneous volatility process during trading hours, which leverages both previous days' data and the current day's observed intraday data. Our approach imposes an interday-by-intraday matrix representation of the instantaneous volatility, which is decomposed into a low-rank conditional expectation component and a noise matrix. To predict the future conditional expected volatility vector, we exploit this low-rank structure and propose the Structural Intraday-volatility Prediction (SIP) procedure. We establish the asymptotic properties of the SIP estimator and demonstrate its effectiveness through an out-of-sample prediction study using real high-frequency trading data.

**Key words:** Diffusion process, high-frequency financial data, low-rank matrix, matrix completion.

---

\*Department of Economics, University of Connecticut, Storrs, CT 06269, USA. E-mail: [sung\\_hoon.choi@uconn.edu](mailto:sung_hoon.choi@uconn.edu).

†Department of Economics, University of California, Riverside, CA 92521, USA. Email: [donggyu.kim@ucr.edu](mailto:donggyu.kim@ucr.edu).

# 1 Introduction

The analysis of volatility is essential in financial econometrics and statistics, and it has broad applications in hedging, option pricing, risk management, and portfolio allocation. In recent years, the widespread availability of high-frequency financial data has led to the development of several nonparametric integrated volatility estimation methods ([Aït-Sahalia et al., 2010](#); [Barndorff-Nielsen et al., 2008, 2011](#); [Bibinger et al., 2014](#); [Christensen et al., 2010](#); [Fan and Kim, 2018](#); [Fan and Wang, 2007](#); [Jacod et al., 2009](#); [Xiu, 2010](#); [Zhang et al., 2005](#); [Zhang, 2006, 2011](#)). These high-frequency-based estimators have significantly improved our understanding of low-frequency (interday) market dynamics and have motivated the development of various conditional volatility models built on realized measures ([Andersen et al., 2003](#); [Corsi, 2009](#); [Hansen et al., 2012](#); [Kim and Wang, 2016b](#); [Shephard and Sheppard, 2010](#)). Parallel to these developments, nonparametric procedures for estimating instantaneous (or spot) volatility have been proposed to capture intraday volatility dynamics ([Fan and Wang, 2008](#); [Figuerola-López and Wu, 2024](#); [Foster and Nelson, 1996](#); [Kristensen, 2010](#); [Mancini et al., 2015](#); [Todorov, 2019](#); [Todorov and Zhang, 2023](#); [Zu and Boswijk, 2014](#)). Leveraging these well-performing estimators, several studies have illustrated U-shaped or reverse J-shaped intraday volatility patterns across various markets ([Admati and Pfleiderer, 1988](#); [Andersen and Bollerslev, 1997](#); [Andersen et al., 2019](#); [Hong and Wang, 2000](#); [Li and Linton, 2023](#)).

In contrast to the extensive literature on predicting daily volatility processes, relatively less attention has been given to forecasting intraday volatility. For example, [Engle and Sokalska \(2012\)](#) proposed a modified GARCH model in which the conditional variance is expressed as the product of daily, diurnal, and stochastic intraday components to capture intraday volatility dynamics, while [Zhang et al. \(2024\)](#) explored several nonparametric machine learning methods for forecasting intraday realized volatility, leveraging the commonality observed in intraday volatility patterns. Recently, [Choi and Kim \(2025\)](#) considered the interday-by-intraday instantaneous volatility matrix process and proposed a semiparametric

prediction procedure for the one-day-ahead instantaneous volatility process by leveraging both interday and intraday volatility dynamics. However, existing methods for forecasting intraday volatility suffer from notable limitations. For example, many methods rely on strong distributional assumptions inherent in parametric models and often fail to incorporate interday dynamics effectively. In practice, practitioners are often interested in forecasting the remaining future intraday volatility process conditional on the observed intraday data up to the present time. Therefore, it is both practically relevant and methodologically important to develop a prediction procedure that adapts to current intraday information and is robust to potential model misspecification.

This paper introduces a novel nonparametric prediction approach based on a low-rank structure for forecasting the remaining future instantaneous volatility process during intraday trading. Specifically, we represent the entire instantaneous volatility process as a matrix, where each row corresponds to a day and each column represents an intraday time point. We impose a low-rank plus noise structure on this matrix. The parameter of interest is the instantaneous volatility process after the current intraday time. That is, on the current day (denoted as the  $D$ th day), only the initial portion of the row is observed, while the remaining part is the target of prediction. To forecast this remaining portion, we exploit the low-rank matrix structure. For example, the future instantaneous volatility vector can be represented by the singular components of the observed part of the low-rank matrix. We use this structure to construct a nonparametric prediction of the future volatility vector. This method is called the Structural Intraday-volatility Prediction (SIP) procedure. It is worth noting that as long as the interday-by-intraday instantaneous volatility matrix exhibits a low-rank structure, the SIP method can effectively predict the remaining future volatility vector without relying on any specific parametric dynamics such as autoregressive models. Therefore, the SIP procedure is robust to model misspecification arising from incorrect parametric assumptions. Furthermore, we derive the convergence rate of the predicted instantaneous volatility under the SIP framework and assess its empirical performance via an out-of-sample prediction study

using high-frequency financial data. The SIP method consistently outperforms alternative approaches in terms of both out-of-sample prediction accuracy and Value-at-Risk (VaR) forecasting performance.

The remainder of the paper is organized as follows. In Section 2, we introduce the model and propose the SIP prediction procedure. Section 3 presents the asymptotic properties of the SIP estimator. In Section 4, we evaluate the finite sample performance of the proposed method through simulation. Section 5 applies the proposed method to real high-frequency financial data to forecast the remaining future instantaneous volatility process during trading hours. Finally, the conclusion is provided in Section 6. All technical proofs are provided in the Appendix A.

## 2 Model Setup and Estimation Procedure

Throughout this paper, we denote by  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{A}\|_2$  (or  $\|\mathbf{A}\|$  for short),  $\|\mathbf{A}\|_1$ ,  $\|\mathbf{A}\|_\infty$ , and  $\|\mathbf{A}\|_{\max}$  the Frobenius norm, operator norm,  $l_1$ -norm,  $l_\infty$ -norm, and element-wise norm, which are defined, respectively, as  $\|\mathbf{A}\|_F = \text{tr}^{1/2}(\mathbf{A}^\top \mathbf{A})$ ,  $\|\mathbf{A}\|_2 = \lambda_{\max}^{1/2}(\mathbf{A}^\top \mathbf{A})$ ,  $\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}|$ ,  $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$ , and  $\|\mathbf{A}\|_{\max} = \max_{i,j} |a_{ij}|$ . When  $\mathbf{A}$  is a vector, the maximum norm is denoted as  $\|\mathbf{A}\|_\infty = \max_i |a_i|$ , and both  $\|\mathbf{A}\|$  and  $\|\mathbf{A}\|_F$  are equal to the Euclidean norm.

### 2.1 A Model Setup

We consider the following jump diffusion process: for the  $i$ -th day and intraday time  $t \in [0, 1]$ ,

$$dX_{i,t} = \mu_{i,t}dt + \sigma_{i,t}dB_{i,t} + J_{i,t}dP_{i,t}, \quad (2.1)$$

where  $X_{i,t}$  is the log price of an asset,  $\mu_{i,t}$  is a drift process,  $B_{i,t}$  is a one-dimensional standard Brownian motion,  $J_{i,t}$  is the jump size, and  $P_{i,t}$  is the Poisson process with the intensity  $\mu_j$ . For a given intraday time sequence, for each  $i = 1, \dots, D$  and  $j = 1, \dots, n$ , we denote

the instantaneous volatility process as  $c_{i,j} := \sigma_{i,t_j}^2$ , where  $0 < t_1 < \dots < t_n = 1$ . Then, we assume that the discrete-time instantaneous volatility process satisfies the following low-rank structure:

$$\boldsymbol{\Sigma}_{D,n} = (c_{i,j})_{D \times n} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top + \mathbf{E} \equiv \mathbf{A} + \mathbf{E}, \quad (2.2)$$

where  $\mathbf{U} = (u_{i,k})_{i=1,\dots,D,k=1,\dots,r}$  is the left singular vector matrix,  $\mathbf{V} = (v_{j,k})_{j=1,\dots,n,k=1,\dots,r}$  is the right singular vector matrix,  $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_r)$  is the singular value matrix, and  $\mathbf{E} = (e_{i,j})_{D \times n}$  is the random noise matrix. We only impose the above low-rank structure in (2.2) to predict future remaining intraday volatility. It allows for considerable modeling flexibility and can implicitly incorporate various time-series features. For example, the left singular vector can represent interday time-series dynamics, while the right singular vector can account for intraday patterns, such as U-shaped or reverse J-shaped (Choi and Kim, 2025). It is worth noting that, unlike Choi and Kim (2025), we do not assume any explicit dynamic or parametric form for the volatility evolution. We note that  $\boldsymbol{\Sigma}_{D,n}$  is a  $D \times n$  matrix, which is distinct from a covariance matrix and contains only positive elements.

In this paper, our target is to predict the remaining future instantaneous volatility vector for the  $D$ th day. Specifically, we assume that trading data is available up to a fraction  $\omega = \frac{n_1}{n}$  of the  $D$ th day. The objective is to predict the remaining  $n_2 \times 1$  instantaneous vector,  $(c_{D,n_1+1}, \dots, c_{D,n})$ , where  $n = n_1 + n_2$ . Given the current information  $\mathcal{F}_{D,n_1}$ , we can predict the instantaneous volatility as follows: for  $j = n_1 + 1, \dots, n$ ,

$$E[\sigma_{D,j}^2 | \mathcal{F}_{D,n_1}] = \sum_{k=1}^r \lambda_k u_{D,k} v_{j,k} \quad \text{a.s.} \quad (2.3)$$

We can write the model (2.2) in a partitioned matrix form as follows:

$$\boldsymbol{\Sigma}_{D,n} = \begin{bmatrix} & n_1 & n_2 \\ \Sigma_{11} & \Sigma_{12} & \\ \Sigma_{21} & \Sigma_{22} & \end{bmatrix} \begin{matrix} D-1 \\ 1 \end{matrix}, \mathbf{A} = \begin{bmatrix} & n_1 & n_2 \\ A_{11} & A_{12} & \\ A_{21} & A_{22} & \end{bmatrix} \begin{matrix} D-1 \\ 1 \end{matrix}, \mathbf{E} = \begin{bmatrix} & n_1 & n_2 \\ E_{11} & E_{12} & \\ E_{21} & E_{22} & \end{bmatrix} \begin{matrix} D-1 \\ 1 \end{matrix}.$$

We note that  $\Sigma_{11}$ ,  $\Sigma_{12}$ , and  $\Sigma_{21}$  are estimable using high-frequency log-price observations. Since  $\mathbf{A}$  is the low rank matrix,  $A_{22}$  is given by

$$A_{22} = A_{21}(A_{11})^\dagger A_{12} = A_{21}(V_{11}\Lambda_{11}^{-1}U_{11}^\top)A_{12}, \quad (2.4)$$

where  $U_{11}$  is the  $(D - 1) \times r$  left singular vector matrix,  $V_{11}$  is the  $n_1 \times r$  right singular vector matrix, and  $\Lambda_{11}$  is the  $r \times r$  singular value matrix of  $A_{11}$ . Since  $A_{11}$ ,  $A_{21}$ , and  $A_{12}$  are quantities related to the observed period, using this relationship (2.4), we can predict  $A_{22}$  without a specific parametric modeling assumption. Specifically, as long as we can estimate  $A_{11}$ ,  $A_{21}$ , and  $A_{12}$  well using the available data,  $A_{22}$  can be estimated using the plug-in scheme. We note that if we treat the remaining intraday volatility vector  $A_{22}$  as missing components, this problem becomes similar to the matrix completion (Cai et al., 2010; Candes and Recht, 2012; Candes and Plan, 2010). Especially in this problem, the missing has a determined structure as in Bai and Ng (2021); Cai et al. (2016); Fan and Kim (2019); Yan and Wainwright (2024). From this point of view, this paper extends the matrix completion concept to predict time series patterns.

Due to the imperfections in the trading process, practitioners cannot directly observe the true underlying log-stock price  $X_{i,t}$  in (2.1) (Aït-Sahalia and Yu, 2009). To account for this, we assume that the high-frequency intraday observations  $X_{i,t_s}$ ,  $s = 1, \dots, m$ , are contaminated by microstructure noise as follows:

$$Y_{i,t_s} = X_{i,t_s} + e_{i,t_s}, \quad i = 1, \dots, D, s = 1, \dots, m, \quad (2.5)$$

where  $e_{i,t_s}$  is the microstructure noise process with a mean of zero and a variance of  $\eta_{ii}$ .

## 2.2 Structural Intraday-volatility Prediction

To predict the remaining future instantaneous volatility vector  $A_{22}$  nonparametrically, we use the low-rank structural equation (2.4). Since  $A_{11}$ ,  $A_{12}$ , and  $A_{21}$  are latent low-rank matrices,

we need to estimate them using a well-performing instantaneous volatility estimator based on the observed high-frequency data. For example, we need a large dimension to estimate the latent low-rank matrix and use the blessing of dimensionality (Li et al., 2018). To do this, we choose the large block matrices such as  $(\Sigma_{11}, \Sigma_{12})$  and  $(\Sigma_{11}^\top, \Sigma_{21}^\top)$ . On the other hand, we employ a well-performing nonparametric instantaneous volatility estimator to estimate the instantaneous volatility. Details can be found in Remark 2.1 below. The specific estimation procedure is as follows:

1. Using high-frequency log-price observations, we compute the instantaneous volatility estimators  $\widehat{c}_{i,j}$  up to the  $D$ th day  $n_1$  intraday time point (see Remark 2.1). Then, we define the following submatrices:  $\widehat{\Sigma}_{11} = (\widehat{c}_{i,j})_{(D-1) \times n_1}$ ,  $\widehat{\Sigma}_{12} = (\widehat{c}_{i,j})_{(D-1) \times n_2}$ , and  $\widehat{\Sigma}_{21} = (\widehat{c}_{i,j})_{1 \times n_1}$ , which are the initial estimated matrices for  $\Sigma_{11}, \Sigma_{12}$ , and  $\Sigma_{21}$ , respectively.
2. We conduct the singular value decomposition to the large submatrices  $\widehat{\Sigma}_{1\bullet} \equiv [\widehat{\Sigma}_{11} \ \widehat{\Sigma}_{12}]$  and  $\widehat{\Sigma}_{\bullet 1} \equiv [\widehat{\Sigma}_{11}^\top \ \widehat{\Sigma}_{21}^\top]^\top$ . The columns of  $\widehat{U}_{11}$  are defined as the  $r$  leading left singular vector of  $\widehat{\Sigma}_{1\bullet}$ , and the columns of  $\widehat{V}_{11}$  are defined as the  $r$  leading right singular vector of  $\widehat{\Sigma}_{\bullet 1}$ . Note that  $\widehat{U}_{11}$  and  $\widehat{V}_{11}$  are estimators for the orthonormal basis of column vectors of  $U_{11}$  and  $V_{11}$  defined in (2.4), respectively.
3. Finally, we estimate the conditional expectation of the remaining instantaneous volatility vector on the  $D$ th day,  $E[(c_{D,n_1+1}, \dots, c_{D,n}) | \mathcal{F}_{D,n_1}]$ , by

$$\widetilde{\Sigma}_{22} := (\widetilde{c}_{D,n_1+1}, \dots, \widetilde{c}_{D,n}) = \widehat{\Sigma}_{21} \widehat{V}_{11} (\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11})^{-1} \widehat{U}_{11}^\top \widehat{\Sigma}_{12}.$$

**Remark 2.1.** For the SIP procedure, we first need to estimate the instantaneous volatilities using the observed data. In the high-frequency literature, several nonparametric methods for instantaneous volatility estimation have been proposed (Fan and Wang, 2008; Figueroa-López and Wu, 2024; Foster and Nelson, 1996; Kristensen, 2010; Mancini et al., 2015; Todorov, 2019; Todorov and Zhang, 2023; Zu and Boswijk, 2014). Any well-performing

estimator of the instantaneous volatility, denoted by  $\widehat{c}_{i,j}$ , that satisfies Assumption 3.1(iii) can be used within our framework. In the numerical study, we adopt the jump-robust pre-averaging method proposed by [Figueroa-López and Wu \(2024\)](#), which is described explicitly in (4.1) in Section 4.

**Remark 2.2.** To implement the SIP procedure, we need to select the number of factors. The number of latent factors,  $r$ , can be determined using data-driven methods ([Ahn and Horenstein, 2013](#); [Bai and Ng, 2002](#); [Onatski, 2010](#)). For instance,  $r$  can be chosen by maximizing the singular value ratio or the largest singular value gap, given by  $\max_{k \leq r_{\max}} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_{k+1}}$  or  $\max_{k \leq r_{\max}} (\widehat{\lambda}_k - \widehat{\lambda}_{k+1})$ , where  $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_{r_{\max}}$  are the leading singular values of  $\widehat{\Sigma}_{1\bullet} \equiv (\widehat{c}_{i,j})_{(D-1) \times n}$  for a predetermined maximum number of factors,  $r_{\max}$ . In this paper, we choose  $r = 1$  in the numerical study (Sections 4 and 5), following the eigenvalue ratio method proposed by [Ahn and Horenstein \(2013\)](#).

In summary, given the estimated instantaneous volatility process, we construct a structured matrix that includes the remaining future intraday volatility vector. To forecast the future intraday volatility vector, we utilize the low-rank structure described above. We refer to this procedure as the Structural Intraday-volatility Prediction (SIP) method. The SIP method nonparametrically predicts the future instantaneous volatility vector using historical trading data along with current intraday market observations. Crucially, by incorporating current intraday information, the prediction is formulated under a low-rank matrix assumption without relying on any specific parametric model, such as an autoregressive structure. This feature makes the SIP procedure robust to model misspecification arising from incorrect parametric assumptions. The numerical results in Sections 4 and 5 demonstrate that the SIP method effectively predicts the remaining future instantaneous volatility vector.

### 3 Asymptotic Properties

In this section, we develop the asymptotic properties of the SIP estimator. To achieve this, the following technical conditions are needed.

**Assumption 3.1.**

(i) For  $k \leq r$ , the eigengap satisfies  $|\lambda_{k+1} - \lambda_k| = O_P(\sqrt{nD})$  and  $\lambda_{r+1} = 0$ .

(ii) For some bounded  $\mu_1$  and  $\mu_2$ , the left and right singular vectors satisfy

$$\max_{i \leq D, k \leq r} |u_{i,k}|^2 \leq \frac{\mu_1}{D}, \text{ and } \max_{j \leq n, k \leq r} |v_{j,k}|^2 \leq \frac{\mu_2}{n}.$$

In addition, for each  $k \leq r$ ,  $u_k = (u_{1,k}, \dots, u_{D,k})^\top$  and  $v_k = (v_{1,k}, \dots, v_{n,k})^\top$  are assumed to be constant vectors.

(iii) For each  $i \leq D$  and  $j \leq n$ , the estimated instantaneous volatility  $\widehat{c}_{i,j}$  satisfies

$$\widehat{c}_{i,j} - c_{i,j} = v_{i,j} + \varsigma_{i,j},$$

where  $v_{i,j}$  follows the martingale difference sequence and  $\varsigma_{i,j}$  is the estimation bias term such that  $E(v_{i,j} | \mathcal{F}_{i,t_j}) = 0$  a.s. For any  $s \leq D$  and  $j' \leq n$ , we assume  $E(c_{i,j'} v_{i,j} | \mathcal{F}_{i,t_j}) = 0$  and  $E(c_{s,j} c_{s,j'} v_{i,j} | \mathcal{F}_{i,t_j}) = 0$  a.s. In addition, for all  $i \neq s$  and any  $j, j' \leq n$ , we assume  $E(v_{i,j} v_{s,j'}) = 0$ . We further assume that  $c_{i,j}$ ,  $v_{i,j}$ , and  $\varsigma_{i,j}$  are sub-Gaussian random variables; that is, there exists a constant  $K > 0$  such that for all  $i \leq D$ ,  $j \leq n$ , and all  $\tau \in \mathbb{R}$ ,

$$E[\exp(\tau \alpha_{i,j})] \leq \exp\left(\frac{K^2 \tau^2}{2}\right) \quad \text{for } \alpha_{i,j} \in \{c_{i,j}, m^{1/8} v_{i,j}, \rho_m^{-1} \varsigma_{i,j}\},$$

where  $\rho_m$  is the convergence rate of the bias term.

(iv) Let  $\mathbf{\Omega}_{e_j, D \times D}$  denote the  $D \times D$  covariance matrix of  $\mathbf{e}_j = (e_{1,j}, \dots, e_{D,j})^\top$  for each

$j = 1, \dots, n$ . Define  $\mathbf{\Omega}_{j,D \times D} = \frac{1}{n_1} U_{\bullet 1} \Lambda_{\bullet 1}^2 U_{\bullet 1}^\top + \mathbf{\Omega}_{e_j, D \times D}$ . The instantaneous volatility matrix,  $\mathbf{\Sigma}_{\bullet 1} \equiv [\Sigma_{11}^\top \Sigma_{21}^\top]^\top$ , satisfies

$$\|\mathbf{\Sigma}_{\bullet 1} \mathbf{\Sigma}_{\bullet 1}^\top - \sum_{j=1}^{n_1} \mathbf{\Omega}_{j,D \times D}\|_{\max} = O_P(\sqrt{n_1 \log D}).$$

Similarly, let  $\mathbf{\Omega}_{e_i, n \times n}$  denote the  $n \times n$  covariance matrix of  $\mathbf{e}_i = (e_{i,1}, \dots, e_{i,n})^\top$  for each  $i = 1, \dots, D$ . Define  $\mathbf{\Omega}_{i,n \times n} = \frac{1}{D} \mathbf{V} \Lambda^2 \mathbf{V}^\top + \mathbf{\Omega}_{e_i, n \times n}$ . The instantaneous volatility matrix,  $\mathbf{\Sigma}$ , satisfies

$$\|\mathbf{\Sigma}^\top \mathbf{\Sigma} - \sum_{i=1}^D \mathbf{\Omega}_{i,n \times n}\|_{\max} = O_P(\sqrt{D \log n}).$$

(v) The covariance matrices  $\mathbf{\Omega}_{e_j, D \times D} = \text{cov}(\mathbf{e}_j)$  for each  $j = 1, \dots, n$  and  $\mathbf{\Omega}_{e_i, n \times n} = \text{cov}(\mathbf{e}_i)$  for each  $i = 1, \dots, D$ , where  $\mathbf{e}_j = (e_{1,j}, \dots, e_{D,j})^\top$  and  $\mathbf{e}_i = (e_{i,1}, \dots, e_{i,n})^\top$ , satisfy

$$\max_{j \leq n} \|\mathbf{\Omega}_{e_j, D \times D}\|_1 = O(\varphi_D) \quad \text{and} \quad \max_{i \leq D} \|\mathbf{\Omega}_{e_i, n \times n}\|_1 = O(\varphi_n),$$

where  $\varphi_D = o(D)$  and  $\varphi_n = o(n)$ , respectively. In addition,  $\|\mathbf{E}\| = O_P(\max(\sqrt{D}, \sqrt{n}))$ .

(vi) There exist positive constants  $C_1$  and  $C_2$  such that, for each  $k \leq r$ ,  $j, l \leq n$ , and  $i, q \leq D$ ,

$$\begin{aligned} |\text{cov}(\sum_{i=1}^D u_{i,k} e_{i,j}, \sum_{i=1}^D u_{i,k} e_{i,l})| &\leq C_1 \varphi_D \rho_1(|j-l|), \quad \text{where} \quad \sum_{h=1}^{\infty} \rho_1(h) < \infty, \\ |\text{cov}(\sum_{j=1}^n v_{j,k} e_{i,j}, \sum_{j=1}^n v_{j,k} e_{q,j})| &\leq C_2 \varphi_n \rho_2(|i-q|), \quad \text{where} \quad \sum_{h=1}^{\infty} \rho_2(h) < \infty. \end{aligned}$$

In addition, for each  $k \leq r$ ,  $j \leq n$ , and  $i \leq D$ , we have  $E[(\sum_{i=1}^D u_{i,k} e_{i,j})^4] = O(\varphi_D^2)$  and  $E[(\sum_{j=1}^n v_{j,k} e_{i,j})^4] = O(\varphi_n^2)$ .

**Remark 3.1.** Assumption 3.1(i) ensures a pervasive assumption, which is essential for the

analysis of low-rank matrix structures (see, e.g., [Candes and Plan, 2010](#); [Cho et al., 2017](#); [Fan et al., 2018a](#)). Given that our instantaneous volatility matrix is of size  $D \times n$ , this eigengap condition suggests that the leading eigenvalues are associated with the low-rank structure scale on the order of  $\sqrt{nD}$ . Assumption 3.1(ii) imposes the conventional incoherence condition, which ensures effective entrywise control in low-rank matrix estimation. Assumption 3.1(iii) is the sub-Gaussian conditions that can be justified under some mild assumptions on the process  $X$ , the microstructure noise, and the kernel function ([Figuerola-López and Wu, 2024](#); [Kim et al., 2016](#); [Kim and Wang, 2016a](#)). Moreover, this condition also holds for heavy-tailed data given the bounded fourth moments condition using a truncated estimation scheme ([Fan and Kim, 2018](#); [Shin et al., 2023](#)). In practice, estimation of instantaneous volatility at time  $t$  typically uses data up to and including time  $t$ , leading to the martingale difference property  $E(v_{i,j}|\mathcal{F}_{i,t_j}) = 0$  almost surely. The resulting instantaneous volatility estimator is asymptotically unbiased, and the bias vanishes at a rate faster than  $m^{-\frac{1}{8}}$ , which is the optimal convergence rate of the instantaneous volatility estimator in the presence of microstructure noise. To justify the effectiveness of the smoothing technique, we require an uncorrelatedness condition:  $E(c_{i,j}'v_{i,j}|\mathcal{F}_{i,t_j}) = E(c_{s,j}c_{s,j}'v_{i,j}|\mathcal{F}_{i,t_j}) = 0$  almost surely. This condition holds when the spot volatility process is adapted to  $\mathcal{F}_{i,t_j}$  and the noise is independent. Such conditions are mild and are commonly satisfied in standard time series settings like ARMA models. Assumption 3.1(iv) is the element-wise convergence condition for analyzing large matrix inference. Under the sub-Gaussian condition and mixing time dependency, we can easily obtain this condition (see [Fan et al., 2018a,b](#); [Vershynin, 2010](#); [Wang and Fan, 2017](#)). Assumption 3.1(v) imposes the conventional condition on the idiosyncratic covariance matrix, such as sparsity, which is commonly assumed in empirical applications ([Boivin and Ng, 2006](#); [Fan et al., 2016](#)). We note that our framework allows for heterogeneity across both the intraday and interday dimensions. Assumption 3.1(vi) requires weak temporal dependence in both the linear and quadratic forms of the projected error terms, as well as bounded fourth moments.

We obtain the following elementwise convergence rate of the predicted instantaneous volatility using the SIP method.

**Theorem 3.1.** *Suppose that  $r$  is fixed,  $\log D = o(n_1)$ ,  $\log n = o(D)$ ,  $\varphi_D \sqrt{n_1} \leq D$ ,  $\varphi_n \sqrt{D} \leq n_1$  and Assumption 3.1 hold. As  $D, n_1, n, m \rightarrow \infty$ , we have*

$$\max_{j \leq n_2} |\tilde{c}_{D, n_1+j} - E[c_{D, n_1+j} | \mathcal{F}_{D, n_1}]| = O_P \left( \rho_m + \frac{1}{m^{1/4}} + \frac{\varphi_D}{D} + \sqrt{\frac{\log D}{n_1}} + \frac{\varphi_n}{n_1} + \sqrt{\frac{\log n}{D}} \right).$$

Theorem 3.1 shows that the proposed SIP estimator consistently predicts the future instantaneous volatility process. The convergence rate of the SIP estimator is given by  $m^{-1/4} + \frac{1}{\sqrt{D}} + \frac{1}{\sqrt{n_1}}$  up to the log order and the sparsity level, when  $\rho_m < m^{-1/4}$ . Each term in this rate corresponds to a distinct component of the estimation challenge:  $D^{-1/2}$  reflects the statistical cost of learning the interday (daily) time-series dynamics;  $n_1^{-1/2}$  captures the cost of estimating intraday volatility patterns;  $m^{-1/4}$  arises from estimating the unobserved instantaneous volatility using high-frequency returns. Notably, while the optimal nonparametric convergence rate for estimating instantaneous volatility is  $m^{-1/8}$ , the SIP method achieves a faster rate of  $m^{-1/4}$ . This improvement stems from the low-rank structure, which allows the SIP estimator to benefit from a law-of-large-numbers-type averaging across the volatility surface. In addition, the term  $\rho_m$  arises from the bias component in the spot volatility estimator. Specifically, the bias originates from the drift term, which typically decays at the rate  $m^{-1}$ ; however, due to the subsampling approach used to mitigate microstructure noise, the effective convergence rate  $\rho_m$  is generally of order  $m^{-1/2}$ .

## 4 Simulation Study

In this section, we examined the finite sample performance of the proposed SIP method through simulations. We generated high-frequency observations that incorporate both interday HAR model feature and the intraday periodic pattern (Choi and Kim, 2025) as follows:

for  $i = 1, \dots, D$ ,  $s = 0, \dots, m$ , and  $t_s = s/m$ ,

$$\begin{aligned} Y_{i,t_s} &= X_{i,t_s} + e_{i,t_s}, \\ dX_{i,t} &= (\mu - \sigma_{i,t}^2/2)dt + \sigma_{i,t}dB_{i,t} + J_{i,t}dP_{i,t}, \\ \sigma_{i,t_s}^2 &= \tilde{\sigma}_i^2 h(t_s) + \varepsilon_{i,t_s}, \end{aligned}$$

where we set microstructure noise as  $e_{i,t_s} \sim \mathcal{N}(0, 0.0005^2)$  and the initial value as  $X_{1,0} = 1$ ;  $B_{i,t}$  is a standard Brownian motion; for the jump part, we set  $J_{i,t} \sim \mathcal{N}(-0.01, 0.02^2)$  and  $P_{i,t+\Delta} - P_{i,t} \sim \text{Poisson}(36\Delta/252)$ ;  $\tilde{\sigma}_i = b_0 + b_1\tilde{\sigma}_{i-1} + b_2\frac{1}{5}\sum_{s=1}^5\tilde{\sigma}_{i-s} + b_3\frac{1}{22}\sum_{s=1}^{22}\tilde{\sigma}_{i-s} + \zeta_i$ ,  $\zeta_i \sim \mathcal{N}(0, 1)$ ,  $h(t_s) = \gamma_0 + \gamma_1(t_s - 0.6)^2$ , and  $\varepsilon_{i,t_s} = q(t_s)\xi_{i,t_s}$ , where  $q(t_s)^2 = 0.1 + 0.5(2t_s - 1)^2$  and  $\xi_{i,t_s} \sim \mathcal{N}(0, 0.01^2)$ . The model parameters were set to be

$$\begin{aligned} \mu &= 0.05/252, \quad \gamma_0 = 0.04/252, \quad \gamma_1 = 0.5/252, \\ b_0 &= 0.5, \quad b_1 = 0.372, \quad b_2 = 0.343, \quad b_3 = 0.224. \end{aligned}$$

The normalized parameter values above imply the daily time unit, and we adapted the estimated coefficients studied in [Corsi \(2009\)](#) to generate  $\tilde{\sigma}_i$ . We note that in each simulation, the instantaneous volatility process was generated until all instantaneous volatility values were positive, based on the data-generating process described above. We set  $m = 23,400$ , which indicates that the data are observed every second over a period of 6.5 trading hours per day.

For each simulation, we used the jump robust pre-averaging method ([Figueroa-López and Wu, 2024](#)) to estimate the instantaneous volatility,  $c_{i,\tau} = \sigma_{i,\frac{\tau}{n}}^2$ , at a frequency of every 5 minutes (i.e.,  $n = 78$ ) for each  $i$ -th day as follows: for  $\tau = 1, \dots, n$ ,

$$\hat{c}_{i,\tau} = \frac{1}{\phi_{k_n}^{k_n}(g)} \sum_{s=1}^{m-k_n+1} K_{b_m}(t_{s-1} - \frac{\tau}{n}) \left( \bar{Y}_{i,s}^2 - \frac{1}{2}\hat{Y}_{i,s} \right) \mathbf{1}_{\{|\bar{Y}_{i,s}| \leq \nu_m\}}, \quad (4.1)$$

where  $K_b(x) = K(x/b)/b$ , the bandwidth size  $b_m = 1/n$ , the weight function  $g(x) = 2x \wedge$

$(1 - x)$ ,

$$\begin{aligned}\bar{Y}_{i,s} &= \sum_{l=1}^{k_m-1} g\left(\frac{l}{k_m}\right) (Y_{i,t_s+l} - Y_{i,t_s+l-1}), & \phi_{k_m}(g) &= \sum_{i=1}^{k_m} g\left(\frac{i}{k_m}\right)^2, \\ \hat{Y}_{i,s} &= \sum_{l=1}^{k_m} \left\{ g\left(\frac{l}{k_m}\right) - g\left(\frac{l-1}{k_m}\right) \right\}^2 (Y_{i,t_s+l} - Y_{i,t_s+l-1})^2,\end{aligned}$$

$\mathbf{1}_{\{\cdot\}}$  is an indicator function, and  $\nu_m = 1.8\sqrt{\text{BPV}}(k_m/m)^{0.47}$ , where the bipower variation  $\text{BPV} = \frac{\pi}{2} \sum_{s=2}^m |Y_{i,t_{s-1}} - Y_{i,t_{s-2}}| |Y_{i,t_s} - Y_{i,t_{s-1}}|$ . We used the uniform kernel function and the data-driven approach to obtain the preaveraging window size,  $k_m$ , as suggested in Section 3.1 of [Figuroa-López and Wu \(2024\)](#).

With the instantaneous volatility estimates available up to the  $n_1$ -th intraday data point on the  $D$ th day, we examined the out-of-sample performance of estimating the remaining  $(1 \times n_2)$  instantaneous volatility vector, where  $n_1 = \omega n$  and  $n_2 = n - n_1$ . For comparison, we considered the SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA methods in predicting  $c_{D,\tau}$  for  $\tau = n_1 + 1, \dots, n$ . AVE represents estimates obtained from the column mean of  $\hat{\Sigma}_{D-1,n}$ , while AR generates predictions using an autoregressive model of order 1, applied to each column of  $\hat{\Sigma}_{D-1,n}$ . SARIMA provides  $n_2$ -step-ahead forecasts using a seasonal ARIMA(1,1,1) model ([Sheppard, 2010](#)), where the seasonal cycle length is set to  $n$ . HAR-D extends the HAR model by incorporating diurnal effects and prior intraday components in addition to daily, weekly, and monthly realized volatilities ([Zhang et al., 2024](#)). XGBoost ([Chen and Guestrin, 2016](#)), a decision-tree-based ensemble learning method, accounts for nonlinear dependencies, using the same hyperparameters as in [Zhang et al. \(2024\)](#). Since SARIMA, HAR-D, and XGBoost require a time series format, we first vectorized the estimates as  $(\hat{c}_{1,1}, \dots, \hat{c}_{D,n_1})$  before applying these methods. PC corresponds to the last row of the estimated low-rank matrix obtained from the best rank- $r$  approximation of  $\hat{\Sigma}_{D-1,n}$ . TIP-PCA uses ex-post daily, weekly, and monthly realized volatilities and the intraday time sequence as observable covariates that explain the left and right singular vectors, as proposed by [Choi and Kim \(2025\)](#), given  $\hat{\Sigma}_{D-1,n}$ . For SIP, TIP-PCA, and PC, we set rank 1

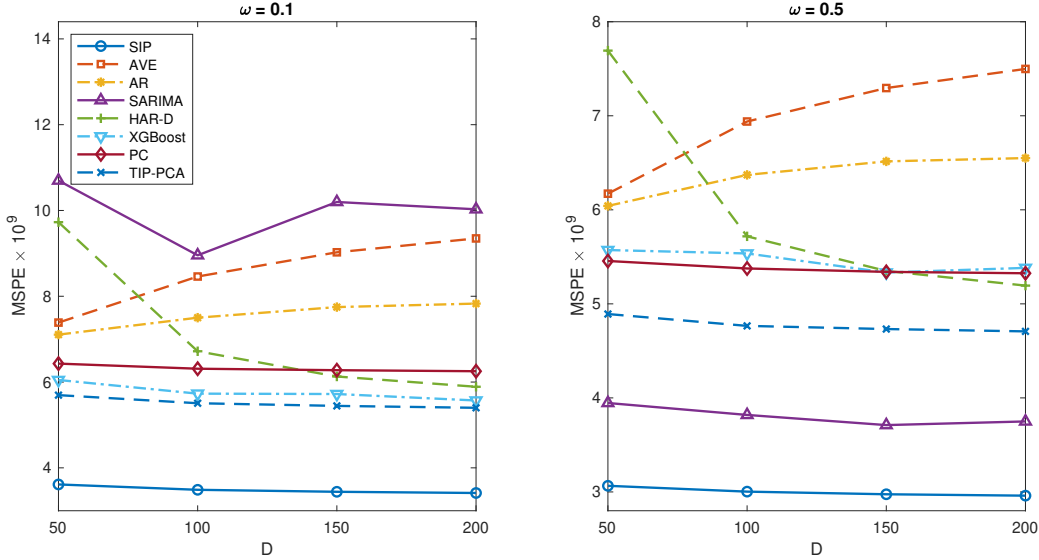


Figure 1: MSPE $\times 10^9$  for the SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA against  $D$  with fixed  $\omega = \{0.1, 0.5\}$ .

as determined by the eigenvalue ratio method (Ahn and Horenstein, 2013). We generated high-frequency data with  $m = 23,400$  over 200 consecutive days, using subsampled log prices from the last  $D = 50, 100, 150,$  and  $200$  days. To assess prediction accuracy, we computed the mean squared prediction error (MSPE) as

$$\frac{1}{n} \sum_{\tau=n_1+1}^n (\tilde{c}_{D,\tau} - c_{D,\tau})^2,$$

where  $\tilde{c}_{D,\tau}$  denotes the instantaneous volatility estimator of each method above. Finally, we calculated the sample averages of MSPEs across 500 simulations.

Figure 1 presents the average MSPEs of the estimators for the remaining future intraday instantaneous volatility process across different values of dimensionality,  $D = 50, 100, 150, 200$ , with fixed  $n$  and  $\omega$ . Figure 2, on the other hand, plots the average MSPEs against varying values of  $\omega = 0.05, 0.1, 0.2, \dots, 0.9$ , while keeping  $D$  and  $n$  fixed. Note that the target future volatility remains consistent across different  $D$  values, as the same subsampled data are used in each simulation setting. In contrast, in Figure 2, the MSPEs exhibit a U-shaped pattern as  $\omega$  varies. This behavior arises because the target future volatility differs with each

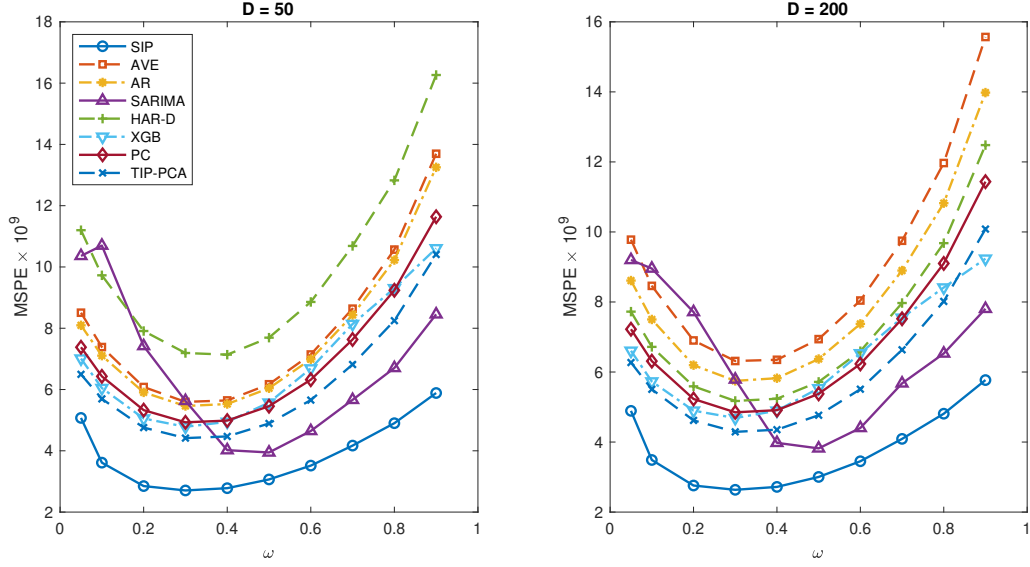


Figure 2:  $\text{MSPE} \times 10^9$  for the SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA against  $\omega$  with fixed  $D = \{50, 100\}$ .

value of  $\omega$ , and the underlying data-generating process follows a U-shaped intraday volatility pattern. As a result, direct comparisons of MSPE values across different  $\omega$  values are not appropriate. Instead, MSPEs should be interpreted relatively across competing methods for each fixed value of  $\omega$ .

Figures 1 and 2 demonstrate that the SIP method consistently achieves the best performance among the competing approaches. This may be because by incorporating information from previous intraday volatility patterns on the  $D$ th day, SIP effectively predicts the remaining future instantaneous volatility process through a nonparametric framework. Moreover, the MSPEs of SIP tend to decrease as the number of daily observations  $D$  increases, which supports the theoretical results established in Section 3. In contrast, the TIP-PCA, PC, AVE, and AR methods perform poorly, as they are unable to leverage current intraday observations on the  $D$ th day within their estimation procedures. In Figure 2, we observe that the performance of SARIMA improves as  $\omega$  increases. This suggests that, given sufficient real-time intraday information, SARIMA can better incorporate both periodic intraday patterns and daily autoregressive dynamics. Nevertheless, SIP outperforms SARIMA and all

other competing methods across the entire range of  $\omega$ , particularly when  $\omega$  is small.

## 5 Empirical Study

In this section, we applied the proposed SIP method to predict intraday instantaneous volatility vector using real high-frequency trading data. Specifically, we obtained intraday data of the S&P 500 index ETF (SPY) and the 11 Global Industry Classification Standard (GICS) sector ETFs (XLC, XLY, XLP, XLE, XLF, XLV, XLI, XLB, XLRE, XLK, and XLU) from July 2021 to June 2022, sourced from the Trade and Quote (TAQ) database in the Wharton Research Data Services (WRDS) system. We used high-frequency data subsampled at 1-second intervals, excluding trading days with early market closures. This subsampling mitigates the effects of irregular observation times (Li and Linton, 2023). Using the log prices, we implemented the jump robust pre-averaging estimation procedure, as defined in Section 4, to estimate the instantaneous variance at 5-minute intervals. Using the in-sample period data, we applied SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA, as described in Section 4, to predict the remaining instantaneous volatilities given each  $\omega = \{0.1, 0.5, 0.9\}$ . We employed the rolling window scheme with a 63-day (i.e., one quarter) in-sample period, where the last  $n_2 = (1 - \omega)n$  instantaneous volatilities on the final in-sample day are the target volatility. The out-of-sample period was from October 2021 to June 2022 (i.e.,  $q = 189$  days).

To evaluate the performance of the predicted instantaneous volatility, we employed the mean squared prediction errors (MSPE) and QLIKE loss function (Patton, 2011), defined as follows: for each  $\omega$  and  $n_2 = (1 - \omega)n$ ,

$$\text{MSPE} = \frac{1}{qn_2} \sum_{i=1}^q \sum_{j=1}^{n_2} (\tilde{c}_{i,j} - \hat{c}_{i,j})^2,$$

$$\text{QLIKE} = \frac{1}{qn_2} \sum_{i=1}^q \sum_{j=1}^{n_2} \left( \log \tilde{c}_{i,j} + \frac{\hat{c}_{i,j}}{\tilde{c}_{i,j}} \right),$$

where  $\tilde{c}_{i,j}$  represents the predicted instantaneous volatility on the  $j$ th intraday time point of the  $i$ th out-of-sample day, obtained from one of the SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA methods. We predicted the remaining future conditional expected instantaneous volatilities using in-sample period data. Since the true conditional expected instantaneous volatility is unknown, we assessed the significance of differences in prediction performances using the Diebold and Mariano (DM) test (Diebold and Mariano, 1995) based on both MSPE and QLIKE. We compared the proposed SIP method against alternative approaches. Given that multiple hypothesis tests were conducted repeatedly throughout this section, it is crucial to control the False Discovery Rate (FDR). To account for this, we applied the Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg, 1995) at a significance level of  $\alpha = 0.05$  to adjust the  $p$ -values and mitigate the risk of false positives across all hypothesis tests performed in this section.

Tables 1 and 2 report the MSPE and QLIKE results, respectively, under varying values of  $\omega \in \{0.1, 0.5, 0.9\}$ . As discussed in Section 4, the target future volatility changes with  $\omega$ . Consequently, direct comparisons of MSPE and QLIKE values across different  $\omega$  values are inappropriate. Instead, performance should be evaluated relative to other methods for each given  $\omega$ . Tables 1–2 show that the SIP method consistently outperforms other methods across different values of  $\omega$ . This implies that incorporating the early intraday trading data enhances the accuracy of instantaneous volatility prediction. We note that the performance of SARIMA and XGBoost improves when  $\omega = 0.9$ . This may be attributed to the availability of a larger amount of information on the last trading day, which improves their prediction performance for a smaller set of target volatilities.

To further evaluate the performance of the proposed method, we conducted a 5-minute frequency Value at Risk (VaR) estimation for the remainder of the trading day, given each  $\omega$ . Specifically, we first predicted the remaining conditional expected instantaneous volatilities on the  $D$ th day using the SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA, based on the same in-sample period data as in the previous analysis. We then computed

Table 1: MSPEs ( $\times 10^9$ ) for SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA across 12 ETFs under varying values of  $\omega \in \{0.1, 0.5, 0.9\}$ .

	SIP	AVE	AR	SARIMA	HAR-D	XGBoost	PC	TIP-PCA
$\omega = 0.1$								
SPY	<b>2.259</b>	3.555***	3.647***	3.099***	3.590***	2.858***	2.839***	2.693***
XLC	<b>2.981</b>	4.669***	5.073***	4.757***	5.299***	4.478***	3.873***	3.681***
XLY	<b>6.959</b>	12.196***	11.554***	11.769***	13.272***	11.328***	9.838***	9.472***
XLP	<b>0.569</b>	0.833***	0.816***	0.793***	0.784***	0.818***	0.641***	0.634***
XLE	<b>7.089</b>	10.871***	10.362***	12.145***	10.327***	8.715***	8.418***	8.282***
XLF	<b>1.772</b>	3.534***	3.079***	2.661***	3.101***	2.678***	2.595***	2.500***
XLV	<b>0.884</b>	1.316***	1.197***	1.071***	1.251***	1.226***	0.994***	0.995***
XLI	<b>1.642</b>	2.626***	2.626***	2.204***	2.511***	2.358***	1.900***	1.911***
XLB	<b>2.024</b>	3.114***	3.115***	2.836***	3.005***	2.778***	2.434***	2.393***
XLRE	<b>1.558</b>	2.422***	2.343***	2.087***	2.325***	2.188***	1.945***	1.918***
XLK	<b>6.256</b>	10.667***	10.154***	8.183***	10.816***	10.222***	8.273***	8.072***
XLU	<b>0.814</b>	1.175***	1.139***	1.108***	1.160***	1.223***	0.988***	0.981***
$\omega = 0.5$								
SPY	<b>3.012</b>	4.114***	4.488***	3.125***	4.212***	3.446***	3.345***	3.234***
XLC	<b>3.675</b>	4.857***	5.831***	3.907***	5.868***	4.668***	4.223***	4.010***
XLY	<b>8.327</b>	12.452***	12.470***	9.307***	13.671***	10.188***	10.316***	10.136***
XLP	<b>0.703</b>	1.005***	1.041***	0.742***	0.945***	0.972***	0.771***	0.768***
XLE	<b>6.196</b>	8.504***	8.301***	6.747***	8.793***	6.685***	7.097***	7.035***
XLF	<b>2.214</b>	3.892***	3.587***	2.627***	3.545***	2.861***	2.902***	2.834***
XLV	<b>1.022</b>	1.534***	1.439***	1.044	1.459***	1.757***	1.179***	1.177***
XLI	<b>1.911</b>	2.904***	3.171***	2.013***	2.750***	2.457***	2.106***	2.129***
XLB	<b>2.442</b>	3.421***	3.611***	2.601***	3.245***	3.029***	2.667***	2.656***
XLRE	<b>1.850</b>	2.765***	2.777***	2.070***	2.573***	2.536***	2.218***	2.190***
XLK	<b>7.727</b>	11.448***	11.655***	8.401***	12.085***	9.823***	9.191***	9.067***
XLU	<b>0.937</b>	1.344***	1.334***	0.968**	1.335***	1.250***	1.113***	1.126***
$\omega = 0.9$								
SPY	<b>2.302</b>	4.383***	3.623***	2.715	4.240***	2.555	3.295***	3.387***
XLC	<b>2.703</b>	4.956***	4.519***	3.286*	5.361***	2.704	3.892***	3.955***
XLY	<b>5.483</b>	11.780***	9.936***	7.261**	13.361***	5.768	9.009***	9.712***
XLP	<b>0.592</b>	1.228***	1.016***	0.763**	1.116***	0.749**	0.852***	0.910***
XLE	4.317	6.357***	5.428***	<b>3.936</b>	6.476***	4.050	5.601***	6.117***
XLF	<b>1.548</b>	4.240***	3.015***	1.908**	3.312***	1.690	2.822***	2.892***
XLV	<b>0.943</b>	1.924***	1.555***	1.102*	1.722***	1.271***	1.356***	1.445***
XLI	<b>1.491</b>	3.456***	2.417***	1.679*	2.889***	1.984***	2.258***	2.478***
XLB	<b>2.011</b>	4.099***	3.362***	2.405**	3.710***	2.172	2.931***	3.188***
XLRE	<b>1.722</b>	3.692***	3.138***	1.959	3.373***	2.652***	2.860***	2.925***
XLK	<b>5.594</b>	10.868***	9.580***	7.009*	11.661***	7.044	8.139***	8.830***
XLU	<b>0.916</b>	1.803***	1.481***	1.211**	1.748***	1.182**	1.398***	1.523***

Note: Bold numbers indicate the lowest MSPE for each ETF, while \*\*\*, \*\*, and \* indicate rejection of the null hypothesis against SIP at the 1%, 5%, and 10% significance levels, respectively, based on the Diebold-Mariano (DM) test.

Table 2: QLIKEs for SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA across 12 ETFs under varying values of  $\omega \in \{0.1, 0.5, 0.9\}$ .

	SIP	AVE	AR	SARIMA	HAR-D	XGBoost	PC	TIP-PCA
$\omega = 0.1$								
SPY	<b>-9.287</b>	-8.970***	-9.114***	-5.377***	-7.974***	-9.208***	-9.213***	-9.071***
XLC	<b>-9.116</b>	-8.863***	-8.942***	0.942***	-8.255***	-8.990***	-9.046***	-8.955***
XLY	<b>-8.521</b>	-8.243***	-8.347***	10.962***	-6.636***	-8.285***	-8.452***	-8.435***
XLP	<b>-9.576</b>	-9.436***	-9.489***	-5.412***	-9.166***	-9.531***	-9.547***	-9.543***
XLE	<b>-8.210</b>	-8.095***	-8.131***	-4.371***	-7.621***	-8.175***	-8.179***	-8.179***
XLF	<b>-8.490</b>	-8.385***	-8.434***	1.489***	-8.037***	-8.465***	-8.454***	-8.455***
XLV	<b>-9.563</b>	-9.375***	-9.452***	-5.433***	-8.594***	-9.451***	-9.539***	-9.495***
XLI	<b>-9.286</b>	-9.057***	-9.167***	-6.441***	-8.464***	-9.201***	-9.269***	-8.926***
XLB	<b>-9.239</b>	-9.016***	-9.088***	-7.276***	-7.454***	-9.135***	-9.202***	-9.184***
XLRE	<b>-9.197</b>	-9.036***	-9.077***	-5.944***	-7.678***	-9.082***	-9.137***	-9.140***
XLK	<b>-8.661</b>	-8.374***	-8.501***	-6.775***	-7.012***	-8.569***	-8.587***	-8.563***
XLU	<b>-9.222</b>	-9.119***	-9.144***	-7.066***	-8.418***	-9.160***	-9.178***	-9.185***
$\omega = 0.5$								
SPY	<b>-9.337</b>	-9.006***	-9.161***	-9.123***	-8.207***	-9.225***	-9.271***	-9.083***
XLC	<b>-9.181</b>	-8.918***	-9.001***	-8.955***	-8.168***	-8.986***	-9.111***	-9.001***
XLY	<b>-8.640</b>	-8.340***	-8.453***	-8.033**	-5.726***	-8.543***	-8.572***	-8.541***
XLP	<b>-9.623</b>	-9.456***	-9.520***	-9.577***	-9.082***	-9.559***	-9.591***	-9.582***
XLE	<b>-8.388</b>	-8.265***	-8.306***	-8.257***	-7.639***	-8.355***	-8.348***	-8.346***
XLF	<b>-8.559</b>	-8.439***	-8.493***	-8.532***	-8.046**	-8.529***	-8.517***	-8.517***
XLV	<b>-9.657</b>	-9.417***	-9.512***	-9.430**	-8.067***	-9.524***	-9.608***	-9.547***
XLI	<b>-9.369</b>	-9.100***	-9.223***	-9.276***	-8.140***	-9.287***	-9.341***	-9.061***
XLB	<b>-9.316</b>	-9.064***	-9.147***	-9.177***	-6.553***	-9.212***	-9.279***	-9.257***
XLRE	<b>-9.232</b>	-9.048***	-9.091***	-8.969***	-6.833***	-9.111***	-9.158***	-9.161***
XLK	<b>-8.764</b>	-8.446***	-8.590***	-8.561***	-6.418***	-8.429*	-8.688***	-8.652***
XLU	<b>-9.294</b>	-9.170***	-9.203***	-9.231***	-7.461***	-9.237***	-9.241***	-9.245***
$\omega = 0.9$								
SPY	<b>-9.065</b>	-8.754***	-8.892***	-9.063	-8.485***	-9.013**	-8.947***	-8.795***
XLC	<b>-8.858</b>	-8.606***	-8.702***	-8.819**	-8.057	-8.803***	-8.760***	-8.662***
XLY	-8.425	-8.138***	-8.261***	-8.413	-5.270***	<b>-8.435</b>	-8.333***	-8.304***
XLP	-9.250	-9.089***	-9.155***	-9.253	<b>-10.141</b>	-9.224***	-9.199***	-9.187***
XLE	-8.289	-8.187***	-8.224***	-8.281	-8.118***	<b>-8.300**</b>	-8.229***	-8.229***
XLF	<b>-8.404</b>	-8.268***	-8.340***	-8.402	-8.303	-8.400	-8.349***	-8.347***
XLV	<b>-9.265</b>	-9.049***	-9.124***	-9.245	-6.789*	-9.236***	-9.196***	-9.146***
XLI	<b>-8.993</b>	-8.738***	-8.869***	-8.962	-8.556***	-8.972***	-8.930***	-8.623***
XLB	<b>-8.955</b>	-8.706***	-8.807***	-8.924	-8.773	-8.861	-8.893***	-8.858***
XLRE	<b>-8.804</b>	-8.621***	-8.645***	-8.680	-1.301	-8.709***	-8.722***	-8.702***
XLK	<b>-8.554</b>	-8.263***	-8.391***	-8.481*	-8.245***	-8.544	-8.454***	-8.418***
XLU	<b>-8.919</b>	-8.794***	-8.824***	-8.914	-8.757***	-8.893***	-8.853***	-8.847***

Note: Bold numbers indicate the lowest QLIKE for each ETF, while \*\*\*, \*\*, and \* indicate rejection of the null hypothesis against SIP at the 1%, 5%, and 10% significance levels, respectively, based on the Diebold-Mariano (DM) test.

Table 3: Number of cases where the adjusted  $p$ -value exceeds 0.05 for SIP, AVE, AR, SARIMA, HAR-D, XGBoost, PC, and TIP-PCA across 12 ETFs at each  $q_0 = \{0.01, 0.02, 0.05, 0.1, 0.2\}$ , based on the LRuc, LRcc, and DQ tests under varying values of  $\omega \in \{0.1, 0.5, 0.9\}$ .

$q_0$	LRuc					LRcc					DQ				
	0.01	0.02	0.05	0.1	0.2	0.01	0.02	0.05	0.1	0.2	0.01	0.02	0.05	0.1	0.2
$\omega = 0.1$															
SIP	12	12	12	12	12	11	12	12	12	12	7	6	7	7	9
AVE	2	1	2	11	12	2	1	5	7	12	1	1	1	2	5
AR	4	3	6	12	12	6	4	5	12	12	1	2	3	3	6
SARIMA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
HAR-D	0	0	0	4	12	0	0	0	3	12	1	1	1	2	4
XGBoost	7	5	6	9	12	6	3	4	8	12	2	2	2	4	5
PC	7	7	12	12	12	9	11	10	12	12	4	3	3	4	7
TIP-PCA	12	10	12	12	12	9	10	11	12	12	4	3	4	5	7
$\omega = 0.5$															
SIP	12	12	12	12	12	11	12	12	12	12	6	7	7	9	10
AVE	3	3	5	12	12	5	3	8	11	12	2	2	3	4	6
AR	7	6	11	12	12	7	8	11	12	12	3	3	4	5	8
SARIMA	5	6	11	12	12	5	7	12	12	12	2	3	4	7	10
HAR-D	0	0	1	4	12	0	1	1	3	12	1	1	2	3	5
XGBoost	5	5	4	8	12	6	5	5	11	12	2	2	3	5	7
PC	10	12	12	12	12	10	12	12	12	12	5	5	5	6	8
TIP-PCA	10	12	12	12	12	10	10	11	12	12	5	4	5	6	8
$\omega = 0.9$															
SIP	12	12	12	12	12	12	12	12	12	12	8	9	10	11	12
AVE	12	12	12	12	12	11	11	12	12	12	6	6	7	9	10
AR	12	12	12	12	12	12	12	12	12	12	7	8	8	10	11
SARIMA	12	12	12	12	12	12	12	12	12	12	7	7	9	11	11
HAR-D	6	8	12	12	12	6	8	12	12	12	4	4	6	9	11
XGBoost	12	12	12	12	12	12	12	12	12	12	7	8	8	10	11
PC	12	12	12	12	12	12	12	12	12	12	6	7	8	10	11
TIP-PCA	12	12	12	12	12	11	12	12	12	12	7	7	9	10	11

quantiles using historical standardized 5-minute returns. Specifically, we first standardized in-sample 5-minute returns with the estimated conditional instantaneous volatilities; then, we derived sample quantiles at levels 0.01, 0.02, 0.05, 0.1, and 0.2. Using these sample quantiles and predicted instantaneous volatilities, we obtained 5-minute frequency VaR values for each prediction method. We used a fixed in-sample period as one quarter and employed a rolling window scheme with the out-of-sample period spanning 189 days, which is consistent with the previous analysis.

Based on the estimated VaR values, we conducted the likelihood ratio unconditional coverage (LRuc) test (Kupiec, 1995), the likelihood ratio conditional coverage (LRcc) test (Christoffersen, 1998), and the dynamic quantile (DQ) test with lag 4 (Engle and Man-ganelli, 2004). Table 3 reports the number of cases where the adjusted  $p$ -value using the BH procedure exceeds 0.05 for the 12 ETFs at each  $q_0 \in \{0.01, 0.02, 0.05, 0.1, 0.2\}$ , based on the LRuc, LRcc, and DQ tests under varying values of  $\omega \in \{0.1, 0.5, 0.9\}$ . From Table 3, we find that the SIP method consistently outperforms other methods across all hypothesis tests. We note that when  $\omega$  is small, the TIP-PCA method also performs well relative to other methods. This may be because when  $\omega$  is small, the amount of current trading day’s information is small. Thus, TIP-PCA, which uses only the information up to the previous day’s information, can account for the interday and intraday dynamics by incorporating the interday HAR structure and the U-shaped intraday volatility feature. Additionally, Table 3 shows that the performance of SARIMA gets better as  $\omega$  grows, aligning with the results in Table 1. However, SARIMA exhibits poor performance when  $\omega$  is small in terms of VaR estimation. Table 3 indicates that SIP outperforms TIP-PCA and other methods, particularly for lower quantiles, which are generally more difficult to predict. This result confirms that the proposed nonparametric prediction method, SIP, by leveraging current intraday volatility information, significantly enhances both the prediction accuracy of future instantaneous volatilities and the effectiveness of risk management.

## 6 Conclusion

This paper proposes a novel nonparametric prediction procedure for forecasting the remaining future instantaneous volatility process during the current intraday trading period. The proposed method, Structural Intraday-volatility Prediction (SIP), leverages both previous days' data and partially observed current-day data by considering the missing component structure of the approximately low-rank matrix representation of the instantaneous volatility process. That is, this paper extends the low-rank matrix completion to predict time series patterns, which makes it possible to forecast future intraday volatility without relying on parametric modeling assumptions. As a result, SIP is robust to model misspecification. We establish the asymptotic properties of the SIP estimator and prove its consistency.

In the empirical analysis, the SIP method outperforms existing alternatives in terms of out-of-sample forecasting accuracy and Value at Risk (VaR) estimation for the remaining intraday volatility. Notably, SIP demonstrates strong predictive performance even when only a small fraction of intraday observations is available. This finding highlights that early intraday information plays a crucial role in forecasting the remaining volatility, and that the low-rank structure effectively captures intraday volatility dynamics without requiring a specific model form such as an autoregressive process.

## References

- ADMATI, A. R. AND P. PFLEIDERER (1988): “A theory of intraday patterns: Volume and price variability,” *The Review of Financial Studies*, 1, 3–40.
- AHN, S. C. AND A. R. HORENSTEIN (2013): “Eigenvalue ratio test for the number of factors,” *Econometrica*, 81, 1203–1227.
- AÏT-SAHALIA, Y., J. FAN, AND D. XIU (2010): “High-frequency covariance estimates with

- noisy and asynchronous financial data,” *Journal of the American Statistical Association*, 105, 1504–1517.
- AÏT-SAHALIA, Y. AND J. YU (2009): “High frequency market microstructure noise estimates and liquidity measures,” *Annals of Applied Statistics*, 3, 422–457.
- ANDERSEN, T. G. AND T. BOLLERSLEV (1997): “Intraday periodicity and volatility persistence in financial markets,” *Journal of Empirical Finance*, 4, 115–158.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): “Modeling and forecasting realized volatility,” *Econometrica*, 71, 579–625.
- ANDERSEN, T. G., M. THYRSGAARD, AND V. TODOROV (2019): “Time-varying periodicity in intraday volatility,” *Journal of the American Statistical Association*, 114, 1695–1707.
- BAI, J. AND S. NG (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.
- (2021): “Matrix completion, counterfactuals, and factor analysis of missing data,” *Journal of the American Statistical Association*, 116, 1746–1763.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise,” *Econometrica*, 76, 1481–1536.
- (2011): “Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading,” *Journal of Econometrics*, 162, 149–169.
- BENJAMINI, Y. AND Y. HOCHBERG (1995): “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.

- BIBINGER, M., N. HAUTSCH, P. MALEC, AND M. REISS (2014): “Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency,” *The Annals of Statistics*, 42, 1312–1346.
- BOIVIN, J. AND S. NG (2006): “Are more data always better for factor analysis?” *Journal of Econometrics*, 132, 169–194.
- CAI, J.-F., E. J. CANDÈS, AND Z. SHEN (2010): “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, 20, 1956–1982.
- CAI, T., T. T. CAI, AND A. ZHANG (2016): “Structured matrix completion with applications to genomic data integration,” *Journal of the American Statistical Association*, 111, 621–633.
- CANDES, E. AND B. RECHT (2012): “Exact matrix completion via convex optimization,” *Communications of the ACM*, 55, 111–119.
- CANDES, E. J. AND Y. PLAN (2010): “Matrix completion with noise,” *Proceedings of the IEEE*, 98, 925–936.
- CHEN, T. AND C. GUESTRIN (2016): “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- CHO, J., D. KIM, AND K. ROHE (2017): “Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise,” *Statistica Sinica*, 1921–1948.
- CHOI, S. H. AND D. KIM (2025): “Matrix-based prediction approach for intraday instantaneous volatility vector,” *Journal of Business & Economic Statistics*, in press.

- CHRISTENSEN, K., S. KINNEBROCK, AND M. PODOLSKIJ (2010): “Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data,” *Journal of Econometrics*, 159, 116–133.
- CHRISTOFFERSEN, P. F. (1998): “Evaluating interval forecasts,” *International Economic Review*, 841–862.
- CORSI, F. (2009): “A simple approximate long-memory model of realized volatility,” *Journal of Financial Econometrics*, 7, 174–196.
- DAVIS, C. AND W. M. KAHAN (1970): “The rotation of eigenvectors by a perturbation. III,” *SIAM Journal on Numerical Analysis*, 7, 1–46.
- DIEBOLD, F. X. AND R. S. MARIANO (1995): “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, 13, 253–265.
- ENGLE, R. F. AND S. MANGANELLI (2004): “CAViaR: Conditional autoregressive value at risk by regression quantiles,” *Journal of Business & Economic Statistics*, 22, 367–381.
- ENGLE, R. F. AND M. E. SOKALSKA (2012): “Forecasting intraday volatility in the US equity market. Multiplicative component GARCH,” *Journal of Financial Econometrics*, 10, 54–83.
- FAN, J., A. FURGER, AND D. XIU (2016): “Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data,” *Journal of Business & Economic Statistics*, 34, 489–503.
- FAN, J. AND D. KIM (2018): “Robust high-dimensional volatility matrix estimation for high-frequency factor model,” *Journal of the American Statistical Association*, 113, 1268–1283.
- (2019): “Structured volatility matrix estimation for non-synchronized high-frequency financial data,” *Journal of Econometrics*, 209, 61–78.

- FAN, J., H. LIU, AND W. WANG (2018a): “Large covariance estimation through elliptical factor models,” *Annals of Statistics*, 46, 1383.
- FAN, J., W. WANG, AND Y. ZHONG (2018b): “An  $\ell_\infty$  eigenvector perturbation bound and its application to robust covariance estimation,” *Journal of Machine Learning Research*, 18, 1–42.
- FAN, J. AND Y. WANG (2007): “Multi-scale jump and volatility analysis for high-frequency financial data,” *Journal of the American Statistical Association*, 102, 1349–1362.
- (2008): “Spot volatility estimation for high-frequency data,” *Statistics and its Interface*, 1, 279–288.
- FIGUEROA-LÓPEZ, J. E. AND B. WU (2024): “Kernel estimation of spot volatility with microstructure noise using pre-averaging,” *Econometric Theory*, 40, 558–607.
- FOSTER, D. P. AND D. B. NELSON (1996): “Continuous record asymptotics for rolling sample variance estimators,” *Econometrica*, 64, 139–174.
- HANSEN, P. R., Z. HUANG, AND H. H. SHEK (2012): “Realized GARCH: a joint model for returns and realized measures of volatility,” *Journal of Applied Econometrics*, 27, 877–906.
- HONG, H. AND J. WANG (2000): “Trading and returns under periodic market closures,” *The Journal of Finance*, 55, 297–354.
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): “Microstructure noise in the continuous case: the pre-averaging approach,” *Stochastic Processes and their Applications*, 119, 2249–2276.
- KIM, D. AND Y. WANG (2016a): “Sparse PCA-based on high-dimensional Itô processes with measurement errors,” *Journal of Multivariate Analysis*, 152, 172–189.

- (2016b): “Unified discrete-time and continuous-time models and statistical inferences for merged low-frequency and high-frequency financial data,” *Journal of Econometrics*, 194, 220–230.
- KIM, D., Y. WANG, AND J. ZOU (2016): “Asymptotic theory for large volatility matrix estimation based on high-frequency financial data,” *Stochastic Processes and their Applications*, 126, 3527–3577.
- KRISTENSEN, D. (2010): “Nonparametric filtering of the realized spot volatility: A kernel-based approach,” *Econometric Theory*, 26, 60–93.
- KUPIEC, P. H. (1995): “Techniques for verifying the accuracy of risk measurement models,” *The Journal of Derivatives*, 3, 73–84.
- LI, Q., G. CHENG, J. FAN, AND Y. WANG (2018): “Embracing the blessing of dimensionality in factor models,” *Journal of the American Statistical Association*, 113, 380–389.
- LI, Z. M. AND O. LINTON (2023): “Robust estimation of integrated and spot volatility,” *Journal of Econometrics*, 105614.
- MANCINI, C., V. MATTIUSI, AND R. RENÒ (2015): “Spot volatility estimation using delta sequences,” *Finance and Stochastics*, 19, 261–293.
- ONATSKI, A. (2010): “Determining the number of factors from empirical distribution of eigenvalues,” *The Review of Economics and Statistics*, 92, 1004–1016.
- PATTON, A. J. (2011): “Volatility forecast comparison using imperfect volatility proxies,” *Journal of Econometrics*, 160, 246–256.
- SHEPHARD, N. AND K. SHEPPARD (2010): “Realising the future: forecasting with high-frequency-based volatility (HEAVY) models,” *Journal of Applied Econometrics*, 25, 197–231.

- SHEPPARD, K. (2010): “Financial Econometrics Notes,” *University of Oxford*, 333–426.
- SHIN, M., D. KIM, AND J. FAN (2023): “Adaptive robust large volatility matrix estimation based on high-frequency financial data,” *Journal of Econometrics*, 237, 105514.
- TODOROV, V. (2019): “Nonparametric spot volatility from options,” *The Annals of Applied Probability*, 29, 3590–3636.
- TODOROV, V. AND Y. ZHANG (2023): “Bias reduction in spot volatility estimation from options,” *Journal of Econometrics*, 234, 53–81.
- VERSHYNIN, R. (2010): “Introduction to the non-asymptotic analysis of random matrices,” *arXiv preprint arXiv:1011.3027*.
- WANG, W. AND J. FAN (2017): “Asymptotics of empirical eigenstructure for high dimensional spiked covariance,” *The Annals of Statistics*, 45, 1342.
- XIU, D. (2010): “Quasi-maximum likelihood estimation of volatility with high frequency data,” *Journal of Econometrics*, 159, 235–250.
- YAN, Y. AND M. J. WAINWRIGHT (2024): “Entrywise inference for causal panel data: A simple and instance-optimal approach,” *arXiv preprint arXiv:2401.13665*, 6.
- ZHANG, C., Y. ZHANG, M. CUCURINGU, AND Z. QIAN (2024): “Volatility forecasting with machine learning and intraday commonality,” *Journal of Financial Econometrics*, 22, 492–530.
- ZHANG, L. (2006): “Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach,” *Bernoulli*, 12, 1019–1043.
- (2011): “Estimating covariation: Epps effect, microstructure noise,” *Journal of Econometrics*, 160, 33–47.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100, 1394–1411.

ZU, Y. AND H. P. BOSWIJK (2014): “Estimating spot volatility with high-frequency financial data,” *Journal of Econometrics*, 181, 117–135.

# A Appendix

## A.1 Related Lemmas

We first present useful lemmas below. Let  $\widehat{A}_{\bullet 1}$  be the best rank- $r$  approximation to  $\widehat{\Sigma}_{\bullet 1}$  such that  $\widehat{A}_{\bullet 1} = \sum_{i=1}^r \widehat{\lambda}_{1,i} \widehat{u}_{1,i} \widehat{v}_{1,i}^\top$ , where  $\{\widehat{\lambda}_{1,i}, \widehat{u}_{1,i}, \widehat{v}_{1,i}\}_{i=1}^{D \wedge n_1}$  are the ordered singular values, left-singular vectors and right-singular vectors of  $\widehat{\Sigma}_{\bullet 1}$  in decreasing order. Similarly, we let  $\widehat{A}_{1\bullet}$  be the best rank- $r$  approximation to  $\widehat{\Sigma}_{1\bullet}$ . Then, we can write the SIP estimator as follows:

$$\widetilde{\Sigma}_{22} = \widehat{\Sigma}_{21} \widehat{V}_{11} (\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11})^{-1} \widehat{U}_{11}^\top \widehat{\Sigma}_{12} = \widehat{A}_{21} \widehat{V}_{11} (\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11})^{-1} \widehat{U}_{11}^\top \widehat{A}_{12},$$

where  $\widehat{A}_{21}$  is the last row of  $\widehat{A}_{\bullet 1}$ , and  $\widehat{A}_{12}$  corresponds to the  $(D-1) \times n_2$  right block of  $\widehat{A}_{1\bullet}$ .

Let  $\Lambda_{\bullet 1}$  and  $\Lambda_{1\bullet}$  denote the  $r \times r$  diagonal matrices of the leading  $r$  singular values of  $A_{\bullet 1} \equiv [A_{11}^\top A_{21}^\top]^\top$  and  $A_{1\bullet} \equiv [A_{11} A_{12}]$ , respectively. Define  $\widehat{\Lambda}_{\bullet 1} = \text{Diag}(\widehat{\lambda}_{\bullet 1}, \dots, \widehat{\lambda}_{\bullet r})$ , where  $\widehat{\lambda}_{\bullet 1} \geq \widehat{\lambda}_{\bullet 2} \geq \dots \geq \widehat{\lambda}_{\bullet r}$  are the square root of leading eigenvalues of  $\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top$ . Similarly, define  $\widehat{\Lambda}_{1\bullet} = \text{Diag}(\widehat{\lambda}_{1\bullet}, \dots, \widehat{\lambda}_{1\bullet r})$ , where  $\widehat{\lambda}_{1\bullet} \geq \widehat{\lambda}_{1\bullet 2} \geq \dots \geq \widehat{\lambda}_{1\bullet r}$  are the square root of leading eigenvalues of  $\widehat{\Sigma}_{1\bullet}^\top \widehat{\Sigma}_{1\bullet}$ . The following lemma presents the individual convergence rate of singular value estimators.

**Lemma A.1.** *Under Assumption 3.1, we have*

$$\begin{aligned} (i) \quad & \|\widehat{\Lambda}_{\bullet 1} - \Lambda_{\bullet 1}\|_{\max} = O_P \left( \sqrt{\frac{D}{n_1}} + \sqrt{n_1 D} \rho_m + \sqrt{\varphi_D} + \sqrt{\frac{n_1}{D}} \varphi_D \right), \\ (ii) \quad & \|\widehat{\Lambda}_{1\bullet} - \Lambda_{1\bullet}\|_{\max} = O_P \left( \sqrt{\frac{n}{D}} + \sqrt{n D} \rho_m + \sqrt{\varphi_n} + \sqrt{\frac{D}{n}} \varphi_n \right). \end{aligned}$$

*Proof.* We first consider (i). We have

$$\begin{aligned} & E \|\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top\|_F^2 \\ &= \sum_{i=1}^D \sum_{s=1}^D E \left[ \left( \sum_{j=1}^{n_1} (\widehat{c}_{i,j} \widehat{c}_{s,j} - c_{i,j} c_{s,j}) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^D E \left[ \left( \sum_{j=1}^{n_1} (\widehat{c}_{i,j}^2 - c_{i,j}^2) \right)^2 \right] + \sum_{i=1}^D \sum_{s \neq i}^D E \left[ \left( \sum_{j=1}^{n_1} (\widehat{c}_{i,j} \widehat{c}_{s,j} - c_{i,j} c_{s,j}) \right)^2 \right] \\
&= \sum_{i=1}^D \sum_{j=1}^{n_1} E [(\widehat{c}_{i,j}^2 - c_{i,j}^2)^2] + \sum_{i=1}^D \sum_{j=1}^{n_1} \sum_{k \neq j}^{n_1} E [(\widehat{c}_{i,j}^2 - c_{i,j}^2)(\widehat{c}_{i,k}^2 - c_{i,k}^2)] \\
&\quad + \sum_{i=1}^D \sum_{s \neq i}^D E \left[ \left( \sum_{j=1}^{n_1} (\widehat{c}_{i,j} \widehat{c}_{s,j} - c_{i,j} c_{s,j}) \right)^2 \right] \\
&:= I + II + III.
\end{aligned}$$

For  $I$ , by substituting  $\widehat{c}_{i,j} = c_{i,j} + v_{i,j} + \varsigma_{i,j}$  and using the Cauchy–Schwarz inequality with the sub-Gaussian conditions from Assumption 3.1(iii), we can show

$$\begin{aligned}
I &= \sum_{i=1}^D \sum_{j=1}^{n_1} E [(\widehat{c}_{i,j}^2 - c_{i,j}^2)^2] = \sum_{i=1}^D \sum_{j=1}^{n_1} E [(2c_{i,j}v_{i,j} + 2c_{i,j}\varsigma_{i,j} + v_{i,j}^2 + \varsigma_{i,j}^2 + 2v_{i,j}\varsigma_{i,j})^2] \\
&= O(Dn_1m^{-1/4} + Dn_1\rho_m^2).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
II &= \sum_{i=1}^D \sum_{j=1}^{n_1} \sum_{k \neq j}^{n_1} E [(\widehat{c}_{i,j}^2 - c_{i,j}^2)(\widehat{c}_{i,k}^2 - c_{i,k}^2)] \\
&\leq \sum_{i=1}^D \sum_{j=1}^{n_1} \sum_{k \neq j}^{n_1} E [(\widehat{c}_{i,j}^2 - c_{i,j}^2)^2]^{1/2} E [(\widehat{c}_{i,k}^2 - c_{i,k}^2)^2]^{1/2} \\
&= O(Dn_1^2m^{-1/4} + Dn_1^2\rho_m^2),
\end{aligned}$$

where the Hölder’s inequality is used for the first inequality.

For  $III$ , we first consider the martingale part only (i.e.,  $\varsigma_{i,j}$  is zero). Then, we can extend the result to the general case. Without loss of generality, we assume that  $i > s$ . We have

$$\widehat{c}_{i,j} \widehat{c}_{s,j} - c_{i,j} c_{s,j} = (\widehat{c}_{i,j} - c_{i,j})(\widehat{c}_{s,j} - c_{s,j}) + (\widehat{c}_{i,j} - c_{i,j})c_{s,j} + c_{i,j}(\widehat{c}_{s,j} - c_{s,j}).$$

Then, for any  $j > j'$ , using the martingale difference property, we can show

$$\begin{aligned}
& E[(\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j})(\widehat{c}_{i,j'}\widehat{c}_{s,j'} - c_{i,j'}c_{s,j'})] \\
&= E[(\widehat{c}_{i,j'} - c_{i,j'})(\widehat{c}_{s,j'} - c_{s,j'})c_{i,j}(\widehat{c}_{s,j} - c_{s,j})] + E[(\widehat{c}_{i,j'} - c_{i,j'})c_{s,j'}c_{i,j}(\widehat{c}_{s,j} - c_{s,j})] \\
&\quad + E[c_{i,j'}(\widehat{c}_{s,j'} - c_{s,j'})c_{i,j}(\widehat{c}_{s,j} - c_{s,j})] \\
&= E[E[(\widehat{c}_{i,j'} - c_{i,j'})c_{i,j}|\mathcal{F}_{i,j'}](\widehat{c}_{s,j'} - c_{s,j'})(\widehat{c}_{s,j} - c_{s,j})] \\
&\quad + E[E[(\widehat{c}_{i,j'} - c_{i,j'})c_{i,j}|\mathcal{F}_{i,j'}]c_{s,j'}(\widehat{c}_{s,j} - c_{s,j})] + E[E[c_{i,j'}c_{i,j}(\widehat{c}_{s,j} - c_{s,j})|\mathcal{F}_{s,j}](\widehat{c}_{s,j'} - c_{s,j'})] \\
&= 0,
\end{aligned}$$

where the last equality is due to Assumption 3.1(iii), and we have

$$\sum_{i=1}^D \sum_{s \neq i}^D \sum_{j=1}^{n_1} E[(\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j})^2] = O(D^2 n_1 m^{-1/4}).$$

For the general case, since  $\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j}$  can be decomposed into the martingale difference term and the bias term, with their respective effects of  $m^{-1/8}$  and  $\rho_m$ , we have

$$\begin{aligned}
III &= \sum_{i=1}^D \sum_{s \neq i}^D E \left[ \left( \sum_{j=1}^{n_1} (\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j}) \right)^2 \right] \\
&= \sum_{i=1}^D \sum_{s \neq i}^D \sum_{j=1}^{n_1} E[(\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j})^2] \\
&\quad + \sum_{i=1}^D \sum_{s \neq i}^D \sum_{j=1}^{n_1} \sum_{k \neq j}^{n_1} E[(\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j})(\widehat{c}_{i,k}\widehat{c}_{s,k} - c_{i,k}c_{s,k})] \\
&= O(D^2 n_1 m^{-1/4} + D^2 n_1 \rho_m^2) + O(D^2 n_1^2 m^{-\frac{1}{8}} \rho_m + D^2 n_1^2 \rho_m^2) \\
&= O(D^2 n_1 m^{-1/4} + D^2 n_1^2 m^{-1/8} \rho_m + D^2 n_1^2 \rho_m^2),
\end{aligned}$$

where the cross product terms such as  $E(c_{i,j}\varsigma_{i,j'}\upsilon_{s,j}c_{s,j'})$  and  $E(c_{i,j}c_{i,j'}\varsigma_{s,j}\varsigma_{s,j'})$  yield the rate

above. Therefore, we have

$$E\|\widehat{\Sigma}_{\bullet 1}\widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1}\Sigma_{\bullet 1}^\top\|_F^2 = O(Dn_1^2m^{-\frac{1}{4}} + D^2n_1m^{-\frac{1}{4}} + D^2n_1^2m^{-\frac{1}{8}}\rho_m + D^2n_1^2\rho_m^2). \quad (\text{A.1})$$

We define  $U_{\bullet 1} = (u_{\bullet 1}, \dots, u_{\bullet r})$  and  $V_{\bullet 1} = (v_{\bullet 1}, \dots, v_{\bullet r})$  as the left and right singular vector matrices of  $A_{\bullet 1} \equiv [A_{11}^\top A_{21}^\top]^\top$ , respectively, and  $E_{\bullet 1} = [E_{11}^\top E_{21}^\top]^\top$ . The columns of  $\widehat{U}_{\bullet 1} = (\widehat{u}_{\bullet 1}, \dots, \widehat{u}_{\bullet r})$  are defined as the top  $r$  eigenvectors of  $\widehat{\Sigma}_{\bullet 1}\widehat{\Sigma}_{\bullet 1}^\top$ . Let  $\Lambda_{\bullet 1} = \text{Diag}(\lambda_{\bullet 1}, \dots, \lambda_{\bullet r})$ . For  $k \leq r$ , we have

$$\begin{aligned} |\widehat{\lambda}_{\bullet k}^2 - \lambda_{\bullet k}^2| &= |\widehat{u}_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top \widehat{u}_{\bullet k} - u_{\bullet k}^\top U_{\bullet 1} \Lambda_{\bullet 1}^2 U_{\bullet 1}^\top u_{\bullet k}| \\ &\leq |\widehat{u}_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top \widehat{u}_{\bullet k} - u_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top u_{\bullet k} - u_{\bullet k}^\top \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top u_{\bullet k}| \\ &\quad + |u_{\bullet k}^\top \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top u_{\bullet k} - u_{\bullet k}^\top U_{\bullet 1} \Lambda_{\bullet 1}^2 U_{\bullet 1}^\top u_{\bullet k}| \\ &:= I + II + III. \end{aligned}$$

We note that, by using the  $\sin(\theta)$  theorem of [Davis and Kahan \(1970\)](#), for  $k \leq r$ , we have

$$\begin{aligned} \|\widehat{u}_{\bullet k} - u_{\bullet k}\|_2 &= O\left(\frac{\|\widehat{\Sigma}_{\bullet 1}\widehat{\Sigma}_{\bullet 1}^\top - U_{\bullet 1}\Lambda_{\bullet 1}^2U_{\bullet 1}^\top\|}{n_1D}\right) \\ &= O\left(\frac{\|\widehat{\Sigma}_{\bullet 1}\widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1}\Sigma_{\bullet 1}^\top\|_F + \|\Sigma_{\bullet 1}\Sigma_{\bullet 1}^\top - \sum_{j=1}^{n_1}\Omega_{j,D \times D}\| + \|\sum_{j=1}^{n_1}\Omega_{e_j,D \times D}\|}{n_1D}\right) \\ &= O_P\left(\sqrt{\frac{1}{Dm^{\frac{1}{4}}} + \frac{1}{n_1m^{\frac{1}{4}}} + m^{-\frac{1}{8}}\rho_m + \rho_m^2}\right) + O_P\left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{D}}\right) + O_P\left(\frac{\varphi_D}{D}\right) \\ &= O_P\left(\sqrt{\frac{\rho_m}{m^{\frac{1}{8}}} + \rho_m} + \frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{D}} + \frac{\varphi_D}{D}\right), \end{aligned} \quad (\text{A.2})$$

where the third equality follows from (A.1), Assumption 3.1(v), and the following bound:

$$\begin{aligned} \|\Sigma_{\bullet 1}\Sigma_{\bullet 1}^\top - \sum_{j=1}^{n_1}\Omega_{j,D \times D}\| &\leq 2\|U_{\bullet 1}\Lambda_{\bullet 1}V_{\bullet 1}^\top E_{\bullet 1}^\top\| + \|E_{\bullet 1}E_{\bullet 1}^\top - \sum_{j=1}^{n_1}\Omega_{e_j,D \times D}\| \\ &= O_P(\sqrt{Dn_1}\max(\sqrt{D}, \sqrt{n_1})) + O_P(\varphi_D\sqrt{Dn_1}) = O_P(D\sqrt{n_1} + n_1\sqrt{D}). \end{aligned}$$

For  $I$ , we can show

$$\begin{aligned}
I &\leq |(\widehat{u}_{\bullet k} - u_{\bullet k})^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top (\widehat{u}_{\bullet k} - u_{\bullet k})| + 2|\widehat{u}_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top u_{\bullet k} - \widehat{u}_{\bullet k}^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top \widehat{u}_{\bullet k}| \\
&= |(\widehat{u}_{\bullet k} - u_{\bullet k})^\top \widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top (\widehat{u}_{\bullet k} - u_{\bullet k})| + 2|\widehat{\lambda}_{\bullet k}^2 \widehat{u}_{\bullet k}^\top u_{\bullet k} - \widehat{\lambda}_{\bullet k}^2| \\
&\leq \widehat{\lambda}_{\bullet k}^2 \|\widehat{u}_{\bullet k} - u_{\bullet k}\|_2^2 + 2\widehat{\lambda}_{\bullet k}^2 |1 - \widehat{u}_{\bullet k}^\top u_{\bullet k}| \\
&\leq 2\widehat{\lambda}_{\bullet k}^2 \|\widehat{u}_{\bullet k} - u_{\bullet k}\|_2^2 \\
&= O_P\left(\frac{Dn_1\rho_m}{m^{\frac{1}{8}}} + Dn_1\rho_m^2 + D + n_1 + \frac{n_1\varphi_D^2}{D}\right),
\end{aligned}$$

where the last equality is due to (A.2) above.

We define  $\Upsilon = (v_{i,j})_{D \times n_1}$  and  $\Theta = (\varsigma_{i,j})_{D \times n_1}$ . Let  $u_{\bullet,ik}$  and  $v_{\bullet,jk}$  denote the  $(i, k)$  entry of  $U_{\bullet 1}$  and the  $(j, k)$  entry of  $V_{\bullet 1}$ , respectively. By Assumption 3.1(ii)–(iii), we have

$$E(u_{\bullet k}^\top \Upsilon v_{\bullet k}) = E\left(\sum_{i=1}^D \sum_{j=1}^{n_1} u_{\bullet,ik} v_{\bullet,jk} v_{i,j}\right) = 0,$$

and for  $j < j'$ ,

$$E(v_{i,j} v_{i,j'}) = E(v_{i,j} E(v_{i,j'} | \mathcal{F}_{i,t_{j'}})) = 0.$$

Therefore, we have

$$\begin{aligned}
E\left[\left(\sum_{i=1}^D \sum_{j=1}^{n_1} u_{\bullet,ik} v_{\bullet,jk} v_{i,j}\right)^2\right] &= \sum_{i,i'=1}^D \sum_{j,j'=1}^{n_1} u_{\bullet,ik} u_{\bullet,i'k} v_{\bullet,jk} v_{\bullet,j'k} E[v_{i,j} v_{i',j'}] \\
&= \sum_{i=1}^D \sum_{j=1}^{n_1} u_{\bullet,ik}^2 v_{\bullet,jk}^2 E[v_{i,j}^2] = O(m^{-1/4}).
\end{aligned}$$

Hence, by Chebyshev's inequality, we obtain

$$|u_{\bullet k}^\top \Upsilon v_{\bullet k}| = O_P(m^{-1/8}).$$

In addition, we have

$$E(\|u_{\bullet k}^\top \Upsilon\|_2^2) = \sum_{j=1}^{n_1} E[(\sum_{i=1}^D u_{\bullet, ik} v_{i,j})^2] = \sum_{j=1}^{n_1} \sum_{i=1}^D u_{\bullet, ik}^2 E(v_{i,j}^2) = O(n_1 m^{-1/4})$$

and

$$E(\|E_{\bullet 1}^\top u_{\bullet k}\|_2^2) = \sum_{j=1}^{n_1} E[(e_j^\top u_{\bullet k})^2] = \sum_{j=1}^{n_1} u_{\bullet k}^\top \Omega_{e_j, D \times D} u_{\bullet k} \leq n_1 \|\Omega_{e_j, D \times D}\| \|u_{\bullet k}\|^2 = O(n_1 \varphi_D).$$

Moreover,

$$|u_{\bullet k}^\top \Theta v_{\bullet k}| \leq \sum_{i=1}^D \sum_{j=1}^{n_1} |u_{\bullet, ik}| |v_{\bullet, jk}| |\varsigma_{i,j}| \leq \frac{C}{\sqrt{n_1 D}} \sum_{i=1}^D \sum_{j=1}^{n_1} |\varsigma_{i,j}| = O_P(\sqrt{n_1 D} \rho_m).$$

Since  $E(\sum_{i=1}^D u_{\bullet, ik} \varsigma_{i,j}) \leq E(\sum_{i=1}^D |u_{\bullet, jk}| |\varsigma_{i,j}|) = O(\sqrt{D} \rho_m)$ , we obtain, by Cauchy-Schwarz inequality,

$$E(\|u_{\bullet k}^\top \Theta\|_2^2) = \sum_{j=1}^{n_1} E[(\sum_{i=1}^D u_{\bullet, ik} \varsigma_{i,j})^2] \leq \sum_{j=1}^{n_1} (E(\sum_{i=1}^D u_{\bullet, ik} \varsigma_{i,j}))^2 = O(n_1 D \rho_m^2).$$

Using the facts above, we have

$$\begin{aligned} |u_{\bullet k}^\top \Upsilon \Sigma_{\bullet 1}^\top u_{\bullet k}| &\leq |u_{\bullet k}^\top \Upsilon V \Lambda U^\top u_{\bullet k}| + |u_{\bullet k}^\top \Upsilon E_{\bullet 1}^\top u_{\bullet k}| = |u_{\bullet k}^\top \Upsilon v_{\bullet k} \lambda_{\bullet k}| + |u_{\bullet k}^\top \Upsilon E_{\bullet 1}^\top u_{\bullet k}| \\ &\leq |u_{\bullet k}^\top \Upsilon v_{\bullet k}| \lambda_{\bullet k} + \|u_{\bullet k}^\top \Upsilon\|_2 \|E_{\bullet 1}^\top u_{\bullet k}\|_2 = O_P(\sqrt{n_1 D} m^{-1/8} + n_1 m^{-1/8} \sqrt{\varphi_D}), \\ |u_{\bullet k}^\top \Theta \Sigma^\top u_{\bullet k}| &\leq |u_{\bullet k}^\top \Theta V \Lambda U^\top u_{\bullet k}| + |u_{\bullet k}^\top \Theta E_{\bullet 1}^\top u_{\bullet k}| \leq |u_{\bullet k}^\top \Theta v_{\bullet k}| \lambda_{\bullet k} + \|u_{\bullet k}^\top \Theta\|_2 \|E_{\bullet 1}^\top u_{\bullet k}\|_2 \\ &= O_P(n_1 D \rho_m + n_1 \rho_m \sqrt{D \varphi_D}) = O_P(n_1 D \rho_m), \\ |u_{\bullet k}^\top \Upsilon \Theta^\top u_{\bullet k}| &\leq \|u_{\bullet k}^\top \Upsilon\|_2 \|u_{\bullet k}^\top \Theta\|_2 = O_P(n_1 \sqrt{D} m^{-1/8} \rho_m), \\ |u_{\bullet k}^\top \Upsilon \Upsilon^\top u_{\bullet k}| &= \|u_{\bullet k}^\top \Upsilon\|_2^2 = O_P(n_1 m^{-1/4}), \text{ and } |u_{\bullet k}^\top \Theta \Theta^\top u_{\bullet k}| = \|u_{\bullet k}^\top \Theta\|_2^2 = O_P(n_1 D \rho_m^2). \end{aligned}$$

Therefore, by substituting  $\widehat{\Sigma}_{\bullet 1} = \Sigma_{\bullet 1} + \Upsilon + \Theta$ , we have

$$\begin{aligned}
II &= |u_{\bullet k}^\top (\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top) u_{\bullet k}| \\
&\leq |u_{\bullet k}^\top \Sigma_{\bullet 1} \Upsilon^\top u_{\bullet k}| + |u_{\bullet k}^\top \Sigma_{\bullet 1} \Theta^\top u_{\bullet k}| + |u_{\bullet k}^\top \Upsilon \Sigma_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top \Upsilon \Upsilon^\top u_{\bullet k}| \\
&\quad + |u_{\bullet k}^\top \Upsilon \Theta^\top u_{\bullet k}| + |u_{\bullet k}^\top \Theta \Sigma_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top \Theta \Upsilon^\top u_{\bullet k}| + |u_{\bullet k}^\top \Theta \Theta^\top u_{\bullet k}| \\
&= O_P \left( \frac{\sqrt{n_1 D}}{m^{\frac{1}{8}}} + \frac{n_1 \sqrt{\varphi_D}}{m^{\frac{1}{8}}} + n_1 D \rho_m \right).
\end{aligned}$$

Note that, for each  $k \leq r$  and  $j \leq n$ ,  $\text{Var}(\mathbf{e}_j^\top u_{\bullet k}) = u_{\bullet k}^\top \boldsymbol{\Omega}_{e_j, D \times D} u_{\bullet k} \leq \|\boldsymbol{\Omega}_{e_j, D \times D}\| \|u_{\bullet k}\|^2 = O(\varphi_D)$  by using Assumption 3.1(v). Then, we have

$$\begin{aligned}
\text{Var}(v_{\bullet k}^\top E_{\bullet 1}^\top u_{\bullet k}) &= \sum_{j=1}^{n_1} v_{\bullet, jk}^2 \text{Var}(\mathbf{e}_j^\top u_{\bullet k}) + 2 \sum_{1 \leq j < l \leq n_1} v_{\bullet, jk} v_{\bullet, lk} \text{cov}(\mathbf{e}_j^\top u_{\bullet k}, \mathbf{e}_l^\top u_{\bullet k}) \\
&\leq \sum_{j=1}^{n_1} v_{\bullet, jk}^2 \text{Var}(\mathbf{e}_j^\top u_{\bullet k}) + \sum_{j=1}^{n_1} \sum_{l=1}^{n_1} |v_{\bullet, jk}| |v_{\bullet, lk}| |\text{cov}(\mathbf{e}_j^\top u_{\bullet k}, \mathbf{e}_l^\top u_{\bullet k})| \\
&\leq \sum_{j=1}^{n_1} v_{\bullet, jk}^2 \text{Var}(\mathbf{e}_j^\top u_{\bullet k}) + 2C\varphi_D \sum_{h=1}^{n_1} \rho(h) \sum_{j=1}^{n_1-h} |v_{\bullet, jk}| |v_{\bullet, (j+h)k}| = O(\varphi_D), \quad (\text{A.3})
\end{aligned}$$

where the Cauchy-Schwarz inequality and Assumption 3.1(vi) is used for the last equality. Since  $E(v_{\bullet k}^\top E_{\bullet 1}^\top u_{\bullet k}) = 0$ , we can obtain  $|v_{\bullet k}^\top E_{\bullet 1}^\top u_{\bullet k}| = O_P(\sqrt{\varphi_D})$  by Chebyshev's inequality. We note that  $|u_{\bullet k}^\top E_{\bullet 1} E_{\bullet 1}^\top u_{\bullet k}| = \|u_{\bullet k}^\top E_{\bullet 1}\|_2^2 = O_P(n_1 \varphi_D)$ . Therefore, we have

$$\begin{aligned}
III &\leq |u_{\bullet k}^\top U_{\bullet 1} \Lambda_{\bullet 1} V_{\bullet 1}^\top E_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top E_{\bullet 1} V_{\bullet 1} \Lambda_{\bullet 1} U_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top E_{\bullet 1} E_{\bullet 1}^\top u_{\bullet k}| \\
&= |\lambda_{\bullet k} v_{\bullet k}^\top E_{\bullet 1}^\top u_{\bullet k}| + |u_{\bullet k}^\top E_{\bullet 1} v_{\bullet k} \lambda_{\bullet k}| + |u_{\bullet k}^\top E_{\bullet 1} E_{\bullet 1}^\top u_{\bullet k}| = O_P(\sqrt{D n_1 \varphi_D} + n_1 \varphi_D).
\end{aligned}$$

Combining the terms  $I$ ,  $II$  and  $III$  together, for  $k \leq r$ , we have

$$|\widehat{\lambda}_{\bullet k}^2 - \lambda_{\bullet k}^2| = O_P \left( D + n_1 D \rho_m + \sqrt{D n_1 \varphi_D} + n_1 \varphi_D \right). \quad (\text{A.4})$$

Therefore, for  $k \leq r$ , we have

$$|\widehat{\lambda}_{\bullet k} - \lambda_{\bullet k}| \leq \frac{|\widehat{\lambda}_{\bullet k}^2 - \lambda_{\bullet k}^2|}{\widehat{\lambda}_{\bullet k} + \lambda_{\bullet k}} = O_P\left(\sqrt{\frac{D}{n_1}} + \sqrt{n_1 D} \rho_m + \sqrt{\varphi_D} + \sqrt{\frac{n_1}{D}} \varphi_D\right).$$

Consider (ii). Using the similar proof of (A.1), we can obtain

$$E\|\widehat{\Sigma}_{1\bullet}^\top \widehat{\Sigma}_{1\bullet} - \Sigma_{1\bullet}^\top \Sigma_{1\bullet}\|_F^2 = O(Dn^2m^{-\frac{1}{4}} + D^2nm^{-\frac{1}{4}} + D^2n^2m^{-\frac{1}{8}}\rho_m + D^2n^2\rho_m^2). \quad (\text{A.5})$$

Let  $\Lambda_{1\bullet} = \text{Diag}(\lambda_{1\bullet}, \dots, \lambda_{r\bullet})$ . Using the similar proof of (A.4), for  $k \leq r$ , we can obtain

$$|\widehat{\lambda}_{k\bullet}^2 - \lambda_{k\bullet}^2| = O_P\left(n + nD\rho_m + \sqrt{Dn\varphi_n} + D\varphi_n\right).$$

Therefore, for  $k \leq r$ , we have

$$|\widehat{\lambda}_{k\bullet} - \lambda_{k\bullet}| \leq \frac{|\widehat{\lambda}_{k\bullet}^2 - \lambda_{k\bullet}^2|}{\widehat{\lambda}_{k\bullet} + \lambda_{k\bullet}} = O_P\left(\sqrt{\frac{n}{D}} + \sqrt{nD}\rho_m + \sqrt{\varphi_n} + \sqrt{\frac{D}{n}}\varphi_n\right).$$

□

We define  $U_{\bullet 1}$  and  $V_{\bullet 1}$  as the left and right singular vector matrices of  $A_{\bullet 1} \equiv [A_{11}^\top A_{21}^\top]^\top$ , and  $U_{1\bullet}$  and  $V_{1\bullet}$  as the left and right singular vector matrices of  $A_{1\bullet} \equiv [A_{11} A_{12}]$ . The columns of  $\widehat{U}_{\bullet 1}$  and  $\widehat{V}_{\bullet 1}$  are defined as the top  $r$  eigenvectors of  $\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top$  and  $\widehat{\Sigma}_{\bullet 1}^\top \widehat{\Sigma}_{\bullet 1}$ , respectively. Similarly, the columns of  $\widehat{U}_{1\bullet}$  and  $\widehat{V}_{1\bullet}$  are defined as the top  $r$  eigenvectors of  $\widehat{\Sigma}_{1\bullet} \widehat{\Sigma}_{1\bullet}^\top$  and  $\widehat{\Sigma}_{1\bullet}^\top \widehat{\Sigma}_{1\bullet}$ , respectively.

The following lemma presents the individual convergence rate of the singular vector estimators, up to sign. We note that the SIP procedure does not require sign alignment; therefore, we focus on the convergence rates of the singular vector estimators, which are equivalent to those of the eigenvector estimators based on the sample covariance matrices.

**Lemma A.2.** *Under Assumption 3.1, if  $\log D = o(n_1)$  and  $\log n = o(D)$ , we have the*

following results:

$$\begin{aligned}
(i) \quad & \|\widehat{U}_{\bullet 1} - U_{\bullet 1}\|_{\max} = O_P \left( \frac{\rho_m^2}{\sqrt{D}} + \frac{1}{\sqrt{D}m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{n_1 D}} + \frac{\varphi_D}{D\sqrt{D}} \right), \\
(ii) \quad & \|\widehat{V}_{1\bullet} - V_{1\bullet}\|_{\max} = O_P \left( \frac{\rho_m^2}{\sqrt{n}} + \frac{1}{\sqrt{n}m^{\frac{1}{4}}} + \sqrt{\frac{\log n}{nD}} + \frac{\varphi_n}{n\sqrt{n}} \right), \\
(iii) \quad & \|\widehat{U}_{1\bullet} - U_{1\bullet}\|_{\max} = O_P \left( \frac{\rho_m^2}{\sqrt{D}} + \frac{1}{\sqrt{D}m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{nD}} + \frac{\varphi_D}{D\sqrt{D}} \right), \\
(iv) \quad & \|\widehat{V}_{\bullet 1} - V_{\bullet 1}\|_{\max} = O_P \left( \frac{\rho_m^2}{\sqrt{n_1}} + \frac{1}{\sqrt{n_1}m^{\frac{1}{4}}} + \sqrt{\frac{\log n_1}{n_1 D}} + \frac{\varphi_n}{n_1\sqrt{n_1}} \right).
\end{aligned}$$

*Proof.* We first consider (i). We have

$$\begin{aligned}
\|\widehat{\Sigma}_{\bullet 1}\widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1}\Sigma_{\bullet 1}^\top\|_{\max} &= \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} (\widehat{c}_{i,j}\widehat{c}_{s,j} - c_{i,j}c_{s,j}) \right| \\
&= \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} (v_{i,j}c_{s,j} + v_{s,j}c_{i,j}) \right| + \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} (\varsigma_{i,j}c_{s,j} + \varsigma_{s,j}c_{i,j}) \right| \\
&\quad + \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} (v_{i,j}v_{s,j} + \varsigma_{i,j}\varsigma_{s,j} + v_{i,j}\varsigma_{s,j} + \varsigma_{i,j}v_{s,j}) \right| \\
&:= I + II + III.
\end{aligned}$$

For each  $(i, s)$ , define  $Z_j^{(i,s)} = v_{i,j}c_{s,j}$ . Recall that both  $v_{i,j}$  and  $c_{s,j}$  are sub-Gaussian random variables by Assumption 3.1(iii). Then,  $Z_j^{(i,s)}$  is sub-exponential, satisfying

$$\|Z_j^{(i,s)}\|_{\psi_1} \leq C\|v_{i,j}\|_{\psi_2}\|c_{s,j}\|_{\psi_2} = O(m^{-1/8}) := \nu.$$

Since  $Z_j^{(i,s)}$  forms a martingale difference sequence, we apply a Freedman-type inequality:

$$P \left( \left| \sum_{j=1}^{n_1} Z_j^{(i,s)} \right| > t \right) \leq 2 \exp \left( -c \cdot \min \left( \frac{t^2}{n_1\nu^2}, \frac{t}{\nu} \right) \right).$$

Take  $t = Cm^{-1/8}\sqrt{n_1 \log D}$ . Then, we have  $\frac{t^2}{n_1\nu^2} = O(\log D)$  and  $\frac{t}{\nu} = O(\sqrt{n_1 \log D})$ . Under

the condition  $\log D = o(n_1)$ , we have  $\min(\frac{t^2}{n_1\nu^2}, \frac{t}{\nu}) = \frac{t^2}{n_1\nu^2}$ . Then, for some large constant  $C > 0$ , we have

$$P\left(\max_{i,s}\left|\sum_{j=1}^{n_1} Z_j^{(i,s)}\right| > t\right) \leq 2D^2 \exp\left(-c \cdot \frac{t^2}{n_1\nu^2}\right) \leq D^{-1}.$$

Therefore, we have

$$I \leq 2 \cdot \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} v_{i,j} c_{s,j} \right| = O_P(m^{-1/8} \sqrt{n_1 \log D}).$$

Similarly, define  $W_j^{(i,s)} = \varsigma_{i,j} c_{s,j}$  for each  $(i, s)$ . Under Assumption 3.1(iii), we have

$$\|W_j^{(i,s)}\|_{\psi_1} \leq C \|\varsigma_{i,j}\|_{\psi_2} \|c_{s,j}\|_{\psi_2} = O(\rho_m) := \mu.$$

For each fixed  $(i, s)$ , we apply Bernstein's inequality:

$$P\left(\left|\sum_{j=1}^{n_1} W_j^{(i,s)}\right| > t\right) \leq 2 \exp\left(-c \cdot \min\left(\frac{t^2}{n_1\mu^2}, \frac{t}{\mu}\right)\right).$$

Take  $t = C\rho_m \sqrt{n_1 \log D}$ . Since  $\log D = o(n_1)$ , it follows that  $\min(\frac{t^2}{n_1\mu^2}, \frac{t}{\mu}) = \frac{t^2}{n_1\mu^2}$ . Then, for some large constant  $C > 0$ , we have

$$P\left(\max_{i,s}\left|\sum_{j=1}^{n_1} W_j^{(i,s)}\right| > t\right) \leq 2D^2 \exp\left(-c \cdot \frac{t^2}{n_1\mu^2}\right) \leq D^{-1}.$$

Therefore, we have

$$II \leq 2 \cdot \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} \varsigma_{i,j} c_{s,j} \right| = O_P(\rho_m \sqrt{n_1 \log D}).$$

To bound  $III$ , we handle the off-diagonal ( $i \neq s$ ) and diagonal ( $i = s$ ) cases separately. We

note that for  $i \neq s$ ,  $E[v_{i,j}v_{s,j}] = 0$  by Assumption 3.1(ii). Since  $v_{i,j}$  is sub-Gaussian, we have

$$\|v_{i,j}v_{s,j}\|_{\psi_1} \leq C\|v_{i,j}\|_{\psi_2}\|v_{s,j}\|_{\psi_2} = O(m^{-1/4}) := \eta_1.$$

For each fixed  $(i, s)$ , we apply Bernstein's inequality for sub-exponential variables:

$$P\left(\left|\sum_{j=1}^{n_1} v_{i,j}v_{s,j}\right| > t\right) \leq 2\exp\left(-c \cdot \min\left(\frac{t^2}{n_1\eta_1^2}, \frac{t}{\eta_1}\right)\right).$$

Take  $t = Cm^{-1/4}\sqrt{n_1 \log D}$ . Since  $\log D = o(n_1)$ , it follows that  $\min(\frac{t^2}{n_1\eta_1^2}, \frac{t}{\eta_1}) = \frac{t^2}{n_1\eta_1^2}$ . Then, for some large constant  $C > 0$ , we have

$$P\left(\max_{i \neq s} \left|\sum_{j=1}^{n_1} v_{i,j}v_{s,j}\right| > t\right) \leq 2D^2 \exp\left(-c \cdot \frac{t^2}{n_1\eta_1^2}\right) \leq D^{-1}.$$

Therefore, we have

$$\max_{i \neq s} \left|\sum_{j=1}^{n_1} v_{i,j}v_{s,j}\right| = O(m^{-1/4}\sqrt{n_1 \log D}).$$

We also note that

$$E\left(\sum_{j=1}^{n_1} v_{i,j}^2\right) = \sum_{j=1}^{n_1} E(v_{i,j}^2) = O(n_1 m^{-1/4}).$$

For each  $i$ , we apply Bernstein's inequality:

$$P\left(\left|\sum_{j=1}^{n_1} v_{i,j}^2 - E\left[\sum_{j=1}^{n_1} v_{i,j}^2\right]\right| > t\right) \leq 2\exp\left(-c \cdot \min\left(\frac{t^2}{n_1\eta_2^2}, \frac{t}{\eta_2}\right)\right),$$

where  $\eta_2 := \|v_{i,j}^2\|_{\psi_1} = O(m^{-1/4})$ . Take  $t = Cm^{-1/4}\sqrt{n_1 \log D}$ . Since  $\log D = o(n_1)$ , it follows that  $\min(\frac{t^2}{n_1\eta_2^2}, \frac{t}{\eta_2}) = \frac{t^2}{n_1\eta_2^2}$ . Then, for some large constant  $C > 0$ , we have

$$P\left(\max_{i \leq D} \left|\sum_{j=1}^{n_1} v_{i,j}^2 - E\left[\sum_{j=1}^{n_1} v_{i,j}^2\right]\right| > t\right) \leq 2D \exp\left(-c \cdot \frac{t^2}{n_1\eta_2^2}\right) \leq D^{-1}.$$

Therefore, we have

$$\max_{i \leq D} \left| \sum_{j=1}^{n_1} v_{i,j}^2 - E \left[ \sum_{j=1}^{n_1} v_{i,j}^2 \right] \right| = O_P(m^{-1/4} \sqrt{n_1 \log D}),$$

and thus,

$$\max_{i \leq D} \left| \sum_{j=1}^{n_1} v_{i,j}^2 \right| = O_P(m^{-1/4} (n_1 + \sqrt{n_1 \log D})) = O_P(m^{-1/4} n_1).$$

By combining the diagonal and off-diagonal terms, we obtain

$$\max_{i,s \leq D} \left| \sum_{j=1}^{n_1} v_{i,j} v_{s,j} \right| = O_P(m^{-1/4} n_1). \quad (\text{A.6})$$

By using the sub-Gaussianity of  $\varsigma_{i,j}$  and the same argument as in (A.6), we also obtain

$$\max_{i,s \leq D} \left| \sum_{j=1}^{n_1} \varsigma_{i,j} \varsigma_{s,j} \right| = O_P(\rho_m^2 n_1).$$

Lastly, define  $Q_j^{(i,s)} = v_{i,j} \varsigma_{s,j}$  for each  $(i, s)$ . Then, we have

$$\|Q_j^{(i,s)}\|_{\psi_1} \leq C \|v_{i,j}\|_{\psi_2} \|\varsigma_{s,j}\|_{\psi_2} = O(m^{-1/8} \rho_m) := \pi.$$

For each fixed  $(i, s)$ , we apply a Bernstein's inequality:

$$P \left( \left| \sum_{j=1}^{n_1} Q_j^{(i,s)} \right| > t \right) \leq 2 \exp \left( -c \cdot \min \left( \frac{t^2}{n_1 \pi^2}, \frac{t}{\pi} \right) \right).$$

Take  $t = Cm^{-1/8} \rho_m \sqrt{n_1 \log D}$ . Since  $\log D = o(n_1)$ , it follows that  $\min(\frac{t^2}{n_1 \pi^2}, \frac{t}{\pi}) = \frac{t^2}{n_1 \pi^2}$ .

Then, for some large constant  $C > 0$ , we have

$$P \left( \max_{i,s} \left| \sum_{j=1}^{n_1} v_{i,j} \varsigma_{s,j} \right| > t \right) \leq 2D^2 \exp \left( -c \cdot \frac{t^2}{n_1 \pi^2} \right) \leq D^{-1}.$$

Therefore, we have

$$\max_{i,s} \left| \sum_{j=1}^{n_1} v_{i,j} \varsigma_{s,j} \right| = O_P(m^{-1/8} \rho_m \sqrt{n_1 \log D}).$$

Hence, we obtain

$$\begin{aligned} III &\leq \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} v_{i,j} v_{s,j} \right| + \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} \varsigma_{i,j} \varsigma_{s,j} \right| + 2 \cdot \max_{i,s \leq D} \left| \sum_{j=1}^{n_1} v_{i,j} \varsigma_{s,j} \right| \\ &= O_P(m^{-1/4} n_1 + \rho_m^2 n_1 + m^{-1/8} \rho_m \sqrt{n_1 \log D}). \end{aligned}$$

Combining the bounds for terms  $I, II$  and  $III$  together, we obtain

$$\|\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top\|_{\max} = O_P \left( m^{-\frac{1}{8}} \sqrt{n_1 \log D} + \rho_m \sqrt{n_1 \log D} + m^{-\frac{1}{4}} n_1 + \rho_m^2 n_1 \right). \quad (\text{A.7})$$

Let  $\widehat{U}_{\bullet 1} = (\widehat{u}_{\bullet 1}, \dots, \widehat{u}_{\bullet r})$ ,  $U_{\bullet 1} = (u_{\bullet 1}, \dots, u_{\bullet r})$ , and  $\Lambda_{\bullet 1} = \text{Diag}(\lambda_{\bullet 1}, \dots, \lambda_{\bullet r})$ . We denote the eigengap  $\bar{\gamma} = \min\{\lambda_{\bullet k}^2 - \lambda_{\bullet k+1}^2 : 1 \leq k \leq r\}$  and  $\lambda_{\bullet r+1} = 0$ . For  $k \leq r$ ,  $\lambda_{\bullet k} \asymp \sqrt{n_1 D}$  and  $\|u_{\bullet k}\|_\infty \leq C/\sqrt{D}$ . In addition, the coherence  $\mu(U_{\bullet 1}) = D \max_i \sum_{k=1}^r u_{\bullet, ik}^2 / r \leq C$ , where  $u_{\bullet, ik}$  is the  $(i, k)$  entry of  $U_{\bullet 1}$ . Thus, by Theorem 1 of [Fan et al. \(2018b\)](#), we have

$$\begin{aligned} \max_{k \leq r} \|\widehat{u}_{\bullet k} - u_{\bullet k}\|_\infty &\leq C \frac{\|\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - U_{\bullet 1} \Lambda_{\bullet 1}^2 U_{\bullet 1}^\top\|_\infty}{\bar{\gamma} \sqrt{D}} \\ &\leq C \frac{\|\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top\|_\infty + \|\Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top - \sum_{j=1}^{n_1} \Omega_{j, D \times D}\|_\infty + \|\sum_{j=1}^{n_1} \Omega_{j, D \times D} - U_{\bullet 1} \Lambda_{\bullet 1}^2 U_{\bullet 1}^\top\|_\infty}{\bar{\gamma} \sqrt{D}} \\ &= O \left( \frac{D \|\widehat{\Sigma}_{\bullet 1} \widehat{\Sigma}_{\bullet 1}^\top - \Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top\|_{\max}}{n_1 D^{\frac{3}{2}}} + \frac{D \|\Sigma_{\bullet 1} \Sigma_{\bullet 1}^\top - \sum_{j=1}^{n_1} \Omega_{j, D \times D}\|_{\max}}{n_1 D^{\frac{3}{2}}} + \frac{\sum_{j=1}^{n_1} \|\Omega_{e_j, D \times D}\|_1}{n_1 D^{\frac{3}{2}}} \right) \\ &= O_P \left( \sqrt{\frac{\log D}{n_1 D m^{\frac{1}{4}}}} + \rho_m \sqrt{\frac{\log D}{n_1 D}} + \frac{1}{\sqrt{D} m^{\frac{1}{4}}} + \frac{\rho_m^2}{\sqrt{D}} \right) + O_P \left( \sqrt{\frac{\log D}{n_1 D}} \right) + O_P \left( \frac{\varphi_D}{D \sqrt{D}} \right) \\ &= O_P \left( \frac{1}{\sqrt{D} m^{\frac{1}{4}}} + \frac{\rho_m^2}{\sqrt{D}} + \sqrt{\frac{\log D}{n_1 D}} + \frac{\varphi_D}{D \sqrt{D}} \right), \end{aligned}$$

where the second equality follows from (A.7) and Assumptions 3.1(iv)–(v). By using a similar argument, we can obtain the results (ii)–(iv).  $\square$

The following lemma presents the individual convergence rate of singular value estimators for the partitioned matrix  $A_{11}$ .

**Lemma A.3.** *Under Assumption 3.1, if  $\log D = o(n_1)$ ,  $\log n = o(D)$ ,  $\varphi_n \sqrt{D} \leq n_1$  and  $\varphi_D \sqrt{n_1} \leq D$ , we have*

$$\begin{aligned} & \|\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11} - \Lambda_{11}\|_{\max} \\ &= O_P \left( \rho_m \sqrt{Dn_1} + \frac{\sqrt{Dn_1}}{m^{\frac{1}{4}}} + \sqrt{D \log D} + \sqrt{\frac{n_1}{D}} \varphi_D + \sqrt{n_1 \log n_1} + \sqrt{\frac{D}{n_1}} \varphi_n \right). \end{aligned}$$

*Proof.* Since  $\Lambda_{11} = U_{11}^\top A_{11} V_{11}$ , we have

$$\begin{aligned} \|\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11} - \Lambda_{11}\|_{\max} &= \|\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11} - U_{11}^\top A_{11} V_{11}\|_{\max} \\ &\leq \|\widehat{U}_{11}^\top (\widehat{\Sigma}_{11} - A_{11}) \widehat{V}_{11}\|_{\max} + \|(\widehat{U}_{11} - U_{11})^\top A_{11} (\widehat{V}_{11} - V_{11})\|_{\max} \\ &\quad + \|(\widehat{U}_{11} - U_{11})^\top A_{11} V_{11}\|_{\max} + \|U_{11}^\top A_{11} (\widehat{V}_{11} - V_{11})\|_{\max} \\ &:= \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4. \end{aligned}$$

In order to find the convergence rate of elementwise norm, we consider the above terms separately. For  $\Delta_1$ , we have

$$\begin{aligned} \Delta_1 &= \|\widehat{U}_{11}^\top (\widehat{\Sigma}_{11} - \Sigma_{11} + \Sigma_{11} - A_{11}) \widehat{V}_{11}\|_{\max} \\ &\leq \|\widehat{U}_{11}^\top (\widehat{\Sigma}_{11} - \Sigma_{11}) \widehat{V}_{11}\|_{\max} + \|\widehat{U}_{11}^\top E_{11} \widehat{V}_{11}\|_{\max} \\ &\leq \|\widehat{U}_{11}^\top (\widehat{\Sigma}_{11} - \Sigma_{11}) \widehat{V}_{11}\|_{\max} + \|U_{11}^\top E_{11} V_{11}\|_{\max} + \|(\widehat{U}_{11} - U_{11})^\top E_{11} (\widehat{V}_{11} - V_{11})\|_{\max} \\ &\quad + \|(\widehat{U}_{11} - U_{11})^\top E_{11} V_{11}\|_{\max} + \|U_{11}^\top E_{11} (\widehat{V}_{11} - V_{11})\|_{\max} \\ &:= I + II + III + IV + V. \end{aligned}$$

Let  $U_{11} = (u_{i,k})_{(D-1) \times r}$  and  $V_{11} = (v_{j,k})_{n_1 \times r}$ . For each  $(k, l)$ , define  $S_{kl} = \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{i,j} v_{j,l}$ .

We note that, by Assumption 3.1(ii)–(iii),  $E(v_{i,j} v_{i',j'}) = 0$  unless  $i = i'$  and  $j = j'$ . Then,

for any  $k, l \leq r$ , we have

$$E(S_{kl}) = \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{j,l} E(v_{i,j}) = 0$$

and

$$\text{Var}(S_{kl}) = \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k}^2 v_{j,l}^2 E[v_{i,j}^2] = O(m^{-1/4}).$$

Then, by Chebyshev's inequality, we have

$$\left| \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{i,j} v_{j,l} \right| = O_P(m^{-1/8}).$$

By Assumption 3.1(ii)–(iii), we have

$$\left| \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} \varsigma_{i,j} v_{j,l} \right| \leq \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} |u_{i,k}| |\varsigma_{i,j}| |v_{j,l}| = O_P(\rho_m \sqrt{Dn_1}).$$

Since  $r$  is fixed, we have

$$\begin{aligned} \|U_{11}^\top (\widehat{\Sigma}_{11} - \Sigma_{11}) V_{11}\|_{\max} &\leq \max_{k,l \leq r} \left| \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{i,j} v_{j,l} \right| + \max_{k,l \leq r} \left| \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} \varsigma_{i,j} v_{j,l} \right| \\ &= O_P(m^{-1/8} + \rho_m \sqrt{Dn_1}). \end{aligned}$$

Therefore, we have

$$I = O_P(\|U_{11}^\top (\widehat{\Sigma}_{11} - \Sigma_{11}) V_{11}\|_{\max}) = O_P(m^{-1/8} + \rho_m \sqrt{Dn_1}).$$

Using the same argument as (A.3), we have

$$\text{Var}\left(\sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{j,l} e_{i,j}\right) = O(\varphi_D)$$

and hence,

$$\left| \sum_{i=1}^{D-1} \sum_{j=1}^{n_1} u_{i,k} v_{j,l} e_{i,j} \right| = O_P(\sqrt{\varphi_D}).$$

Since  $r$  is fixed, we can apply a union bound over  $r^2$  terms without affecting the rate.

Therefore, we have

$$II = O(\sqrt{\varphi_D}).$$

We note that  $\widehat{U}_{11}$  and  $\widehat{U}_{1\bullet}$  are equal up to column-wise sign changes; the same holds for  $\widehat{V}_{11}$  and  $\widehat{V}_{\bullet 1}$ . We define  $Q_{k,j} = \sum_{i=1}^{D-1} u_{i,k} e_{i,j}$ . We note that

$$\text{Var}(Q_{k,j}) = u_k^\top \boldsymbol{\Omega}_{e_j, D \times D} u_k \leq \max_{j \leq n} \|\boldsymbol{\Omega}_{e_j, D \times D}\| \|u_k\|^2 = O(\varphi_D).$$

Using the bounded fourth moment condition in Assumption 3.1(vi) and Markov's inequality, we have

$$P(|Q_{k,j}| > t) \leq \frac{C\varphi_D^2}{t^4}.$$

By applying the union bound over all  $(k, j)$  pairs, we then have

$$P(\max_{k,j} |Q_{k,j}| > t) \leq Crn_1 \cdot \frac{\varphi_D^2}{t^4}.$$

Therefore, we have

$$\|U_{11}^\top E_{11}\|_{\max} = O_P(\varphi_D^{1/2} n_1^{1/4}) := O_P(\kappa_u).$$

Similarly, we have  $\|E_{11} V_{11}\|_{\max} = O_P(\varphi_n^{1/2} D^{1/4}) := O_P(\kappa_v)$ . Then, by Lemma A.2 (iii)–(iv), we can obtain

$$\begin{aligned} IV &= O_P(\kappa_v D \|\widehat{U}_{1\bullet} - U_{1\bullet}\|_{\max}) \\ &= O_P\left(\kappa_v \sqrt{D} \left(\rho_m^2 + \frac{1}{m^{\frac{1}{4}}}\right) + \kappa_v \sqrt{\frac{D \log D}{n}} + \frac{\kappa_v \varphi_D}{\sqrt{D}}\right), \\ V &= O_P(\kappa_u n_1 \|\widehat{V}_{\bullet 1} - V_{\bullet 1}\|_{\max}) \end{aligned}$$

$$= O_P \left( \kappa_u \sqrt{n_1} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \kappa_u \sqrt{\frac{n_1 \log n_1}{D}} + \frac{\kappa_u \varphi_n}{\sqrt{n_1}} \right).$$

Similarly, we can show that the term *III* is dominated by *IV* and *V*. Therefore, we have

$$\begin{aligned} \Delta_1 &= O_P \left( \rho_m \sqrt{Dn_1} + \sqrt{\varphi_D} + \kappa_v \sqrt{D} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \kappa_v \sqrt{\frac{D \log D}{n}} + \frac{\kappa_v \varphi_D}{\sqrt{D}} \right. \\ &\quad \left. + \kappa_u \sqrt{n_1} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \kappa_u \sqrt{\frac{n_1 \log n_1}{D}} + \frac{\kappa_u \varphi_n}{\sqrt{n_1}} \right). \end{aligned}$$

For  $\Delta_3$  and  $\Delta_4$ , by Lemma A.2 (iii)–(iv), we have

$$\begin{aligned} \Delta_3 &= \|(\widehat{U}_{11} - U_{11})^\top U_{11} \Lambda_{11}\|_{\max} = O_P(D\sqrt{n_1} \|\widehat{U}_{1\bullet} - U_{1\bullet}\|_{\max}) \\ &= O_P \left( \sqrt{n_1 D} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \sqrt{\frac{n_1 D \log D}{n}} + \varphi_D \sqrt{\frac{n_1}{D}} \right), \\ \Delta_4 &= \|\Lambda_{11} V_{11}^\top (\widehat{V}_{11} - V_{11})\|_{\max} = O_P(n_1 \sqrt{D} \|\widehat{V}_{\bullet 1} - V_{\bullet 1}\|_{\max}) \\ &= O_P \left( \sqrt{n_1 D} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \sqrt{n_1 \log n_1} + \varphi_n \sqrt{\frac{D}{n_1}} \right). \end{aligned}$$

Similarly, we can show  $\Delta_2$  is dominated by  $\Delta_3$  and  $\Delta_4$ . Combining the terms  $\Delta_1$ ,  $\Delta_3$  and  $\Delta_4$  together, we have

$$\begin{aligned} &\|\widehat{U}_{11}^\top \widehat{\Sigma}_{11} \widehat{V}_{11} - \Lambda_{11}\|_{\max} \\ &= O_P \left( \rho_m \sqrt{Dn_1} + \sqrt{\varphi_D} + \max(\kappa_v, \sqrt{n_1}) \left( \sqrt{D} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \sqrt{\frac{D \log D}{n}} + \frac{\varphi_D}{\sqrt{D}} \right) \right. \\ &\quad \left. + \max(\kappa_u, \sqrt{D}) \left( \sqrt{n_1} \left( \rho_m^2 + \frac{1}{m^{\frac{1}{4}}} \right) + \sqrt{\frac{n_1 \log n_1}{D}} + \frac{\varphi_n}{\sqrt{n_1}} \right) \right) \\ &= O_P \left( \rho_m \sqrt{Dn_1} + \frac{\sqrt{Dn_1}}{m^{\frac{1}{4}}} + \sqrt{D \log D} + \sqrt{\frac{n_1}{D}} \varphi_D + \sqrt{n_1 \log n_1} + \sqrt{\frac{D}{n_1}} \varphi_n \right). \end{aligned}$$

□

## A.2 Proof of Theorem 3.1

We recall that  $\tilde{\Sigma}_{22} = \hat{\Sigma}_{21} \hat{V}_{11} (\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11})^{-1} \hat{U}_{11}^\top \hat{\Sigma}_{12} = \hat{A}_{21} \hat{V}_{11} (\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11})^{-1} \hat{U}_{11}^\top \hat{A}_{12}$ . By Lemmas A.1–A.2, we can show

$$\begin{aligned} \frac{1}{\sqrt{n_1}} \|\hat{A}_{21} \hat{V}_{11} - A_{21} V_{11}\|_{\max} &\leq \frac{1}{\sqrt{n_1}} \|\hat{A}_{\bullet 1} \hat{V}_{11} - A_{\bullet 1} V_{11}\|_{\max} = \frac{1}{\sqrt{n_1}} \|\hat{U}_{\bullet 1} \hat{\Lambda}_{\bullet 1} - U_{\bullet 1} \Lambda_{\bullet 1}\|_{\max} \\ &\leq \frac{1}{\sqrt{n_1}} \|\hat{U}_{\bullet 1} (\hat{\Lambda}_{\bullet 1} - \Lambda_{\bullet 1})\|_{\max} + \frac{1}{\sqrt{n_1}} \|(\hat{U}_{\bullet 1} - U_{\bullet 1}) \Lambda_{\bullet 1}\|_{\max} \\ &= O_P \left( \frac{1}{\sqrt{n_1 D}} \|\hat{\Lambda}_{\bullet 1} - \Lambda_{\bullet 1}\|_{\max} + \sqrt{D} \|\hat{U}_{\bullet 1} - U_{\bullet 1}\|_{\max} \right) \\ &= O_P \left( \rho_m + \frac{\varphi_D}{D} + \frac{1}{m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{n_1}} \right) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{D}} \|\hat{U}_{11}^\top \hat{A}_{12} - U_{11}^\top A_{12}\|_{\max} &\leq \frac{1}{\sqrt{D}} \|\hat{U}_{11}^\top \hat{A}_{1\bullet} - U_{11}^\top A_{1\bullet}\|_{\max} = \frac{1}{\sqrt{D}} \|\hat{\Lambda}_{1\bullet} \hat{V}_{1\bullet}^\top - \Lambda_{1\bullet} V_{1\bullet}^\top\|_{\max} \\ &\leq \frac{1}{\sqrt{D}} \|(\hat{\Lambda}_{1\bullet} - \Lambda_{1\bullet}) \hat{V}_{1\bullet}^\top\|_{\max} + \frac{1}{\sqrt{D}} \|\Lambda_{1\bullet} (\hat{V}_{1\bullet} - V_{1\bullet})^\top\|_{\max} \\ &= O_P \left( \frac{1}{\sqrt{nD}} \|\hat{\Lambda}_{1\bullet} - \Lambda_{1\bullet}\|_{\max} + \sqrt{n} \|\hat{V}_{1\bullet} - V_{1\bullet}\|_{\max} \right) \\ &= O_P \left( \rho_m + \frac{\varphi_n}{n} + \frac{1}{m^{\frac{1}{4}}} + \sqrt{\frac{\log n}{D}} \right). \end{aligned}$$

By Lemma A.3, we can obtain

$$\begin{aligned} \|(\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11})^{-1} - \Lambda_{11}^{-1}\|_{\max} &= O_P \left( \frac{1}{n_1 D} \|\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11} - \Lambda_{11}\|_{\max} \right) \\ &= O_P \left( \frac{\rho_m}{\sqrt{n_1 D}} + \frac{1}{\sqrt{n_1 D} m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{n_1^2 D}} + \frac{\varphi_D}{\sqrt{n_1 D^3}} + \sqrt{\frac{\log n_1}{n_1 D^2}} + \frac{\varphi_n}{\sqrt{n_1^3 D}} \right). \end{aligned}$$

By using the above results, we have

$$\begin{aligned} \max_{j \leq n_2} |\tilde{c}_{D, n_1+j} - E[c_{D, n_1+j} | \mathcal{F}_{D, n_1}]| &= \|\hat{\Sigma}_{21} \hat{V}_{11} (\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11})^{-1} \hat{U}_{11}^\top \hat{\Sigma}_{12} - A_{21} (V_{11} \Lambda_{11}^{-1} U_{11}^\top) A_{12}\|_{\infty} \\ &= \|\hat{A}_{21} \hat{V}_{11} (\hat{U}_{11}^\top \hat{\Sigma}_{11} \hat{V}_{11})^{-1} \hat{U}_{11}^\top \hat{A}_{12} - A_{21} (V_{11} \Lambda_{11}^{-1} U_{11}^\top) A_{12}\|_{\infty} \end{aligned}$$

$$\begin{aligned}
&\leq \|\widehat{A}_{21}\widehat{V}_{11}((\widehat{U}_{11}^\top\widehat{\Sigma}_{11}\widehat{V}_{11})^{-1} - \Lambda_{11}^{-1})\widehat{U}_{11}^\top\widehat{A}_{12}\|_\infty + \|(\widehat{A}_{21}\widehat{V}_{11} - A_{21}V_{11})\Lambda_{11}^{-1}(\widehat{U}_{11}^\top\widehat{A}_{12} - U_{11}^\top A_{12})\|_\infty \\
&\quad + \|A_{21}V_{11}\Lambda_{11}^{-1}(\widehat{U}_{11}^\top\widehat{A}_{12} - U_{11}^\top A_{12})\|_\infty + \|(\widehat{A}_{21}\widehat{V}_{11} - A_{21}V_{11})\Lambda_{11}^{-1}U_{11}^\top A_{12}\|_\infty \\
&= O_P\left(\sqrt{n_1 D}\|(\widehat{U}_{11}^\top\widehat{\Sigma}_{11}\widehat{V}_{11})^{-1} - \Lambda_{11}^{-1}\|_{\max} + \frac{1}{\sqrt{n_1}}\|\widehat{A}_{21}\widehat{V}_{11} - A_{21}V_{11}\|_{\max}\right. \\
&\quad \left. + \frac{1}{\sqrt{D}}\|\widehat{U}_{11}^\top\widehat{A}_{12} - U_{11}^\top A_{12}\|_{\max}\right) \\
&= O_P\left(\rho_m + \frac{1}{m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{n_1}} + \frac{\varphi_D}{D} + \sqrt{\frac{\log n_1}{D}} + \frac{\varphi_n}{n_1} + \rho_m + \frac{\varphi_D}{D} + \frac{1}{m^{\frac{1}{4}}} + \sqrt{\frac{\log D}{n_1}}\right. \\
&\quad \left. + \rho_m + \frac{\varphi_n}{n} + \frac{1}{m^{\frac{1}{4}}} + \sqrt{\frac{\log n}{D}}\right) \\
&= O_P\left(\rho_m + \frac{1}{m^{\frac{1}{4}}} + \frac{\varphi_D}{D} + \sqrt{\frac{\log D}{n_1}} + \frac{\varphi_n}{n_1} + \sqrt{\frac{\log n}{D}}\right).
\end{aligned}$$

□