Factor Modeling for Volatility

Yi Ding * Robert Engle † Yingying Li ‡ Xinghua Zheng §

November 20, 2022

Abstract

We establish a framework to study the factor structure in stock variance under a high-frequency and high-dimensional setup. We prove the consistency of conducting principal component analysis on realized variances in estimating the factor structure. Moreover, based on strong empirical evidence, we propose a multiplicative volatility factor (MVF) model, where stock variance is represented by a common variance factor and a multiplicative lognormal idiosyncratic component. We further show that our MVF model leads to significantly improved volatility prediction. The favorable performance of the proposed MVF model is seen in both US stocks and global equity indices.

Keywords: Volatility modeling; Factor model; High-frequency data; High-dimension; Principal component analysis.

JEL Codes: C13, C51, C53, C55, C58, G17

*Faculty of Business Administration, University of Macau, Taipa, Macau. Email: yiding@um.edu.mo

†Stern School of Business, New York University, 44 West Fourth Street Suite 9-62, New York. Email: rengle@stern.nyu.edu

‡Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Research is supported in part by RGC grants GRF 16503419, GRF 16502118 and T31-64/18-N of the HKSAR and NSFC 19BM03. Email: yyli@ust.hk

§Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Research is supported in part by RGC grants GRF-16304521, GRF 16304019 and T31-64/18-N of the HKSAR. Email: xhzheng@ust.hk

1 Introduction

Volatilities–intrinsically linked with macro- and microeconomics–play a central role in investments, asset pricing, risk management and monetary policies. Nowadays, we have witnessed the impact of geopolitical turmoils, pandemics and climate change on the clearly more than ever uncertain and connected global economy. The imminent era of long-lasting instability raises challenges and amplifies the imperativeness of a better understanding of the volatilities in the financial market.

The past few decades have seen tremendous progress in modeling time-varying volatilities. Well-known volatility models include the autoregressive conditional heteroskedasticity (ARCH) model and the generalized ARCH (GARCH) model (Engle (1982); Bollerslev (1986)), as well as stochastic volatility models (Clark (1973); Taylor (1982)).

Thanks to the availability of high-frequency data and recent developments in volatility measuring with high-frequency data, we can now estimate daily volatilities with high accuracy. The realized volatility (RV) is a consistent estimator of the integrated volatility (IV) as the sampling frequency increases when microstructure noise is absent; see, for example, Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002). Various robust IV estimators have been proposed when there is microstructure noise, including the two-scale realized volatility (Zhang, Mykland, and Aït-Sahalia (2005)), multi-scale realized volatility (Zhang et al. (2006)), pre-averaging approach (Jacod, Li, Mykland, Podolskij, and Vetter (2009), Jacod, Li, and Zheng (2019)), and quasi-maximum likelihood estimator (Xiu (2010)). Jump-robust volatility estimators have also been proposed, including the bipower variation by Barndorff-Nielsen and Shephard (2004) and the truncated RV by Mancini (2009). Gonçalves and Meddahi (2009) and Hounyo, Gonçalves, and Meddahi (2017) develop bootstrap methods for inference on integrated volatility. Li, Liu, and Xiu (2019) propose an efficient multi-scale jackknife estimator for integrated volatility functionals. Empirically, Liu, Patton, and Sheppard (2015) find that the simple 5-minute RVs achieve high estimation accuracies. RV-based models have been further studied for volatility prediction, including fractionally-integrated Gaussian vector autoregression for log RV (Andersen, Bollerslev, Diebold, and Labys (2003)) and heterogeneous AR

Model (HAR, Corsi (2009)).

To have a concrete idea of daily volatilities in the market, we compute the daily RVs of S&P 500 Index constituent stocks using 5-minute intraday returns between 2003 and 2020, pick out a few stocks, more specifically, the stocks that have their mean RVs on the 30%, 50%, and 70% quantiles, and plot the time series of their RVs in Figure 1.

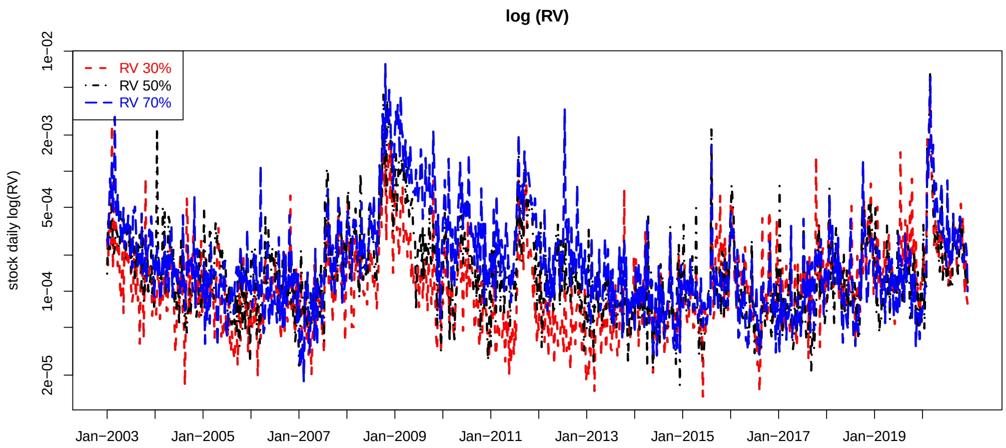

Figure 1: Time series plots (log scale) of three representative S&P500 Index constituent stocks' RVs (RV 30%, RV 50%, RV 70%) based on 5-minute intraday returns from 2003 to 2020 with mean RVs falling on the 30%, 50%, 70% quantiles of all mean RVs.

Figure 1 shows clearly that the stock RVs co-move. Such a co-movement feature in volatilities has been well-documented. For example, Engle, Ito, and Lin (1990) and Calvet, Fisher, and Thompson (2006) examine exchange markets, Susmel and Engle (1994); Da and Schaumburg (2006) and Kelly, Lustig, and Van Nieuwerburgh (2013) study equities, and Bollerslev, Hood, Huss, and Pedersen (2018) and Engle and Martin (2019) study global multiple asset classes. The volatility co-movement has been used in volatility forecasting; see, for example, Luciani and Veredas (2015); Asai and McAleer (2015); Barigozzi and Hallin (2017), and Bollerslev, Hood, Huss, and Pedersen (2018).

The co-movement in volatility is not surprising as it is well known that returns admit a factor structure, such as the capital asset pricing model (CAPM, Sharpe (1964)), Fama-French three-factor/five-factor/six-factor models (FF3/FF5/FF6, Fama and French (1993, 2015, 2018)), multi-factor and approximate factor models (Ross (1976); Chamberlain and Rothschild (1983); Chen, Roll, and Ross (1986)). When volatilities are stochastic, the factor structure in returns will induce a factor structure in variance, which leads to the volatility co-movement. An interesting question then arises:

Is the co-movement in volatility purely due to the factor structure in returns?

We find that it is not the case.

First, factors in volatilities can exist even when the return factor is absent. Consider the following example. Suppose that there are N assets with returns $\mathbf{R}_t = (R_{1t}, ..., R_{Nt})^T$. They follow a single factor model, $\mathbf{R}_t = \boldsymbol{\beta} f_t + \mathbf{U}_t$, where $\mathbf{U}_t = (U_{1t}, ..., U_{Nt})^T$ are the idiosyncratic returns. Suppose $U_{it} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{u;it})$, where $\sigma^2_{u;it} = a + b\delta_t + z_{it}$, $a \geq 0, b > 0$, and δ_t and z_{it} are independent positive random variables. Under such a model, the idiosyncratic returns $(U_{it})_{1 \leq i \leq N}$ are uncorrelated, and hence do not admit a factor structure. The idiosyncratic variances, however, have a common factor δ_t.

An idiosyncratic variance factor structure that is unlikely induced by omitted return factors is indeed what we find in the empirical data. Our sample includes the high-frequency data of 291 constituent stocks in the S&P 500 Index between 2003 and 2020. We take the FF3 model as an example. We obtain 5-minute estimated idiosyncratic returns and the daily idiosyncratic realized variances by regressing the 5-minute intraday stock returns over the FF3 factors. The principal component analysis (PCA) on the idiosyncratic returns does not suggest a clear factor structure. In contrast, the PCA on the idiosyncratic RVs suggests a clear factor structure. The cross-sectional average of the idiosyncratic variances can be approximately considered as the factor in idiosyncratic variances. The results are consistent with the findings of Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016).

Second, the number of factors in stock variance is not necessarily the same as the number of return factors. In fact, we find strong empirical evidence for a single factor in stock variance. We take the FF3 model again as an example. When performing PCA on four variance factor candidates, namely, the RVs of FF3 factors and the common idiosyncratic realized variance factor, we find that they are strongly correlated with each other, and there exists a common component in the four variance factor candidates. The common component in the four variance factors is strongly correlated with the first principal component (PC) in stock RVs, which explains a majority proportion (60%) of the total variation in stock RVs. The evidence strongly suggests a common component in stock variances. Our empirical evidence about a single variance factor in the stock variances agrees with the findings of Kapadia, Linn, and Paye (2020), who find a common factor in both market volatility and market neutral volatilities of a wide range of return factors.

The empirical evidence described above is obtained using PCA on realized variances. Under the high-dimensional setting, PCA is consistent in identifying factor structure in returns; see, for example, Bai and Ng (2002); Bai (2003); Fan, Liao, and Mincheva (2013); Aït-Sahalia and Xiu (2017), and Ding, Li, and Zheng (2021). However, unlike returns, volatilities are not observable and have to be estimated. The resulting estimated variance contains errors that accumulate as the dimension increases. This leads to an important question:

Is PCA-based estimation valid in identifying the factor structure in stock variances?

To address this question, we establish a framework to study the factor structure in stock variance under a high-frequency and high-dimensional setup. In brief, one estimates stock integrated variances using realized variances and then conducts PCA on the sample covariance matrix of the stock realized variances. We prove the consistency of such a procedure in estimating the factor structure in stock variance. Specifically, we develop statistical theories about the explicit convergence rate in estimating the population covariance matrix of the stock variance using the sample covariance matrix of the stock realized variances. Furthermore, we obtain the convergence rate in using PCA to estimate the factor structure when it exists in stock variance. We also obtain the consistency results for the factor structure estimation in idiosyncratic variances using PCA on idiosyncratic realized variances, which are based on estimated idiosyncratic returns from regressing the high-frequency stock returns over factor returns. It is worth pointing out that our setting is different than the usual error-in-variable setting because the errors in RVs are not i.i.d..

Next, we investigate the following question:

What model could suitably describe the volatility co-movements?

We propose a multiplicative volatility factor (MVF) model for stock volatility. We observe a high correlation between the first PC and the cross-sectional average of stock realized variances, which we term common realized variance (CRV). It suggests that the cross-sectional average of variance, or common variance (CV), can be approximately considered as the single factor in stock variance. We also find strong empirical evidence from the analysis of the usual additive linear model that the variance factor model is in a surprisingly neat multiplicative format. In our proposed MVF model, the variance is represented by a multiplicative common factor and a multiplicative lognormal idiosyncratic component, which we term the idiosyncratic variance exposure.

Our proposed MVF model has a number of desirable properties. First, the model captures important volatility characteristics such as non-negativity and heavy-tailedness. Second, it incorporates the co-movement in the volatility in a simple way, which makes the model estimation straightforward. The common multiplicative factor can be well approximated by the common variance. The idiosyncratic variance exposure is then simply the variance divided by the common variance, based on which, the two model coefficients, mean and standard deviation of the idiosyncratic lognormal component can be estimated. The simplicity of the MVF model makes it particularly attractive for the study of volatility in a high-dimensional context. Third, the MVF model enjoys internal model consistency. To be more specific, under the MVF model, the volatility of a stock portfolio inherits the common factor from the underlying stock volatilities, while other factor model structures such as the log-linear factor model do not have such a desirable property.

Finally, we address the following question:

How much could the MVF model be helpful?

We mainly answer this question from the perspective of volatility forecasting. Under our proposed MVF model, we simply predict the CV factor and the multiplicative idiosyncratic components separately using log HAR models. The stock volatility forecast is the multiplication of the forecasts of the two components. We use the MVF model to predict daily volatilities of the S&P 500 Index constituent stocks between 2004 and 2020. We find that our approach outperforms the benchmark models by generating lower Q-like losses. When checking stock by stock, the outperformance of our approach is statistically significant in a majority (around 90%) of the stocks we evaluate.

Our proposed MVF model is built upon empirical evidence from the US stocks. Beyond the US market, we also find that the MVF model applies to the global market based on a parallel analysis using daily realized variances of 31 global equity indices. In addition, the MVF model outperforms in the global market volatility forecasting.

To summarize, our contributions lie in the following aspects:

First, we develop a framework to study the factor structure in stock variance (and idiosyncratic variance) using high-frequency data under a high-dimensional setup.

Second, we establish theoretical support for using PCA on realized variances to estimate the factor structure in stock variance (and idiosyncratic variance).

Third, we propose a single factor volatility model with a multiplicative idiosyncratic component, the MVF model, based on strong empirical evidence in US stocks. Our MVF model has several desirable features that make it attractive in various applications.

Fourth, we utilize the proposed MVF model for volatility forecasting. Our model performs dominantly well compared with various benchmark approaches.

Last but not least, we show that our MVF model applies to the global market and helps predict the volatilities of global equity indices.

The rest of this paper is organized as follows. In Section 2, we discuss the evidence of factor structure in variance. We develop the MVF model in Section 3. Section 4 presents the out-of-sample volatility forecasting results. In Section 5, we examine our MVF model in the global equity indices. Section 6 contains concluding remarks. Proofs and additional empirical results are collected in the Supplementary Materials Ding, Engle, Li, and Zheng (2022).

2 Evidence of Factor Model for Volatility

2.1 Data

We focus on the S&P 500 Index constituent stocks in 2003 and exclude the least liquid stocks that have more than 20% zero 5-minute returns from January 2003 to December 2020. We collect high-frequency stock prices from the TAQ database. Following the common data cleaning procedure (e.g., Aït-Sahalia and Mancini (2008)), "bounce back"s are removed. We sample the log prices starting from 9:35 until 16:00, using the previous-tick approach (Gençay, Dacorogna, Muller, Pictet, and Olsen (2001)). Regarding the sampling frequency, we use 5-minute log-returns for which the market microstructure noise can be safely ignored (Liu, Patton, and Sheppard (2015)). Holidays, half trading days and overnight returns are eliminated. Same as the treatment in Li and Xiu (2016), we also remove May 6, 2010, the day when the "Flash Crash" occurred. After the cleaning procedure, we obtain 291 stocks for 4491 trading days in 2003–2020, and each stock has 77 5-minute intraday log-returns per day. About return factors, we consider the Fama-French three-factor model (Fama and French (1993)) and use 5-minute returns of the market, the small-minus-big (SmB) and the high-minus-low (HmL) portfolios.[1]

Following Bollerslev and Todorov (2011); Aït-Sahalia, Fan, and Li (2013) and Li, Todorov, and Tauchen (2017), for each stock i and each day t, we estimate the continuous component of variance with $RV_{it}^c = \sum_{j=1}^{77} (R_{i;t[j]}^{tr})^2$, where $R_{i;t[j]}^{tr} = R_{i;t[j]} \mathbf{1}_{\{|R_{i;t[j]}| \leq v_{it}\}}$, $1 \leq j \leq 77$, and v_{it} is set to be $v_{it} = 3\sqrt{\min(RV_{it}, BV_{it})} \times \Delta_n^{0.49}$, $\Delta_n = 1/77$, $RV_{it} = \sum_{j=1}^{[1/\Delta_n]} R_{i;t[j]}^2$, and $BV_{it} = \frac{\pi}{2} \frac{[1/\Delta_n]}{[1/\Delta_n]-1} \sum_{j=2}^{[1/\Delta_n]} |R_{i;t[j]} R_{i;t[j-1]}|$ is the bipower variation (Barndorff-Nielsen and Shephard (2004)). We apply the same truncation procedure to the high-frequency factor data to obtain the continuous component of the factor return. The truncated factor returns are denoted by $\boldsymbol{F}^{tr}$. Our analysis is based on the truncated returns $\boldsymbol{R}^{tr}$, truncated factor returns $\boldsymbol{F}^{tr}$ and the continuous component of realized variance, RV^c. For notational ease, when there is no ambiguity, we denote the truncated return $\boldsymbol{R}^{tr}$ by $\boldsymbol{R}$ and the continuous component of realized variance RV^c by RV.

[1]We thank Saketh Aleti for sharing high-frequency factor data from the paper Aleti (2022).

2.2 Factor Structure in Idiosyncratic Variances

We analyze daily idiosyncratic realized variances constructed from 5-minute returns of S&P 500 Index stocks and the Fama-French three-factor model. Specifically, we regress the 5-minute intraday returns over the Fama-French three factors, from which we get the idiosyncratic returns and the corresponding idiosyncratic realized variances. We find that PCA on the idiosyncratic returns shows no evidence of a factor structure. In contrast, PCA on the idiosyncratic realized variances shows that the first PC accounts for a large proportion (49%) of the total variation in idiosyncratic realized variances. In addition, consistent with the results in Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016), we find that the first PC has a high correlation (0.955) with the cross-sectional average of the idiosyncratic realized variances, that is, the common idiosyncratic realized volatility (CiRV). Therefore, the cross-sectional average of idiosyncratic variance, or common idiosyncratic variance (CiV),[2] can be considered as the factor in idiosyncratic variance. Beyond the Fama-French three factors, we also use statistical factors to check the robustness of our findings. The results are similar.[3]

2.3 Factor Structure in Stock Variances

The factor structure in stock returns naturally induces a factor structure in stock variance. For example, under the FF3 model, we have

$$\begin{aligned} V_{it} = & \beta^2_{iMkt} V_{Mkt\,t} + \beta^2_{iHmL} V_{HmL\,t} + \beta^2_{iSmB} V_{SmB\,t} + V_{U_i\,t} \\ & + \text{covariance terms}, \end{aligned}$$

where V_{it} and $V_{U_i\,t}$ denote stock and idiosyncratic variances, respectively, and $V_{Mkt\,t}$, $V_{HmL\,t}$ and $V_{SmB\,t}$ are the factor variances. Hence, the return factor variances (V_{Mkt}, V_{HmL}, V_{SmB}) are potential factors for the stock variance. Moreover, as discussed in

[2]In Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016), CIV refers to the cross-sectional average of standard deviations, while in our paper, we refer to CiRV as the cross-sectional average of idiosyncratic realized variances. We refer to CiV as the cross-sectional average of integrated idiosyncratic variance. Despite such a difference, we still use the name CiV. We find this name nicely summarizes the most important first principle component.

[3]The detailed results are available upon request.

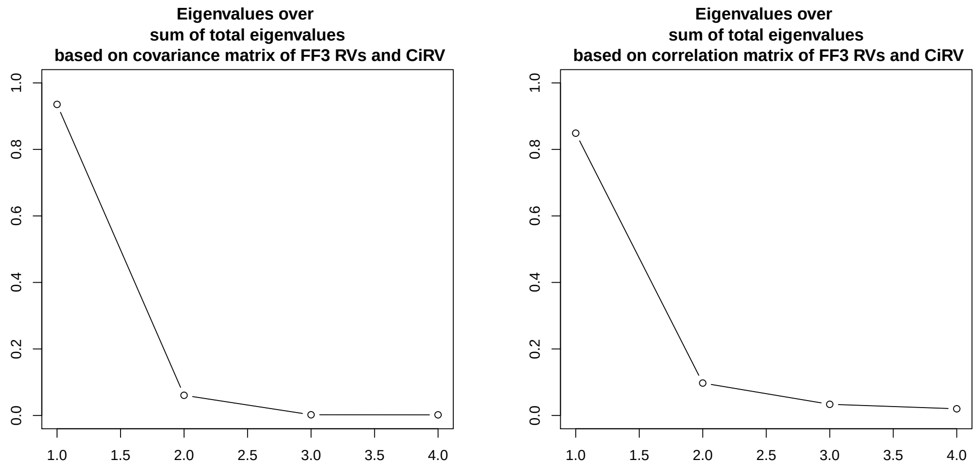

Figure 2: Eigenvalue ratios of the sample covariance matrix (left panel) and sample correlation matrix (right panel) of FF3 factors' RVs and CiRV.

the previous section, a single factor exists in idiosyncratic variance, CiV. Therefore, altogether, there are four potential factors in stock variances. An interesting question arises, namely, are there indeed *four* factors?

To address this question, we first employ PCA on the four variance factor candidates, namely, the three return factor realized variances (RV_{Mkt}, RV_{HmL}, RV_{SmB}), and the common idiosyncratic realized variance, CiRV. We compute the eigenvalue ratios, which are the eigenvalues divided by the sum of the total eigenvalues, and plot the results in Figure 2. Surprisingly, we find that the first PC explains more than 90% of the total variation in the four variance factor candidates, suggesting a single common component.

We then perform PCA directly[4] on the stock RVs, and compute the ratio of the top eigenvalues over the sum of the total eigenvalues, based on both the covariance matrix and the correlation matrix of the stock RVs. The results are plotted in Figure 3.

Figure 3 shows that a high proportion (60%) of the total variation in stock RVs can be explained by the first PC, while the second and other PCs do not account for a proportion substantially higher than the remaining. These observations suggest a single factor model for the stock variances. We also estimate the number of factors

[4]Outliers are removed by 95% winsorization to avoid the effect of extreme variations in the RVs.

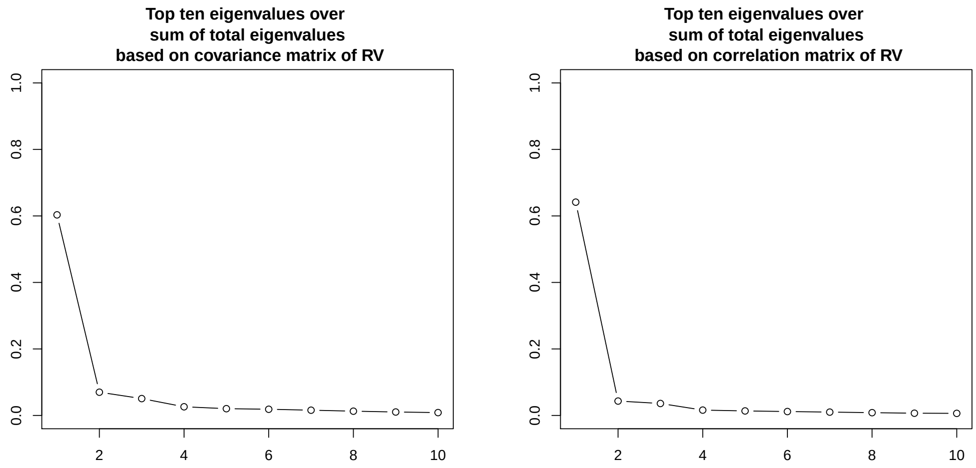

Figure 3: Top ten eigenvalue ratios of the sample covariance matrix (left panel) and sample correlation matrix (right panel) of stock RVs.

using estimators from Bai and Ng (2002) and Ahn and Horenstein (2013), and the results also suggest a single factor model for the stock variances.

We next compute the pairwise correlations among the variance factor candidates, which are the return factor RVs (RV_{Mkt}, RV_{HmL}, RV_{SmB}), the CiRV, and in addition, the first PC (PC_{RV}) in stock RVs. The results are summarized in Table 1.

Table 1: Pairwise correlations among RV_{Mkt}, RV_{HmL}, RV_{SmB}, CiRV, and the first PC in the stock RVs.

	CiRV	RV_{Mkt}	RV_{HmL}	RV_{SmB}
PC_{RV}	0.970	0.868	0.800	0.827
CiRV		0.848	0.788	0.855
RV_{Mkt}			0.680	0.920
RV_{HmL}				0.689

Table 1 shows that all variance factor candidates are highly correlated with an average pairwise correlation of around 0.80, and are also highly correlated with the first PC in the stock RVs. These results are consistent with the findings in Li, Todorov, and Tauchen (2016), which show a high correlation between the spot market factor volatilities and idiosyncratic volatilities of sector portfolios. Kapadia, Linn, and Paye (2020) also have similar findings about the high correlation between market volatility and the market neutral volatilities of a wide range of return factors.

In summary, we find compelling evidence for a single factor structure in stock variance.

2.4 PCA Consistency of Using Realized Variance in Estimating Factors in Integrated Variance

The empirical studies in Sections 2.2 and 2.3 are based on realized (idiosyncratic) variances, which inevitably contain estimation errors. Because both the number of assets and the time span are large, the estimation errors accumulate. In this section, we analyze the consistency of conducting PCA on realized variances in identifying factor structure in integrated variances.

2.4.1 Continuous-time Log Price Process

We consider the following continuous-time factor model for log prices:

$$d\boldsymbol{Y}_t = \boldsymbol{\alpha}_t dt + \boldsymbol{\beta} d\boldsymbol{X}_t + d\boldsymbol{Z}_t, \quad t \in [0, T], \tag{2.1}$$

where $(\boldsymbol{Y}_t)$ is an N-dimensional log price process, $(\boldsymbol{X}_t)$ is a K-dimensional factor process, $(\boldsymbol{Z}_t)$ is the idiosyncratic component, $\boldsymbol{\alpha}_t =: (\alpha_{1t}, ..., \alpha_{NT})^T$ is the drift term, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N)^T =: (\beta_{ik})$ is a factor loading matrix of dimension $N \times K$.

For any matrix $A = (a_{ij})$, we denote the entrywise norm as $\|A\|_{\max} := \max_{i,j} |a_{ij}|$; the spectrum norm is denoted as $\|\mathbf{A}\|_2 := \max_{\|\mathbf{x}\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2$, where $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$; and the minimum singular value and the maximum singular value are denoted as $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively.

We make the following assumptions.

Assumption 1. *$(\boldsymbol{X}_t)$ and $(\boldsymbol{Z}_t)$ are continuous Itô semimartingales:*

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{h}_s ds + \int_0^t \boldsymbol{\eta}_s d\boldsymbol{W}_s, \qquad \boldsymbol{Z}_t = \int_0^t \boldsymbol{\zeta}_s d\boldsymbol{B}_s,$$

where $(\boldsymbol{W}_t)$ *and* $(\boldsymbol{B}_t)$ *are independent Brownian motions, and* $\boldsymbol{h}_t$ *is the drift term for factors. We write the spot covariance matrix of* $(\boldsymbol{X}_t)$ *and* $(\boldsymbol{Z}_t)$ *as* $\boldsymbol{\Phi}_t = \boldsymbol{\eta}_t \boldsymbol{\eta}_t^T$ *and* $\boldsymbol{\Theta}_t = \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^T$*. The processes* $(\boldsymbol{\eta}_t)$ *and* $(\boldsymbol{\zeta}_t)$ *are càdlàg, and* $\boldsymbol{\Phi}_t$*,* $\boldsymbol{\Phi}_{t-}$*,* $\boldsymbol{\Theta}_t$ *and* $\boldsymbol{\Theta}_{t-}$ *are positive definite. In addition,* $\max\left(\|\boldsymbol{\beta}\|_{\max}, \sup_{s\geq 0}\|\boldsymbol{\alpha}_s\|_{\max}, \sup_{s\geq 0}\|\boldsymbol{h}_s\|_{\max}\right) \leq C$*, and* $\lambda_{\min}\left(E(\int_0^1 \boldsymbol{\Phi}_s ds)\right) > c$ *for some constants* $c, C > 0$*.*

Finally, the mixing coefficients $\rho(\chi) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_\chi^\infty} |P(AB) - P(A)P(B)|$*, where* $\mathcal{F}_{-\infty}^0$*,* $\mathcal{F}_\chi^\infty$ *are* σ*-algebras generated by* $\{(\boldsymbol{\Phi}_t, \boldsymbol{\Theta}_t) : -\infty \leq t \leq 0\}$ *and* $\{(\boldsymbol{\Phi}_t, \boldsymbol{\Theta}_t) : \chi \leq t \leq \infty\}$*, respectively, satisfy that* $\rho(\chi) \leq c_1 \exp(-c_2\chi)$ *for some constants* $c_1, c_2 > 0$ *and any positive integer* χ*.*

For each day $t = 1, ..., T$, we denote the integrated variances and integrated idiosyncratic variances of N stocks as $V_t = (V_{1t}, ..., V_{Nt})^T$ and $V_{\boldsymbol{U};t} = (V_{\boldsymbol{U};1t}, ..., V_{\boldsymbol{U};Nt})^T$, respectively, namely,

$$V_{it} = \int_{t-1}^{t} \boldsymbol{\Psi}_{\tau,ii} d\tau, \quad \text{and} \quad V_{\boldsymbol{U};it} = \int_{t-1}^{t} \boldsymbol{\Theta}_{\tau,ii} d\tau, \quad 1 \leq i \leq N, \tag{2.2}$$

where $\boldsymbol{\Psi}_t = \boldsymbol{\beta}\boldsymbol{\Phi}_t\boldsymbol{\beta}^T + \boldsymbol{\Theta}_t$. We define $V_{\boldsymbol{F};kt} = \int_{t-1}^t \boldsymbol{\Phi}_{\tau,kk} d\tau$, $1 \leq k \leq K$, and $V_{\boldsymbol{F};t} = (V_{\boldsymbol{F};1t}, ..., V_{\boldsymbol{F};Kt})^T$.

Suppose that we observe log-returns of stocks and factors at sampling frequency Δ_n. For each $t = 1, ..., T$ and $j = 1, ..., n := [1/\Delta_n]$, we write the log-returns of stocks and factors as $\boldsymbol{R}_{t[j]}$ and $\boldsymbol{F}_{t[j]}$, respectively, where

$$\begin{aligned} \boldsymbol{R}_{t[j]} &= \boldsymbol{Y}_{t-1+\Delta_n j} - \boldsymbol{Y}_{t-1+\Delta_n(j-1)} =: (R_{1,t[j]}, ..., R_{N,t[j]})^T, \quad \text{and} \\ \boldsymbol{F}_{t[j]} &= \boldsymbol{X}_{t-1+\Delta_n j} - \boldsymbol{X}_{t-1+\Delta_n(j-1)} =: (F_{1,t[j]}, ..., F_{K,t[j]})^T. \end{aligned}$$

Model (2.1) induces a factor model for high-frequency returns:

$$\begin{aligned} &\boldsymbol{R}_{t[j]} = \boldsymbol{\alpha}_{n;t[j]} + \boldsymbol{\beta}\boldsymbol{F}_{t[j]} + \boldsymbol{U}_{t[j]}, \quad \boldsymbol{\alpha}_{n;t[j]} = \int_{t-1+\Delta_n(j-1)}^{t-1+\Delta_n j} \boldsymbol{\alpha}_s ds, \quad \text{and} \\ &\boldsymbol{U}_{t[j]} = \boldsymbol{Z}_{t-1+\Delta_n j} - \boldsymbol{Z}_{t-1+\Delta_n(j-1)} =: (U_{1,t[j]}, ..., U_{N,t[j]})^T, \end{aligned} \tag{2.3}$$

where $\boldsymbol{U}_{t[j]}$ is an N-dimensional vector of idiosyncratic returns.

The realized variance of stock i on day t, RV_{it}, is defined as $RV_{it} = \sum_{j=1}^n R_{i,t[j]}^2$. It is consistent in estimating the integrated variance and enjoys $\sqrt{n}$ rate of convergence.

2.4.2 Stock Variance Factor Estimation

Next, we present the theoretical results for the estimation of factor structure in stock variance. We make the following assumptions on the stock volatility processes.

Assumption 2. *The integrated variances* $(\int_{t-1}^{t} \mathbf{\Psi}_{\tau,ii} d\tau)_{1\le i\le N}$ *are stationary, and* $\sup_{t\in\mathbb{N}} E\Big((\sup_{t-1\le s<t} \mathbf{\Psi}_{s,ii})^M\Big) \le k_\delta$ *for some positive constants* k_δ, $M > 0$ *and for all* $t \in \mathbb{N}$, $1 \le i \le N$.

Assumption 3. *The covariance matrix of integrated variances,* $\mathbf{\Sigma}_V = \mathrm{Cov}\left(V_t\right)$, *satisfies that for some constants* $c, C > 0$, $c \le \lambda_{V,i}/N < \lambda_{V,i-1}/N \le C$ *and* $\lambda_{V,i-1}/N - \lambda_{V,i}/N > c$ *for* $1 \le i \le q$, *and* $c \le \lambda_{V,i} \le C$ *for* $q < i \le N$, *where* $\lambda_{V,1} \ge ... \ge \lambda_{V,N}$ *are the eigenvalues of* $\mathbf{\Sigma}_V$, *and* q *is a fixed positive integer.*

Assumption 3 is a standard assumption in factor models (Bai (2003); Fan, Liao, and Mincheva (2013)). It implies that the integrated variances admit a factor structure with q (strong) factors.

To estimate $\mathbf{\Sigma}_V$, we use the sample covariance matrix of the realized variances,

$$\widehat{\mathbf{\Sigma}}_{RV} = \frac{1}{T}\sum_{t=1}^{T}(RV_t - \overline{RV})(RV_t - \overline{RV})^T,$$

where $RV_t = (RV_{1t}, ..., RV_{NT})^T$ and $\overline{RV} = \sum_{t=1}^{T} RV_t/T$. We denote the ith eigenvector of $\mathbf{\Sigma}_{V_U}$ by $\boldsymbol{\xi}_{V_U;i}$, the ith largest eigenvalue of $\widehat{\mathbf{\Sigma}}_{RV}$ by $\widehat{\lambda}_{RV_i}$ and the corresponding eigenvector by $\widehat{\boldsymbol{\xi}}_{RV_i}$, $1 \le i \le N$.

The next theorem gives the error bound of $\widehat{\mathbf{\Sigma}}_{RV}$ in estimating $\mathbf{\Sigma}_V$.

Theorem 1. *Under Assumptions 1 and 2, if* $\log T/|\log \Delta_n| = O(1)$, $N = O(T^\gamma)$ *for some* $\gamma > 0$, *and the* M *in Assumption 2 satisfies* $M > 4(1 + 2\gamma)$, *then*

$$\|\widehat{\mathbf{\Sigma}}_{RV} - \mathbf{\Sigma}_V\|_{\max} = O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}}\right). \tag{2.4}$$

In addition, if Assumption 3 holds, then

$$\max_{1\le i\le q}\left|\frac{\widehat{\lambda}_{RV_i}}{\lambda_{V_i}}-1\right| = O_p\left(\sqrt{\Delta_n}+\sqrt{\frac{\log N}{T}}\right), \quad \textit{and} \tag{2.5}$$

$$\max_{1\le i\le q}\|\widehat{\boldsymbol{\xi}}_{RV_i}-\boldsymbol{\xi}_{V_i}\|_2 = O_p\left(\sqrt{\Delta_n}+\sqrt{\frac{\log N}{T}}\right). \tag{2.6}$$

Theorem 1 guarantees that if a factor structure exists in the stock variance, then it can be consistently estimated by conducting PCA on the stock RV as long as $\max\Big(\Delta_n, (\log N)/T\Big) \to 0$.

2.4.3 Idiosyncratic Variance Factor Estimation

In this subsection, we give the consistency results of conducting PCA on realized idiosyncratic variances in identifying factor structure in integrated idiosyncratic variances.

Under the factor model setup, we obtain the factor loading estimator $\widehat{\boldsymbol{\beta}} =: (\widehat{\boldsymbol{\beta}}_1, ..., \widehat{\boldsymbol{\beta}}_N)^T$, the estimator $\widehat{\boldsymbol{\alpha}_n} = (\widehat{\alpha_{n1}}, ..., \widehat{\alpha_{nN}})^T$ of the average drift $\overline{\boldsymbol{\alpha}_n} = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^T\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{\alpha}_{n;t[}$ and the idiosyncratic return estimator $\widehat{\boldsymbol{U}}_{t[j]} =: (\widehat{U}_{1,t[j]}, ..., \widehat{U}_{N,t[j]})^T$. Specifically,

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{R}_{t[j]}-\overline{\boldsymbol{R}})(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})^T\right)\left(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})^T\right)^{-1},$$
$$\widehat{\boldsymbol{\alpha}_n} = \overline{\boldsymbol{R}} - \widehat{\boldsymbol{\beta}}\,\overline{\boldsymbol{F}}, \quad \text{and} \quad \widehat{\boldsymbol{U}}_{t[j]} = \boldsymbol{R}_{t[j]} - \widehat{\boldsymbol{\alpha}_n} - \widehat{\boldsymbol{\beta}}\boldsymbol{F}_{t[j]}, \tag{2.7}$$

where $\overline{\boldsymbol{R}} = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^T\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{R}_{t[j]}$, and $\overline{\boldsymbol{F}} = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^T\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{F}_{t[j]}$. The feasible idiosyncratic realized variance is defined as follows:

$$RV_{\widehat{\boldsymbol{U}};it} = \sum_{j=1}^{[1/\Delta_n]}\widehat{U}^2_{i,t[j]} = \sum_{j=1}^{[1/\Delta_n]}(R_{i,t[j]} - \widehat{\alpha_{ni}} - \widehat{\boldsymbol{\beta}}_i^T\boldsymbol{F}_{t[j]})^2, \quad 1\le i\le N, 1\le t\le T. \tag{2.8}$$

We then estimate $\boldsymbol{\Sigma}_{V_U}$ using the sample covariance matrix of $(RV_{\widehat{\boldsymbol{U}};it})_{1\le i\le N, 1\le t\le T}$:

$$\widehat{\boldsymbol{\Sigma}}_{RV_{\widehat{U}}} = \frac{1}{T}\sum_{t=1}^{T}(RV_{\widehat{\boldsymbol{U}};t} - \overline{RV}_{\widehat{\boldsymbol{U}}})(RV_{\widehat{\boldsymbol{U}};t} - \overline{RV}_{\widehat{\boldsymbol{U}}})^T,$$

where $RV_{\widehat{\boldsymbol{U}};t} = (RV_{\widehat{\boldsymbol{U}};1t}, ..., RV_{\widehat{\boldsymbol{U}};Nt})^T$ and $\overline{RV}_{\widehat{\boldsymbol{U}}} = \frac{1}{T}\sum_{t=1}^{T} RV_{\widehat{\boldsymbol{U}};t}$.

We make the following assumptions on the integrated factor variances and the integrated idiosyncratic variances.

Assumption 4. *$(\int_{t-1}^{t} \boldsymbol{\Phi}_{\tau,jk} d\tau)_{1\le j,k\le K}$ and $(\int_{t-1}^{t} \boldsymbol{\Theta}_{\tau,ii} d\tau)_{1\le i\le N}$ are stationary, and there exist k_δ and M such that for all $t \ge 1$, $1 \le k \le K$ and $1 \le i \le N$, $\max\Big(\sup_{t\in\mathbb{N}} E(\sup_{t-1\le s\le t} \boldsymbol{\Phi}_{s,kk}^M), \sup_{t\in\mathbb{N}} E(\sup_{t-1\le s\le t} \boldsymbol{\Theta}_{s,ii}^M)\Big) \le k_\delta$.*

Assumption 5. *The covariance matrix of integrated idiosyncratic variance, $\boldsymbol{\Sigma}_{V_{\boldsymbol{U}}} = \mathrm{Cov}\Big(V_{\boldsymbol{U};t}\Big)$, satisfies that for some constants $c, C > 0$, one has $c \le \lambda_{V_{\boldsymbol{U}};i}/N < \lambda_{V_{\boldsymbol{U}};i-1}/N \le C$ for $1 \le i \le r$, and $c \le \lambda_{V_{\boldsymbol{U}};i} \le C$ for $r < i \le N$, where $\lambda_{V_{\boldsymbol{U}};1} \ge \lambda_{V_{\boldsymbol{U}};2} \ge, ..., \ge \lambda_{V_{\boldsymbol{U}};N}$ are the eigenvalues of $\boldsymbol{\Sigma}_{V_{\boldsymbol{U}}}$, and r is a fixed positive integer.*

We denote by $\boldsymbol{\xi}_{V_U;i}$ the ith eigenvector of $\boldsymbol{\Sigma}_{V_U}$, $1 \le i \le N$. For the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{RV_{\widehat{\boldsymbol{U}}}}$, the eigenvectors and eigenvalues are denoted by $\widehat{\boldsymbol{\xi}}_{RV_{\widehat{\boldsymbol{U}};i}}$ and $\widehat{\lambda}_{RV_{\widehat{\boldsymbol{U}};i}}$, $1 \le i \le N$, respectively. The next theorem gives the error bound of using $\widehat{\boldsymbol{\Sigma}}_{RV_{\widehat{\boldsymbol{U}}}}$ to estimate $\boldsymbol{\Sigma}_{V_{\boldsymbol{U}}}$.

Theorem 2. *Under Assumptions 1, 2 and 4, if $\log T/|\log \Delta_n| = O(1)$, $N = O(T^\gamma)$ for some $\gamma > 0$ and the M in Assumption 4 satisfies $M > 4(1 + 2\gamma)$, then*

$$\|\widehat{\boldsymbol{\Sigma}}_{RV_{\widehat{\boldsymbol{U}}}} - \boldsymbol{\Sigma}_{V_{\boldsymbol{U}}}\|_{\max} = O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}}\right). \tag{2.9}$$

In addition, if Assumption 5 holds, then

$$\max_{1\le i\le r}\left|\frac{\widehat{\lambda}_{RV_{\widehat{\boldsymbol{U}}};i}}{\lambda_{V_{\boldsymbol{U}};i}} - 1\right| = O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}}\right), \quad \textit{and} \tag{2.10}$$

$$\max_{1\le i\le r}\|\widehat{\boldsymbol{\xi}}_{RV_{\widehat{\boldsymbol{U}}};i} - \boldsymbol{\xi}_{V_{\boldsymbol{U}};i}\|_2 = O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}}\right). \tag{2.11}$$

Theorem 2 guarantees that if a factor structure exists in the idiosyncratic variance, then the factor structure can be consistently estimated by conducting PCA on the idiosyncratic RV provided that $\max\Big(\Delta_n, (\log N)/T\Big) \to 0$.

3 Factor Modeling for Stock Volatility

3.1 Common Variance Factor

The empirical evidence in Section 2.3 suggests that both return factor variances and idiosyncratic variances are driven by a single variance factor. In order to construct the single factor, by Theorem 1, we can use the first PC in stock RVs. Alternatively, one can take the common variance (CV) of stocks, which is defined as the cross-sectional average of stock integrated variance:

$$CV_t = \frac{1}{N}\sum_{i=1}^{N} V_{it}.$$

Correspondingly, the common realized variance (CRV) is defined as follows:

$$CRV_t = \frac{1}{N}\sum_{i=1}^{N} RV_{it}.$$

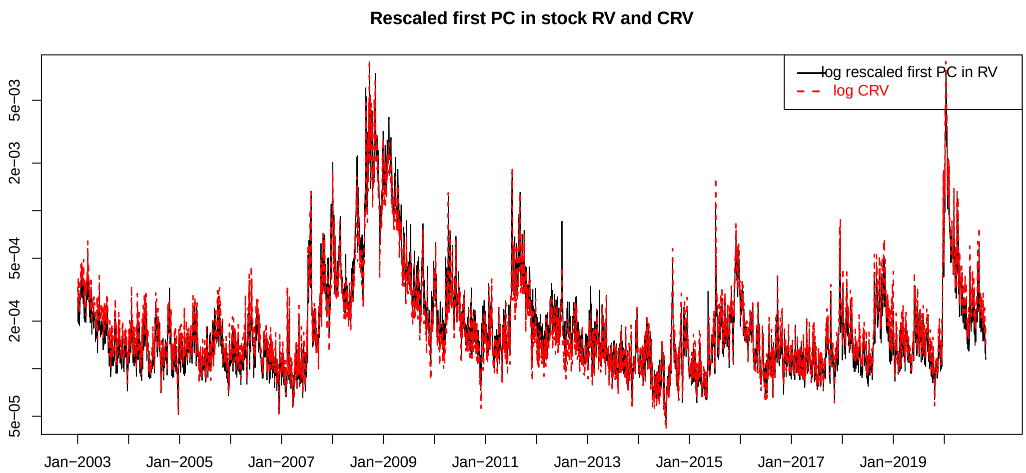

Figure 4: Time series plots of common realized variance (CRV) and rescaled first PC in stock RVs. The plots are drawn on a log scale to improve visibility.

In Figure 4, we plot the time series of CRV and the first PC in stock RV rescaled to have the same mean and standard deviation as CRV. We find that the first PC in RVs largely coincides with CRV. They have a high correlation of 0.979. The results suggest that CV can be approximately considered as the single factor in stock variance. Compared with the first PC, the CV factor is more interpretable and easier to estimate.

Remark 1. *Besides CV, we also evaluate VIX (the CBOE Market Volatility Index, transformed to daily variance) as a possible candidate for the volatility factor. The correlation between VIX and the first PC in stock RVs is lower than the correlation between CRV and the first PC (0.842 vs. 0.979). VIX measures implied volatility for the future and is more informative in longer monthly/yearly horizons (see, e.g., the discussion in Andersen and Benzoni (2010)). In addition, it carries a volatility risk premium, which complicates volatility forecasting. Given the nature of VIX, we consider CV as a more appropriate factor proxy in stock volatility. Nevertheless, in practice, when predicting volatility, people often find VIX to be helpful. In our volatility prediction method to be introduced in Section 4, replacing CRV with VIX leads to similar performance.*

3.2 Evidence about the Multiplicative Factor Structure

3.2.1 Single Factor Model for Variance

The empirical evidence from conducting PCA on stock RVs suggests the following factor model:

$$V_{it} = a_{\xi,i} + b_{\xi,i}\xi_t + \varepsilon_{\xi,it}, \quad 1 \le i \le N, \tag{3.1}$$

where ξ_t is the single (latent) factor. Taking average over i on both sides yields

$$CV_t = \frac{1}{N}\sum_{i=1}^{N} V_{it} = \bar{a}_\xi + \bar{b}_\xi \xi_t + \bar{\varepsilon}_{\xi,t}, \tag{3.2}$$

where $\bar{a}_\xi = \sum_{i=1}^{N} a_{\xi,i}/N$, $\bar{b}_\xi = \sum_{i=1}^{N} b_{\xi,i}/N$, and $\bar{\varepsilon}_{\xi,t} = \sum_{i=1}^{N} \varepsilon_{\xi,it}/N$. We can hence rewrite model (3.1) using CV as the variance factor:

$$V_{it} = a_i + b_i CV_t + \varepsilon_{it}, \quad 1 \le i \le N, \tag{3.3}$$

where $a_i = a_{\xi,i} - \bar{a}_\xi b_{\xi,i}/\bar{b}_\xi + E(\varepsilon_{\xi,it}) - b_{\xi,i}E(\bar{\varepsilon}_{\xi,t})/\bar{b}_\xi$, $b_i = b_{\xi,i}/\bar{b}_\xi$, and $\varepsilon_{it} = \varepsilon_{\xi,it} - E(\varepsilon_{\xi,it}) - b_{\xi,i}\big(\bar{\varepsilon}_{\xi,t} - E(\bar{\varepsilon}_{\xi,t})\big)/\bar{b}_\xi$.

To estimate the coefficients a_i and b_i in model (3.3), we regress RV_{it} over CRV_t, and obtain

$$\widehat{b}_i = \frac{\sum_{t=1}^T (CRV_t - \overline{CRV})(RV_{it} - \overline{RV}_i)}{\sum_{t=1}^T (CRV_t - \overline{CRV})^2}, \quad \widehat{a}_i = \overline{RV}_i - \widehat{b}_i \overline{CRV}, \text{ for } 1 \le i \le N, \tag{3.4}$$

where $\overline{CRV} = \sum_{t=1}^T CRV_t/T$, and $\overline{RV}_i = \sum_{t=1}^T RV_{it}/T$.

The next result shows that the estimators $(\widehat{a}_i, \widehat{b}_i)$ are consistent under the following mild assumptions.

Assumption 6. *The factor process (ξ_t) is stationary. $(\boldsymbol{\varepsilon}_{\xi,t}) = \Big((\varepsilon_{\xi,1t}, \varepsilon_{\xi,2t}, ..., \varepsilon_{\xi,Nt})^T\Big)$ is stationary and uncorrelated with ξ_t. Moreover, $|\bar{b}_\xi| > c$, and $\|\operatorname{Cov}(\boldsymbol{\varepsilon}_{\xi,t})\|_2 \le C$ for some constant $c, C > 0$.*

Proposition 1. *Under the assumptions of Theorem 1 and Assumption 6,*

$$\begin{aligned} \max_{1\le i\le N} |\widehat{b}_i - b_i| &= O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}} + \sqrt{\frac{1}{N}}\right), \\ \max_{1\le i\le N} |\widehat{a}_i - a_i| &= O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}} + \sqrt{\frac{1}{N}}\right). \end{aligned} \tag{3.5}$$

3.2.2 From Additive to Multiplicative

We estimate model (3.3) using the S&P 500 Index constituent stock RVs. We have the following interesting findings.

First, when checking the idiosyncratic component ε_{it}, we find a strong correlation between ε_{it}^2 and CV_t^2, while the correlation between $\varepsilon_{it}^2/CV_t^2$ and CV_t^2 is almost zero. This result suggests that ε_{it} scales with CV_t. In addition, after checking the distribution of ε_{it}/CV_t, we find that ε_{it} can be well modeled by the multiplication of CV_t and a centered lognormal random variable, namely, $\varepsilon_{it} = CV_t\big(\exp(\mu_i + \sigma_i z_{it}) - \exp(\mu_i + \sigma_i^2/2)\big)$. Details are given in Appendix A.1 of the Supplementary Material.

Second, when further analyzing the coefficients a_i and b_i, strong evidence suggests that the intercept terms $(a_i)_{1\le i\le N}$ in (3.3) are close to zero. In addition, the slope term b_i in (3.3) and $\exp(\mu_i + \sigma_i^2/2)$, that is, the expectation of the lognormal term, are approximately equal. These results motivate us to impose the restrictions: $a_i = 0$, and $b_i = \exp(\mu_i + \sigma_i^2/2), 1 \leq i \leq N$.

Combining the findings above yields a multiplicative factor model, which we present in the next subsection. The detailed analysis results are relegated to Appendix A in the Supplementary Material.

3.3 Main Model: Multiplicative Volatility Factor Model

The analysis in Section 3.2.2 leads us to propose the following Multiplicative Volatility Factor (MVF) model:

$$V_{it} = \xi_t \exp(\mu_i + \sigma_i z_{it}), \quad 1 \leq i \leq N, \tag{3.6}$$

where ξ_t is the multiplicative latent factor, and $\exp(\mu_i + \sigma_i z_{it})$ is a multiplicative idiosyncratic component, where $z_{it} \sim \mathcal{N}(0, 1)$ and is independent with ξ_t. We denote $\widetilde{V}_{it} = \exp(\mu_i + \sigma_i z_{it})$ and refer to it as idiosyncratic variance exposure (iE). We allow $\log(\widetilde{V}_{it})$ to be dependent over time. The latent factor ξ_t can be well approximated by the common variance CV_t. Correspondingly, $\widetilde{V}_{it} = V_{it}/CV_t$.

The MVF model (3.6) has several desirable properties: 1) it reflects important volatility characteristics such as non-negativity and heavy-tailedness; 2) it has a simple form that eases the model estimation and volatility prediction; 3) it enjoys *internal model consistency* in the following sense: if a model applies to individual assets, then it also applies to portfolios of the assets. We explain the three properties in more detail below.

First, volatility is usually modeled in a *multiplicative* way. Examples include the lognormal stochastic volatility model (Hull and White (1987)), EGARCH (Nelson (1991)), and realized-GARCH with log-linear specification (Hansen, Huang, and Shek (2012)). The multiplicative model (3.6) naturally captures important features in volatilities such as non-negativity and heavy-tailedness.

Second, the MVF model (3.6) has a surprisingly neat format, which makes the model estimation and volatility prediction very straightforward. The common factor is the common variance, and the only two model parameters, μ_i and σ_i, can be easily estimated using the sample mean and standard deviation of $\log(\widetilde{RV}_{it})$, where $\widetilde{RV}_{it} = RV_{it}/CRV_t$ is the idiosyncratic realized variance exposure (iRE). For volatility forecasting, $\log(CV_t)$ and $\log(\widetilde{V}_{it})$ can be separately modeled by, for example, a HAR model (Corsi (2009)).

Third, the MVF model (3.6) is a special case of the additive linear factor model (3.1), and hence describes the factor structure in variance. When evaluating portfolio risk, the variance of a portfolio involves a linear combination of underlying stock variances. Under our MVF model, the portfolio's volatility inherits the factor component from the underlying stock volatilities. That is, the MVF model enjoys internal model consistency.

We find that PCA on log(RV) also suggests a single factor model structure. The modeling of log variance will lead to a log-linear single factor model:

$$\log(V_{it}) = a_i' + b_i'\xi_t + \varepsilon_{it}'. \tag{3.7}$$

Our MVF model (3.6) is closely related to (3.7). However, there are subtle but important differences between them.

Comparing the MVF model to the single factor model for log variance, we see that model (3.7) is more difficult to interpret and does not enjoy internal model consistency. Note that model (3.7) is equivalent to

$$V_{it} = \xi_t^{b_i'} \; e^{a_i' + \varepsilon_{it}'},$$

where the coefficient b_i' becomes the exponent, which makes the interpretation difficult. One natural choice of the factor ξ_t in (3.7) is the common log variance (ClogV), namely, the cross-sectional average of the log variances. We estimate the coefficients b' in the log-linear model (3.7) by regressing $\log(RV)$ over $ClogRV$, the cross-sectional average of the log RVs. We find that the coefficient b' varies around 1, with an interquartile range of 0.89∼1.08. In particular, b' can deviate from 1. As a result, model (3.7) does not enjoy internal model consistency.

In addition, the MVF model has the same format as the linear model for log variance with the slope term constrained to be one. Note that if all $b'_i = 1$, then model (3.7) becomes our MVF model (3.6). The difference between the two is the choice of factor. Under the linear model for log variance, a natural choice of the factor is ClogV. We find that CRV and the exponential of ClogRV are almost identical with a correlation higher than 0.99. The volatility prediction performance of the MVF model and the constrained log-linear model is also very similar. We conclude that the MVF model is almost equivalent to the log-linear single factor model with the slope term constrained to be one.

As a result of the above comparison, we recommend our MVF model for variance over the linear model for log variance.

Remark 2. *Barigozzi and Hallin (2020) discuss a factor model for log variance and the estimation consistency of the factor structure. However, we find that modeling variance rather than log variance has several advantages as discussed above. In addition, they assume no heteroskadasticity in returns. In contrast, we model the dynamic volatilities and naturally allow heterokasticity in returns. Moreover, our model does not rely on factor structure in returns.*

Remark 3. *The MVF model (3.6) can be easily modified to include multiple factors. When multiple factors exist, the generalized MVF takes the following form:*

$$V_{it} = \Big(\sum_{k=1}^{K} \xi_{kt} \exp(\mu_{ki}) \Big) \exp(\sigma_i z_{it}), \quad 1 \le i \le N, \tag{3.8}$$

where $(\xi_{kt})_{k=1}^{K}$ are K factors, and $\big(\exp(\mu_{ki} + \sigma_i z_{it})\big)_{k=1}^{K}$ are the multiplicative idiosyncratic exposures. Model (3.8) can be analyzed, estimated, and used in volatility forecasting in a similar way to the single factor model.

4 Volatility Forecasting

In this section, we utilize the proposed MVF model (3.6) for volatility forecasting.

4.1 Forecasting Models

4.1.1 Our Approach: MVF Model

We use CV_t as the proxy of the latent factor in our proposed MVF model (3.6).[5] The idiosyncratic variance exposure is $\widetilde{V}_{it} = V_{it}/CV_t$. We estimate CV_t by the common realized variance CRV_t and compute the idiosyncratic realized variance exposures (iRE), $\widetilde{RV}_{it} := RV_{it}/CRV_t$ for $i = 1, ..., N$. Then, we model the $\log(CRV_t)$ and $\log(\widetilde{RV}_{it})$ separately using the HAR model[6] (Corsi (2009)):

$$x_{t+1} = \theta_0 + \theta_d x_t + \theta_w x_{t-5,t} + \theta_m x_{t-22,t} + u_t, \tag{4.1}$$

where x_t represents $\log(CRV_t)$ or $\log(\widetilde{RV}_{it})$, $x_{t-5,t}$ and $x_{t-22,t}$ are the previous one week and one month averages of x_t, respectively, $u_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2)$, and θ_0, θ_d, θ_w, θ_m and σ_u are constants. The parameters in (4.1) and the models in benchmark approaches presented in the next subsection are estimated with a 252-day rolling window. We denote the forecasts of CV and $\widetilde{V}_i$ on day $t+1$ as $\widehat{CV}_{t+1}$ and $\widehat{\widetilde{V}}_{it+1}$ for $1 \leq i \leq N$. Then, the forecast of the volatility of stock i on day $t+1$ is

$$\widehat{V}_{it+1} = \widehat{CV}_{t+1} \times \widehat{\widetilde{V}}_{it+1}, \quad i = 1, \ldots, N.$$

We denote the predictions by our proposed MVF model as $\text{CV}_{\text{logHAR}} \times \text{iE}_{\text{logHAR}}$.

4.1.2 Benchmark Models

We compare our proposed MVF model with the following benchmark models.

BM1: Individual Volatility Modeling

This approach predicts each stock volatility using a logHAR model.[7] Specifically,

[5] We also evaluate the performance of the MVF model by replacing CV with the market volatility as the factor proxy. The forecasting results are much worse than the model with CV factor.

[6] Fitting $\log(\widetilde{RV}_{it})$ with a HAR model is equivalent to first estimating $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ from model (3.6) with sample mean and sample standard deviation of $\log(\widetilde{RV}_{it})$, then fitting $\widehat{z}_{it} = (\log(\widetilde{RV}_{it}) - \widehat{\mu}_i)/\widehat{\sigma}_i$ with a HAR model.

[7] We also evaluate the forecasting performance of the standard HAR model by fitting a HAR model directly on variance. When comparing the logHAR model with the standard HAR model, we

for each stock i, we fit $\log(RV_{it})$ with a HAR model (4.1). The fitted model is then used for prediction. This approach does not incorporate cross-sectional information, neither the factor structure in returns nor in volatilities. We denote the prediction of this method as $\text{Idv}_{\text{logHAR}}$.

BM2: Return Factor model + Individual Idiosyncratic Volatility modeling

This approach predicts systematic and idiosyncratic components of the volatility separately using return factor models. On each day t, we estimate $\boldsymbol{\beta}$ and $(\boldsymbol{U}_{t[j]})_{1\leq j\leq n}$ using a 252-day rolling window under the Fama-French three-factor model or a statistical factor model.[8]

As to volatility forecasting, similar to BM1, the idiosyncratic variance (iV) is predicted with a logHAR model. To forecast the factor covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{F}_{t+1}}$, BM2 uses the realized GARCH-DCC (rG) model. Specifically, we use the realized-GARCH (1,1) model with a log-linear specification (Eqn. (1), (4) and (5) in Hansen, Huang, and Shek (2012)) to forecast the factor variances, and use the DCC model (Engle (2002)) to forecast the correlation matrix of the factor returns. We denote the resulting forecast of the factor covariance matrix as $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{t+1}}$. Then, the forecast of the stock variance is

$$\widehat{V}_{it+1} = \widehat{\boldsymbol{\beta}}_{i(t)}^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{t+1}} \widehat{\boldsymbol{\beta}}_{i(t)} + \widehat{iV}_{it+1}, \quad 1 \leq i \leq N.$$

This approach utilizes the cross-sectional structure in returns but does not incorporate the factor structure in idiosyncratic variances. We denote the prediction of the method under the Fama-French three-factor model as $\text{FF3}_{\text{rG}}+\text{iV}_{\text{logHAR}}$ and the prediction under the statistical factor model as $\text{StatsF}_{\text{rG}}+\text{iV}_{\text{logHAR}}$.

BM3: Return Factor Model + Common Idiosyncratic Volatility Modeling

find that the standard HAR model performs worse.

[8]Specifically, for the statistical factor model, we use five PCs in stock returns as return factors, estimated with a 252-day rolling window based on the high-frequency 5-min returns.

This approach utilizes the return factor model as well as the factor structure in idiosyncratic variances. Same as the BM2 approach, the systematic/idiosyncratic components of the volatility are predicted separately and the systematic component is predicted using the realized GARCH+DCC model. About the forecasting of the idiosyncratic variance, BM3 utilizes the following single factor structure in idiosyncratic variance:

$$iV_{it} = c_{0i} + c_{1i} CiV_t + \varepsilon_{it}, \quad 1 \leq i \leq N.$$

On each day t, c_{0i} and c_{1i} are estimated using a 252-day rolling window regression of the idiosyncratic realized variance iRV_i over the common idiosyncratic realized variance $CiRV$. We employ a logHAR model to predict the CiV factor. The residual of the factor model for idiosyncratic variance, ε, is modeled by an AR(1), $\varepsilon_{i,t} = \xi_i \varepsilon_{i,t-1} + u_{it}$.[9] We predict the idiosyncratic variance of stock i on day $t+1$ with $\widehat{iV}_{it+1} = \widehat{c}_{0i(t)} + \widehat{c}_{1i(t)} \widehat{CiV}_{t+1} + \widehat{\varepsilon}_{it+1}$. The forecast of the stock volatility is

$$\widehat{V}_{it+1} = \widehat{\boldsymbol{\beta}}_{i(t)}^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{t+1}} \widehat{\boldsymbol{\beta}}_{i(t)} + \widehat{c}_{0i(t)} + \widehat{c}_{1i(t)} \widehat{CiV}_{t+1} + \widehat{\varepsilon}_{it+1}, \quad i = 1, ..., N.$$

We denote the prediction of BM3 under FF3 model as $\mathrm{FF3_{rG}}$+$\mathrm{iV_{CiV}}$ and the prediction of BM3 under the statistical factor model as $\mathrm{StatsF_{rG}}$+$\mathrm{iV_{CiV}}$.

4.2 Evaluation Metrics

We evaluate the performance of different models in forecasting daily continuous variance. We use the same S&P 500 Index consitituent stocks data described in Section 2.1. The evaluation period is from January 2004 to December 2020. The performance of different approaches is evaluated by Q-like (Patton (2011))[10]:

$$QLIKE_i = \frac{1}{L} \sum_{t=1}^{L} \left(\log \left(\frac{\widehat{V}_{it}}{RV_{it}} \right) + \frac{RV_{it}}{\widehat{V}_{it}} - 1 \right) \quad i = 1, \ldots, N,$$

[9] We also check the method that predicts ε with 0. The performance is worse than the AR(1) model.

[10] Besides Q-like, we also evaluate the performance using out-of-sample R^2. Our approach performs consistently well compared to the benchmark models in most of the years under evaluation.

where $\widehat{V}_{it}$ is stock i's forecast variance on day t, RV_{it} is the truncated realized variance, and L is the total length of the forecasting period. The Q-like measure is robust to the presence of noise in the volatility proxy; see Patton (2011). A smaller Q-like indicates a better volatility prediction. We use the same formulation of Q-like loss as Bollerslev, Patton, and Quaedvlieg (2016) so that the Q-likes for different stocks are standardized by the stocks' volatilities.

4.3 Forecasting Results

In Table 2, we summarize the Q-likes of different models in forecasting the volatilities of the S&P Index constituent stocks. The results show that our MVF model, $\mathrm{CV}_{\mathrm{logHAR}} \times \mathrm{iE}_{\mathrm{logHAR}}$, outperforms the benchmark models in that its Q-likes have the lowest mean, median 25% and 75% quantiles.

Table 2: Summary statistics of the Q-likes of various forecasting models in predicting S&P 500 Index constituent stocks' daily volatilities from January 2004 to December 2020. The total number of stocks under evaluation is 291, and the length of the evaluation period is $L = 4239$. The reported values are the 25% quantile (Q1), median, mean and the 75% quantile (Q3) of Q-likes across the stocks under evaluation.

Forecasting models	Q1	Median	Mean	Q3
BM1: Individual volatility modeling				
$\mathrm{Idv}_{\mathrm{logHAR}}$	0.126	0.135	0.140	0.149
BM2: Return factor model + Individual idiosyncratic volatility modeling				
$\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$	0.127	0.136	0.142	0.150
$\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$	0.129	0.137	0.144	0.150
BM3: Return factor model + Common idiosyncratic volatility modeling				
$\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$	0.134	0.143	0.159	0.165
$\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$	0.136	0.145	0.161	0.166
MVF model				
$\mathrm{CV}_{\mathrm{logHAR}} \times \mathrm{iE}_{\mathrm{logHAR}}$	**0.120**	**0.129**	**0.135**	**0.142**

We further compare the forecasting performance stock by stock and test the significance in Q-like differences between our MVF model and the benchmark models. Specifically, we compare the Q-like of the MVF model with the benchmark models for each stock, and compute the percentage of stocks for which our model generates lower Q-likes:

$$\text{Out.perf.} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{QLIKE_{MVF,i} < QLIKE_{bm,i}\}}, \tag{4.2}$$

where $N = 291$ is the total number of stocks under evaluation, and $QLIKE_{MVF,i}$ and $QLIKE_{bm,i}$ are the Q-likes of our MVF model and the benchmark model for the ith stock, respectively. Furthermore, we perform the Diebold-Mariano (DM) test (Diebold and Marino (1995)) to examine the significance of the Q-like differences between our MVF model and the benchmark models. Specifically, we perform the following one-sided Q-like difference test:

$$H_0 : E(e_{1,t} - e_{2,t}) \geq 0 \quad \text{vs.} \quad H_1 : E(e_{1,t} - e_{2,t}) < 0,$$

where $e_{m,t} = \log(\widehat{V}_{m,t}/RV_t) + RV_t/\widehat{V}_{m,t} - 1$ is the Q-like loss, $m = 1, 2$, which represent the Q-like loss of our MVF model and the benchmark model, respectively. We write $d_t = e_{1,t} - e_{2,t}$. The DM test statistic is $\bar{d}/\widehat{\sigma}(\bar{d})$, where $\bar{d} = \sum_{t=1}^{L} d_t/L$ and $\widehat{\sigma}(\bar{d})$ is the standard error of $\bar{d}$ estimated by heteroskedasticity-autocorrelation-consistent (HAC) estimator. We then compute the proportion of stocks where our MVF model generates statistically significantly lower Q-likes than the benchmark models:

$$\text{Sig.Out.perf.} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{p_i < \alpha\}}, \tag{4.3}$$

where p_i is the p-value of the DM test for the ith stock, and $\alpha = 5\%$ is the significance level.

The outperformance proportion (Out.perf.) of our proposed MFV model over the benchmark models and the significant outperformance proportion (Sig.Out.perf.) are reported in Table 3. The results show that the MVF model, $\text{CV}_{\text{logHAR}} \times \text{iE}_{\text{logHAR}}$, yields a more accurate prediction than the benchmark models for almost all the stocks. Moreover, the DM test results show that the outperformance of our MVF model is statistically significant for a very high percentage of stocks under evaluation.

Table 3: Outperformance proportion of the proposed MVF model over the benchmark models among the S&P Index constituent stocks in terms of the Q-like measure during January 2004–December 2020. The values are the percentages of stocks for which the MVF model outperforms the benchmark models.

MVF Model: $\mathrm{CV}_{\mathrm{logHAR}} \times \mathrm{iE}_{\mathrm{logHAR}}$	Outperformance proportion	
vs. Benchmark models	Out.perf. (%)	Sig.Out.perf. (%)
BM1: $\mathrm{Idv}_{\mathrm{logHAR}}$	99.3	93.5
BM2: $\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$	96.9	86.6
BM2: $\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$	97.9	93.1
BM3: $\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$	99.3	94.5
BM3: $\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$	99.0	96.6

The prediction results have several implications. First, the outperformance of our proposed MVF model compared with the model that only uses the individual stock's information, BM1:$\mathrm{Idv}_{\mathrm{logHAR}}$, demonstrates the benefit of utilizing cross-sectional information in individual stock volatility prediction. Second, our MVF model has a dominant outperformance compared with the model that uses only return factor models, BM2: $\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$ and BM2: $\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{logHAR}}$, showing the importance of incorporating the stock/idiosyncratic variance factor structure. Third, our MVF model not only simplifies the forecasting but also generates more accurate forecasting compared with the more complex models, the four-factor model (BM3: $\mathrm{FF3}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$ and $\mathrm{StatsF}_{\mathrm{rG}}+\mathrm{iV}_{\mathrm{CiV}}$). These comparisons demonstrate the solid advantages of the MVF model in volatility forecasting.

5 Global Evidence

The MVF model is built upon empirical evidence from the S&P 500 Index constituent stocks. We next examine whether the MVF model also applies to the global market.

We perform a parallel analysis of the factor structure in the global market using

Table 4: Summary statistics of the Q-likes of the forecasting models in predicting global equity indices' daily volatilities from January 2001 to March 2021. The reported values are the 25% quantile (Q1), median, mean and the 75% quantile (Q3) of Q-likes across the indices under evaluation.

Forecasting models	Q1	Median	Mean	Q3
BM1: $\text{Idv}_{\text{logHAR}}$	0.152	0.185	0.189	0.222
MVF: $\text{CV}_{\text{logHAR}} \times \text{iE}_{\text{logHAR}}$	**0.146**	**0.183**	**0.183**	**0.209**

daily realized variances of 31 global equity indices[11] from January 1, 2001 to March 12, 2021 obtained from the Oxford-Man Institute's "realized library". In the global equity indices' volatilities, we also find a strong co-movement feature and high pairwise correlation with a mean pairwise correlation of 0.5. When performing PCA on the indices' RVs, we find that a high proportion (67%) of the total variation in global equity indices' RVs can be explained by the first PC, while the second and other PCs do not account for a proportion substantially higher than the remaining. The number of factors estimators (Bai and Ng (2002), Ahn and Horenstein (2013)) also suggest a single factor in the global indices' variances. We find that CRV (the cross-sectional average of all indices' RVs) has a 0.988 correlation with the first PC and can still be a good proxy for the variance factor in the global market. In addition, similar to the US market, strong evidence suggests that the multiplicative factor structure still holds for the volatilities in the global market.[12]

We next evaluate the performance of our MVF model in forecasting the volatility of the global indices. We forecast the volatility based on our MVF model, $\text{CV}_{\text{logHAR}} \times \text{iE}_{\text{logHAR}}$, and compare the results with the benchmark model BM1:$\text{Idv}_{\text{logHAR}}$. We do not include BM2 and BM3 because there is no well-established factor model for the global indices. In Table 4, we summarize the Q-likes of these two models in forecasting the volatilities of the global equity indices. Table 4 shows that our MVF model outperforms BM1:$\text{Idv}_{\text{logHAR}}$ in that its Q-likes have the lowest mean, median, 25% and 75% quantiles. We also find that our MVF model outperforms 25 out of 31, or 80.6% of all the indices we study. The results confirm the advantage of our MVF

[11]The data from different indices are synchronized by treating GMT 00:00 – GMT 23:59 as the same day. There is no market opening overnight.

[12]The detailed results are available upon request.

model in global market volatility forecasting.

6 Conclusion

This work provides a framework to study the factor structure in stock variance based on high-frequency and high-dimensional price data. We theoretically show that the factor structure in stock and idiosyncratic variance can be consistently estimated by conducting PCA on the stock/idiosyncratic realized variances. Empirically, based on the strong empirical evidence from the analysis of daily volatilities of S&P 500 Index constituent stocks, we propose a multiplicative volatility factor (MVF) model. The MVF model includes a multiplicative variance factor and a multiplicative idiosyncratic component, where the variance factor is approximately the cross-sectional average of stock variances. Based on the proposed MVF model, we develop a forecasting model, which is found to provide more accurate volatility forecasts than various benchmark approaches for a majority of the stocks we evaluate. Finally, we demonstrate that our MVF model also applies to the global market and helps to predict the volatilities of global equity indices.

The volatility factor modeling framework that we propose facilitates a deeper understanding of the financial market. The MVF model achieves dimension reduction in volatility modeling for a large cross-section of assets. Besides volatility forecasting, our framework provides insights into the study of shock spillover and transmission in financial systems and holds promise in applications such as large portfolio allocation, risk management and volatility trading.

References

AHN, SEUNG C. AND ALEX R. HORENSTEIN (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81, 1203–1227.

AÏT-SAHALIA, YACINE, JIANQING FAN, AND YINGYING LI (2013): "The leverage effect puzzle: Disentangling sources of bias at high frequency," *Journal of Financial Economics*, 109, 224–249.

Aït-Sahalia, Yacine and Loriano Mancini (2008): "Out of sample forecasts of quadratic variation," *Journal of Econometrics*, 147, 17–33.

Aït-Sahalia, Yacine and Dacheng Xiu (2017): "Using principal component analysis to estimate a high dimensional factor model with high-frequency data," *Journal of Econometrics*, 201, 384–399.

Aleti, Saketh (2022): "The high-frequency factor zoo," *Working paper*.

Andersen, Torben G and Luca Benzoni (2010): "Do bonds span volatility risk in the US Treasury market? A specification test for affine term structure models," *The Journal of Finance*, 65, 603–653.

Andersen, Torben G, Tim Bollerslev, Francis X Diebold, and Paul Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Asai, Manabu and Michael McAleer (2015): "Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance," *Journal of Econometrics*, 189, 251–262.

Bai, Jushan (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

Bai, Jushan and Serena Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

Barigozzi, Matteo and Marc Hallin (2017): "Generalized dynamic factor models and volatilities: estimation and forecasting," *Journal of Econometrics*, 201, 307–321.

——— (2020): "Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals," *Journal of Econometrics*, 216, 4–34.

Barndorff-Nielsen, Ole E and Neil Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

——— (2004): "Power and bipower variation with stochastic volatility and jumps," *Journal of Financial Econometrics*, 2, 1–37.

Bollerslev, Tim (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, Tim, Benjamin Hood, John Huss, and Lasse Heje Pedersen (2018): "Risk everywhere: Modeling and managing volatility," *The Review of Financial Studies*, 31, 2729–2773.

Bollerslev, Tim, Andrew J Patton, and Rogier Quaedvlieg (2016): "Exploiting the errors: A simple approach for improved volatility forecasting," *Journal of Econometrics*, 192, 1–18.

Bollerslev, Tim and Viktor Todorov (2011): "Estimation of jump tails," *Econometrica*, 79, 1727–1783.

Calvet, Laurent E, Adlai J Fisher, and Samuel B Thompson (2006): "Volatility comovement: a multifrequency approach," *Journal of Econometrics*, 131, 179–215.

Chamberlain, Gary and Michael Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross (1986): "Economic forces and the stock market," *Journal of Business*, 383–403.

Clark, Peter K (1973): "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica: Journal of the Econometric Society*, 135–155.

Corsi, Fulvio (2009): "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Da, Zhi and Ernst Schaumburg (2006): "The factor structure of realized volatility and its implications for option pricing," .

Diebold, Francis X and Roberto S Marino (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13.

Ding, Yi, Robert Engle, Yingying Li, and Xinghua Zheng (2022): "Supplement to "Factor modeling for volatility"," .

Ding, Yi, Yingying Li, and Xinghua Zheng (2021): "High dimensional minimum variance portfolio estimation under statistical factor models," *Journal of Econometrics*, 222, 502–515.

Engle, Robert (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingrom inflation," *Econometrica*, 50, 391–407.

——— (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, 20, 339–350.

Engle, Robert and Susana Martin (2019): "Measuring and hedging geopolitical risks," .

Engle, Robert F, Takatoshi Ito, and Wen-Ling Lin (1990): "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market," *Econometrica*, 58, 525–542.

Fama, Eugene F and Kenneth R French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

——— (2015): "A five-factor asset pricing model," *Journal of financial economics*, 116, 1–22.

——— (2018): "Choosing factors," *Journal of financial economics*, 128, 234–252.

Fan, Jianqing, Yuan Liao, and Martina Mincheva (2013): "Large covariance estimation by thresholding principal orthogonal complements," *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75, 603–680, with 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva.

Gençay, Ramazan, Michel Dacorogna, Ulrich A Muller, Olivier Pictet, and Richard Olsen (2001): *An introduction to high-frequency finance*, Elsevier.

Gonçalves, Sílvia and Nour Meddahi (2009): "Bootstrapping realized volatility," *Econometrica*, 77, 283–306.

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek (2012): "Realized garch: a joint model for returns and realized measures of volatility," *Journal of Applied Econometrics*, 27, 877–906.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh (2016): "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119, 249–283.

Hounyo, Ulrich, Sílvia Gonçalves, and Nour Meddahi (2017): "Bootstrapping pre-averaged realized volatility under market microstructure noise," *Econometric Theory*, 33, 791–838.

Hull, John and Alan White (1987): "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, 42, 281–300.

Jacod, Jean, Yingying Li, Per A Mykland, Mark Podolskij, and Mathias Vetter (2009): "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic processes and their applications*, 119, 2249–2276.

Jacod, Jean, Yingying Li, and Xinghua Zheng (2019): "Estimating the integrated volatility with tick observations," *Journal of Econometrics*, 208, 80–100.

Jacod, Jean and Philip Protter (1998): "Asymptotic error distributions for the Euler method for stochastic differential equations," *The Annals of Probability*, 26, 267–307.

Kapadia, Nishad, Matthew Linn, and Bradley S Paye (2020): "One Vol to Rule Them All: Common Volatility Dynamics in Factor Returns," *Available at SSRN 3606637*.

Kelly, Bryan, Hanno Lustig, and Stijn Van Nieuwerburgh (2013): "Firm volatility in granular networks," Tech. rep., National Bureau of Economic Research.

Li, Jia, Yunxiao Liu, and Dacheng Xiu (2019): "Efficient estimation of integrated volatility functionals via multiscale jackknife," *The Annals of Statistics*, 47, 156–176.

Li, Jia, Viktor Todorov, and George Tauchen (2016): "Inference theory for volatility functional dependencies," *Journal of Econometrics*, 193, 17–34.

——— (2017): "Jump regressions," *Econometrica*, 85, 173–195.

Li, Jia and Dacheng Xiu (2016): "Generalized Method of Integrated Moments for High-Frequency Data," *Econometrica*, 84, 1613–1633.

Liu, Lily Y, Andrew J Patton, and Kevin Sheppard (2015): "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes," *Journal of Econometrics*, 187, 293–311.

Luciani, Matteo and David Veredas (2015): "Estimating and forecasting large panels of volatilities with approximate dynamic factor models," *Journal of Forecasting*, 34, 163–176.

Mancini, Cecilia (2009): "Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps," *Scandinavian Journal of Statistics*, 36, 270–296.

Nelson, Daniel B (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica: Journal of the Econometric Society*, 347–370.

Patton, Andrew J (2011): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 160, 246–256.

Ross, Stephen (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

Sharpe, William (1964): "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Susmel, Raul and Robert F Engle (1994): “Hourly volatility spillovers between international equity markets,” *Journal of International Money and Finance*, 13, 3–25.

Taylor, Stephen John (1982): “Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75,” *Time series analysis: theory and practice*, 1, 203–226.

Xiu, Dacheng (2010): “Quasi-maximum likelihood estimation of volatility with high frequency data,” *Journal of Econometrics*, 159, 235–250.

Zhang, Lan, Per A Mykland, and Yacine 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.

Zhang, Lan et al. (2006): “Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach,” *Bernoulli*, 12, 1019–1043.

Supplement to "Factor Modeling for Volatility"

Yi Ding * Robert Engle † Yingying Li ‡ Xinghua Zheng §

November 20, 2022

This supplement gives additional empirical results and proofs of the theoretical results for Ding, Engle, Li, and Zheng (2022).

A Analysis under Additive Linear CV Factor Model

A.1 Analysis of Idiosyncratic Component

We first investigate the idiosyncratic component ε in (3.3). To do so, we pick a random stock, estimate the coefficients and obtain the residual $\widehat{\varepsilon}$, and then check the time series plots of $\widehat{\varepsilon}$ and $\widehat{\varepsilon}/CRV$ in Figure 1. It suggests that the residual $\widehat{\varepsilon}$ is clearly heteroskedastic. Moreover, $\widehat{\varepsilon}$ seems to scale with CRV. If we scale $\widehat{\varepsilon}$ by CRV, the scaled residual $\widehat{\varepsilon}/CRV$ appears to be more homoskedastic.

To quantitatively evaluate the abovementioned relationship, for each stock i, we compute the correlation between $\widehat{\varepsilon}_i^2$ and CRV^2, and the correlation between $\widehat{\varepsilon}_i^2/CRV^2$ and CRV^2. The results are summarized in Table 1. From Table 1, we see a very

*Faculty of Business Administration, University of Macau, Taipa, Macau. Email: yiding@um.edu.mo

†Stern School of Business, New York University, 44 West Fourth Street Suite 9-62, New York. Email: rengle@stern.nyu.edu

‡Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Research is supported in part by RGC grants GRF 16503419, GRF 16502118 and T31-64/18-N of the HKSAR and NSFC 19BM03. Email: yyli@ust.hk

§Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Research is supported in part by RGC grants GRF-16304521, GRF 16304019 and T31-64/18-N of the HKSAR. Email: xhzheng@ust.hk

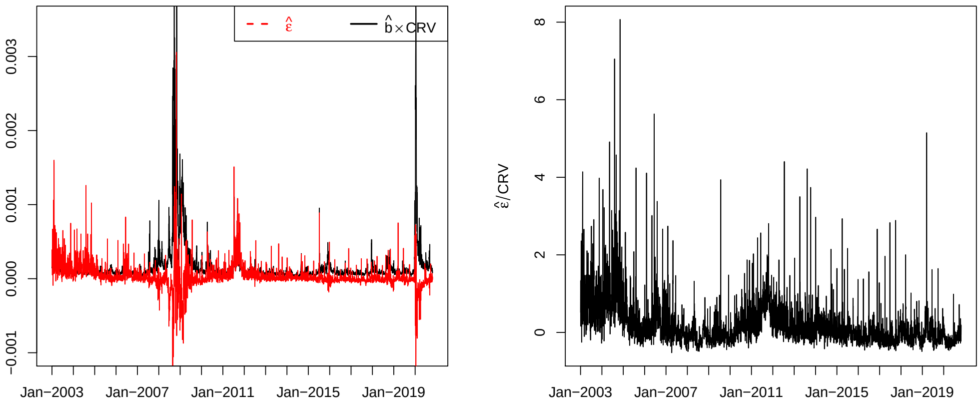

Figure 1: Left: Estimated idiosyncratic component $\widehat{\varepsilon}$ of a typical stock in model (3.3) and $\widehat{b} \times CRV$. Right: Time series plot of $\widehat{\varepsilon}/CRV$.

Table 1: Summary statistics of correlations in absolute value between $\widehat{\varepsilon}^2$ from model (3.3) and CRV^2, and between $\widehat{\varepsilon}^2/CRV^2$ and CRV^2. The values are the 25% quantile (Q1), median, mean and the 75% quantile (Q3).

	Q1	Median	Mean	Q3
$\lvert\text{corr}(\widehat{\varepsilon}^2_{i,\cdot}, CRV^2_{\cdot})\rvert_{1\le i\le N}$	0.276	0.447	0.425	0.568
$\lvert\text{corr}(\widehat{\varepsilon}^2_{i,\cdot}/CRV^2_{\cdot}, CRV^2_{\cdot})\rvert_{1\le i\le N}$	0.006	0.009	0.017	0.017

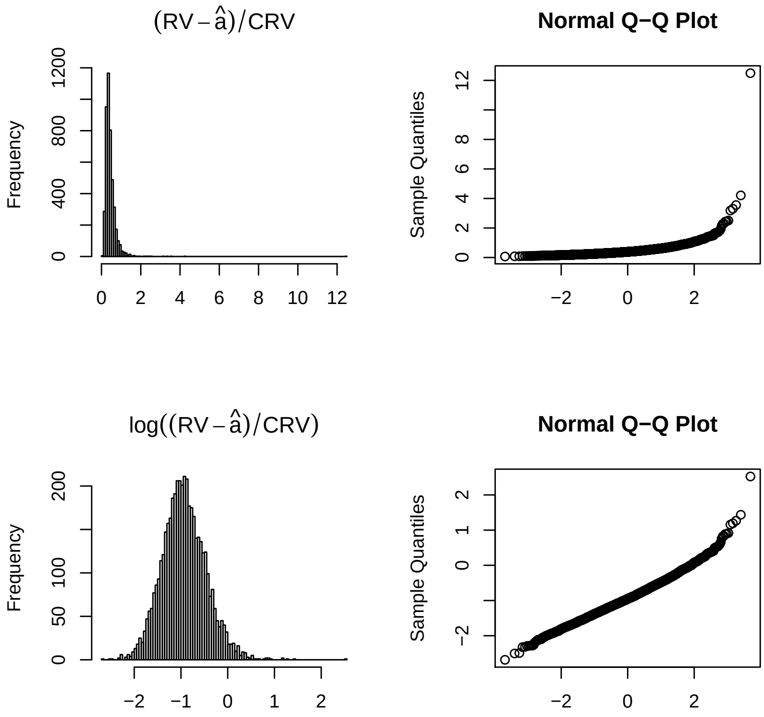

Figure 2: Distributions of $(RV - \widehat{a})/CRV$ and $\log\big((RV - \widehat{a})/CRV\big)$ of a random stock, where $\widehat{a}$ is the weighted regression estimate in model (A.1).

interesting phenomenon: the size of $\widehat{\varepsilon}$ clearly correlates with CRV, and the correlation almost disappears when we scale $\widehat{\varepsilon}$ by CRV.

In summary, the evidence suggests the following model:

$$V_{it} = a_i + b_i CV_t + CV_t\, \tilde{\varepsilon}_{it}, \quad 1 \le i \le N, \tag{A.1}$$

where $\tilde{\varepsilon}_{it}$ is independent with CV. We then estimate the model (A.1) by weighted regression using CRV as weights. We then compare the distribution of $(RV-\widehat{a})/CRV$ and $\log\big((RV-\widehat{a})/CRV\big)$ against normal distribution. The results for a random stock are presented in Figure 2. We observe in Figure 2 that $(RV - \widehat{a})/CRV$ is heavy-tailed, while $\log((RV - \widehat{a})/CRV)$ appears to be roughly normally distributed. We are therefore led to the following refined model:

$$\begin{aligned} V_{it} =& a_i + b_i CV_t + CV_t \tilde{\varepsilon}_{it} \\ =& a_i + b_i CV_t + CV_t \Big(\exp(\mu_i + \sigma_i z_{it}) - \exp(\mu_i + \sigma_i^2/2) \Big), \quad i = 1, ..., N, \end{aligned} \tag{A.2}$$

where the idiosyncratic component scales with CV, and $\tilde{\varepsilon}_{it}$ is modeled by a centered lognormal distribution.

A.2 Analysis of Model Coefficients

We estimate the model (A.2) using MLE on subsampled data. We subsample the data to circumvent possible auto-correlation. Specifically, we conduct MLE of model (A.2) on observations subsampled every ten days, and then take the average of the ten estimates.

We compute the proportion of the estimated intercept $\widehat{a}_i$ over $\overline{RV}_i$ (time series average) for each stock i and find that the proportions are all very small with an average of 0.035. The results suggest that a_i plays little role in the model (A.2). As another check, we note that (A.2) is equivalent to $V_{it}/CV_t = b_i + a_i/CV_t + \tilde{\varepsilon}_{it}$. We then regress RV_{it}/CRV_t over $1/CRV_t$ for each individual stock and find that the R^2s are all nearly zero with an average of 0.024. This evidence suggests that the intercept term a_i in (A.2) can be ignored. We therefore get the simplified model with zero intercept:

$$V_{it} = b_i CV_t + CV_t \Big(\exp(\mu_i + \sigma_i z_{it}) - \exp(\mu_i + \sigma_i^2/2) \Big).$$

Next, we check the relation between b_i and $\exp(\mu_i + \sigma_i^2/2)$. The scatterplot of the MLE of b_i and $\exp(\mu_i + \sigma_i^2/2)$ is presented in Figure 3. It shows that b and $\exp(\mu + \sigma^2/2)$ are strongly linearly related. The correlation between the two reaches 0.99. Moreover, the linear relation fits well with the line $y = x$.

By constraining $a = 0$ and $b = \exp(\mu + \sigma^2/2)$ in (A.2), we reach our final multiplicative volatility factor model (3.6).

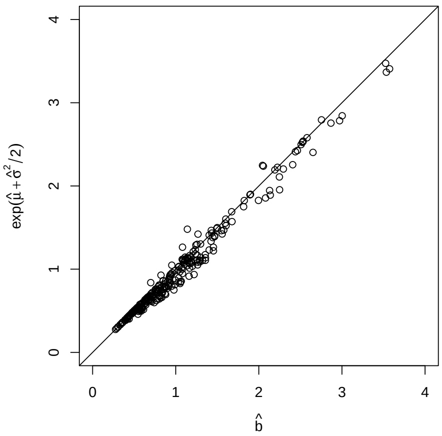

Figure 3: Scatterplot of the MLE of b_i and $\exp(\mu_i+\sigma_i^2/2)$ of all stocks under evaluation in model (A.2). The diagonal line is $y = x$.

B Proofs

In the following, $c, C, C_1, C', C_0, ...$, etc, denote constants which do not depend on T, N, Δ_n, and can vary from place to place.

The first lemma extends the concentration inequality for estimating (co)-integrated variance (Fan, Li, and Yu (2012); Cai, Hu, Li, and Zheng (2020)) to the case when only polynomial tail decay is imposed on the spot volatility.

Lemma 1. *Suppose that* (ν_{1t}) *and* (ν_{2t}) *satisfy* $d\nu_{jt} = \mu_{jt}dt + \sigma_{jt}dW_{j;t}$ *for* $j = 1, 2$, *where* $(W_{1;t})$ *and* $(W_{2;t})$ *are standard Brownian motions that can be dependent with each other, and there exist constants* C_μ, K_σ, $M > 0$, *such that* $\max_{0\le t\le 1} |\mu_t^j| \le C_\mu$, *and for any* $x > 0$ *and* $j = 1, 2$,

$$P\left(\max_{0\le t\le 1} |\sigma_{jt}| > x\right) < \frac{K_\sigma}{x^M}.$$

Suppose also that the observation times (t_i) *satisfy* $\sup_n \max_{1\le i\le n} n|t_i - t_{i-1}| \le C_\Delta$ *for some constant* $C_\Delta > 0$. *For* $j_1, j_2 \in \{1, 2\}$, *denote the realized (co)variance by* $[\nu_{j_1}, \nu_{j_2}]_t = \sum_{\{i:t_i\le t\}}(\nu_{j_1t_i} - \nu_{j_1t_{i-1}})(\nu_{j_2t_i} - \nu_{j_2t_{i-1}})$. *Then for any* $0 < \delta < 1$, *there* *exist* C_1 *and* C_2 *such that for all* $x \in [0, (2C_\Delta\sqrt{n})^{1/(1-\delta)}]$,

$$P\left(\sqrt{n}\left|[\nu_j,\nu_j]_1 - \int_0^1 (\sigma_{js})^2 ds\right| > x\right) \le C_1 x^{-\frac{M\delta}{2}}, \quad j = 1,2, \ \textit{and} \tag{B.1}$$

$$P\left(\sqrt{n}\left|[\nu_1,\nu_2]_1 - \int_0^1 \sigma_{1s}\sigma_{2s}\rho_s ds\right| > x\right) \le C_2 x^{-\frac{M\delta}{2}}, \tag{B.2}$$

where $\rho_s = \mathrm{corr}(dW_{1;s}, dW_{2;s}) := \lim_{h\to 0} \mathrm{corr}(W_{1;s+h} - W_{1;s}, W_{2;s+h} - W_{2;s}|\mathcal{F}_t)$, $\mathcal{F}_t$ *is the information available at time* t*, and the constants* C_1*,* C_2 *depend only on* K_σ*,* C_μ*,* δ*,* C_Δ *and* M*.*

Proof: Define

$$\varphi(x) = \begin{cases} 1, \text{ if } 0 \le x < 1, \quad \text{and} \\ x^{\delta/2}, \text{ if } 1 \le x \le (2C_\Delta\sqrt{n})^{1/(1-\delta)}. \end{cases} \tag{B.3}$$

Note that $0 \le x \le 2\varphi(x)^2 C_\Delta\sqrt{n}$ when $0 \le x \le (2C_\Delta\sqrt{n})^{1/(1-\delta)}$. By the proof of Lemma 1 in Cai, Hu, Li, and Zheng (2020), for any $0 \le x \le (2C_\Delta\sqrt{n})^{1/(1-\delta)}$, we have

$$\begin{aligned}
&P\left(\sqrt{n}\left|[\nu_j,\nu_j]_1 - \int_0^1 (\sigma_{js})^2 ds\right| > x\right) \\
\le &P\left(\left\{\sqrt{n}\left|[\nu_j,\nu_j]_1 - \int_0^1 (\sigma_{js})^2 ds\right| > x\right\} \bigcap \left\{\max_{0\le t\le 1} \sigma_{jt} \le \varphi(x)\right\}\right) + P\left(\max_{0\le t\le 1} \sigma_{jt} > \varphi(x)\right) \\
\le &3\exp\left(\frac{C_\mu^2}{\varphi(x)^2} - \frac{x^2}{32\varphi(x)^4 C_\Delta^2}\right) + \frac{K_\sigma}{\varphi(x)^M}.
\end{aligned}$$

We have $\varphi(x)^{-M} \le x^{-\frac{M\delta}{2}}$. Moreover, when $0 \le x \le 1$, $\exp\big(-x^2/(32\varphi(x)^4 C_\Delta^2)\big) \le 1 \le x^{-\frac{M\delta}{2}}$. When $1 < x \le (2C_\Delta\sqrt{n})^{1/(1-\delta)}$, by the fact that $\exp(-x)x^y \le \exp(-y)y^y$ for all $x, y > 0$, we have

$$\begin{aligned}
\exp\left(-\frac{x^2}{32\varphi(x)^4 C_\Delta^2}\right) &= \exp\left(-\frac{x^{2(1-\delta)}}{32C_\Delta^2}\right) \\
&\le \left(\left(\frac{8M\delta C_\Delta^2}{1-\delta}\right)^{\frac{M\delta}{4(1-\delta)}} \cdot \exp\left(-\frac{M\delta}{4(1-\delta)}\right)\right) \cdot x^{-\frac{M\delta}{2}}.
\end{aligned}$$

Hence for all $0 \leq x \leq (2C_\Delta\sqrt{n})^{1/(1-\delta)}$, we have

$$
\begin{aligned}
&3\exp\left(\frac{C_\mu^2}{\varphi(x)^2} - \frac{x^2}{32\varphi(x)^4 C_\Delta^2}\right) \\
\leq & 3\exp\left(C_\mu^2 - \frac{x^2}{32\varphi(x)^4 C_\Delta^2}\right) \\
\leq & 3\exp(C_\mu^2)\cdot\left(\left(\frac{8M\delta C_\Delta^2}{1-\delta}\right)^{\frac{M\delta}{4(1-\delta)}}\cdot\exp\Big(-\frac{M\delta}{4(1-\delta)}\Big)+1\right)\cdot x^{-\frac{M\delta}{2}}.
\end{aligned}
$$

The desired bound (B.1) follows by setting $C_1 = 3\exp(C_\mu^2)\cdot\left(\left(\frac{8M\delta C_\Delta^2}{1-\delta}\right)^{\frac{M\delta}{4(1-\delta)}}\cdot\exp\Big(-\frac{M\delta}{4(1-\delta)}\Big)+1\right) + K_\sigma$.

The bound (B.2) follows from a similar argument above by using the inequality

$$
\begin{aligned}
&P\left(\sqrt{n}\left|[\nu_1,\nu_2]_1 - \int_0^1 \sigma_{1s}\sigma_{2s}\rho_s ds\right| > x\right) \\
\leq & P\left(\left\{\sqrt{n}\left|[\nu_1,\nu_2]_1 - \int_0^1 \sigma_{1s}\sigma_{2s}\rho_s ds\right| > x\right\}\bigcap\left\{\max_{0\leq t\leq 1}(|\sigma_{1t}|,|\sigma_{2t}|)\leq\varphi(x)\right\}\right) \\
&+ P\left(\max_{0\leq t\leq 1}|\sigma_{1t}| > \varphi(x)\right) + P\left(\max_{0\leq t\leq 1}|\sigma_{2t}| > \varphi(x)\right),
\end{aligned}
$$

Lemma 2 in Cai, Hu, Li, and Zheng (2020), and setting $C_2 = 6\exp(C_\mu^2)\cdot\left(\left(\frac{32M\delta C_\Delta^2}{1-\delta}\right)^{\frac{M\delta}{4(1-\delta)}}\cdot\exp\Big(-\frac{M\delta}{4(1-\delta)}\Big)+1\right) + 2K_\sigma$. □

Lemma 2. *Suppose that* $\mathbf{X}_t = (x_{1t},...,x_{St})^T$, $1 \leq t \leq T$, $S = O(T^\gamma)$ *and* $\max_{1\leq i\leq S, 1\leq t\leq T}|E(x_{it}^M)| < c$ *for some constants* $\gamma > 0$, $M > 2+2\gamma$, *and* $c > 0$. *Assume the strong mixing condition that* $\rho(\chi) \leq c_1\exp(-c_2\chi)$ *for some constants* $c_1, c_2 > 0$ *and any positive integer* χ, *where* $\rho(\chi) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_\chi^\infty}|P(AB) - P(A)P(B)|$, $\mathcal{F}_{-\infty}^0$, $\mathcal{F}_\chi^\infty$ *are* σ*-algebras generated by* $\{\mathbf{X}_t : -\infty \leq t \leq 0\}$ *and* $\{\mathbf{X}_t : \chi \leq t \leq \infty\}$, *respectively. Then for some constant* $C > 0$, *(i) if S is fixed,*

$$P\left(\max_{1\le i\le S}\left(\Big|\frac{1}{T}\sum_{t=1}^{T}x_{it}-E(x_{it})\Big|\right)>C\sqrt{\frac{\log T}{T}}\right)\le C\left(\frac{(\log T)^{M/2}}{T^{M/2-1-\gamma}}+\frac{1}{T}\right);$$

(ii) if $S\to\infty$,

$$P\left(\max_{1\le i\le S}\left(\Big|\frac{1}{T}\sum_{t=1}^{T}x_{it}-E(x_{it}\Big|\right)>C\sqrt{\frac{\log S}{T}}\right)\le C\left(\frac{(\log S)^{M/2}}{T^{M/2-1-\gamma}}+\frac{1}{T}+\frac{1}{S}\right).$$

Proof: We only show the case when $S\to\infty$. For the case when S is fixed, the results can be shown similarly by using truncation level $\sqrt{T/(\log T)}$ instead of $\sqrt{T/(\log S)}$.

We denote $x_{it}^{tr}=x_{it}\mathbf{1}_{\{|x_{it}|\le C_0\sqrt{T/(\log S)}\}}$. By Markov's inequality, we have

$$P\Big(|x_{it}|>C_0\sqrt{T/(\log S)}\Big)\le\frac{C(\log S)^{M/2}}{T^{M/2}}.$$

By Bonferroni's inequality and that $E(x_{it}^M)<c$ and $M>2+2\gamma$, we have

$$\begin{aligned}&P\Big(x_{it}=x_{it}^{tr}\text{ for all }1\le t\le T,1\le i\le S\Big)\\ =&1-P\Big(\max_{1\le t\le T,1\le i\le S}|x_{it}|>C_0\sqrt{T/(\log S)}\Big)\\ \ge&1-\frac{C(\log S)^{M/2}ST}{T^{M/2}}\ge1-\frac{C(\log S)^{M/2}}{T^{M/2-1-\gamma}}.\end{aligned}\tag{B.4}$$

By $E(x_{it}^M)<c$ and the Cauchy-Schwarz inequality, we have that, for any $1\le M_0<M$,

$$\begin{aligned}&\max_{1\le i\le S}E\Big(|x_{it}^{tr}-x_{it}|^{M_0}\Big)\\ =&\max_{1\le i\le S}E\Big(|x_{it}|^{M_0}\cdot\mathbf{1}_{\{x_{it}\ge C_0\sqrt{T/(\log S)}\}}\Big)\\ \le&\max_{1\le i\le S}\Big(E(|x_{it}|^M)\Big)^{M_0/M}\cdot\max_{1\le i\le S}\left(P\Big(x_{it}\ge C_0\sqrt{T/(\log S)}\Big)\right)^{1-M_0/M}\\ \le&\frac{C(\log S)^{(M-M_0)/2}}{T^{(M-M_0)/2}}.\end{aligned}\tag{B.5}$$

Because $M > 2$, (B.5) implies that

$$\max_{1\le i\le S}\left|E(x_{it}^{tr}) - E(x_{it})\right| = o(\sqrt{1/T}). \tag{B.6}$$

By the fact that $(a+b)^g \le 2^g(a^g+b^g)$ for all $a, b > 0$, and $g \ge 1$, for some $2 < M_0 < M$, and $C > 0$,

$$\max_{1\le i\le S} E\left(|x_{it}^{tr}|^{M_0}\right) \le \max_{1\le i\le S} 2^{M_0}\left(E\left(|x_{it}^{tr} - x_{it}|^{M_0}\right) + E\left(|x_{it}|^{M_0}\right)\right) < C. \tag{B.7}$$

By (B.6), (B.7) and the triangle inequality, applying Bernstein's inequality (Theorem 2 Eqn. (2.3) of Merlevède, Peligrad, Rio et al. (2009)) to $x_{it}^{tr} - E(x_{it}^{tr})$ yields

$$P\left(\max_{1\le i\le S}\left|\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}^{tr} - E(x_{it})\right)\right| > C\sqrt{\frac{\log S}{T}}\right) \le C\left(\frac{1}{S} + \frac{1}{T}\right). \tag{B.8}$$

The desired bound follows from (B.4) and (B.8). □

Lemma 3. *Under the assumptions of Theorem 2, for some constant $C_0 > 0$, the $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}_n}$ defined in (2.7) satisfy*

$$P\left(\max_{1\le i\le N}\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|_2 > C_0\sqrt{\Delta_n}\right) \le C_0\left(\frac{1}{N} + \frac{1}{T}\right), \text{ and} \tag{B.9}$$

$$P\left(\|\widehat{\boldsymbol{\alpha}_n} - \overline{\boldsymbol{\alpha}_n}\|_{\max} > C_0\Delta_n\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}}\right)\right) \le C_0\left(\frac{1}{N} + \frac{1}{T}\right). \tag{B.10}$$

Proof: First, we denote

$$\overline{\boldsymbol{U}} =: (\overline{U}_1, ..., \overline{U}_N)^T = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{U}_{t[j]}, \quad \text{and}$$

$$\overline{\boldsymbol{F}} =: (\overline{F}_1, ..., \overline{F}_K)^T = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{F}_{t[j]}.$$

By definition (2.7), we have,

$$\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = \Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})^T\Big)^{-1}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(U_{i,t[j]} - \overline{U}_i)\Big)$$
$$+\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})^T\Big)^{-1}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(\alpha_{n;t[j]i} - \overline{\alpha_{ni}})\Big),$$

where $\alpha_{n;t[j]i}$ and $\overline{\alpha_{ni}}$ are the ith element of $\boldsymbol{\alpha}_{n;t[j]}$ defined in (2.3) and average drift $\overline{\boldsymbol{\alpha}_n} = \frac{1}{T\cdot[1/\Delta_n]}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{\alpha}_{n;t[j]}$, respectively.

We define an event A as follows. For some $c, C > 0$,

$$A = \Big\{\max_{1\le k\le K, 1\le i\le N}\Big|\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(F_{k,t[j]} - \overline{F}_k)(U_{i,t[j]} - \overline{U}_i)\Big| < C\sqrt{\Delta_n T N^{1/\gamma}}\Big\}$$
$$\bigcap\Big\{\lambda_{\min}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})^T\Big) \ge \frac{1}{2}\lambda_{\min}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\Big\}$$
$$\bigcap\Big\{\lambda_{\max}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]} - \overline{\boldsymbol{F}})^T\Big) \le 2\lambda_{\max}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\Big\}$$
$$\bigcap\Big\{cT < \lambda_{\min}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big) < \lambda_{\max}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big) < CT\Big\}.$$

By Assumption 1 that $\sup_{s\ge 0}\|\boldsymbol{\alpha}_s\|_{\max} = O(1)$, we have that

$$\max_{1\le i\le N, 1\le t\le T, 1\le j\le[1/\Delta_n]}|\alpha_{n;t[j]i} - \overline{\alpha_{ni}}| \le C\Delta_n.$$

Under the event A, we have

$$\max_{1\le i\le N}\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|_2^2 \le \frac{8}{c^2T^2}\sum_{k=1}^{K}\max_{1\le i\le N}\left(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(F_{k,t[j]} - \overline{F}_k)(U_{i,t[j]} - \overline{U}_i)\right)^2$$
$$+ \frac{32C^2T^2}{c^2T^2}\cdot TK[1/\Delta_n]\cdot\max_{1\le i\le N, 1\le t\le T, 1\le j\le[1/\Delta_n]}(\alpha_{n;t[j]i} - \overline{\alpha_{ni}})^2$$
$$\le \frac{8KC^2\Delta_n T N^{1/\gamma}}{c^2T^2} + \frac{32C^2T^2KC^2\Delta_n}{c^2T^2}.$$

By the assumption that $N = O(T^\gamma)$, we have $\max_{1\le i\le N} \|\widehat{\boldsymbol\beta}_i - \boldsymbol\beta_i\|_2 = O(\sqrt{\Delta_n})$.

It remains to show that $P(A) \ge 1 - O(1/N + 1/T)$. First, under Assumption 4 and that $M > 4$, by Jensen's inequality, we have, for any $t \le T-1$ and $1 \le j,k \le N$,

$$E\bigg(\Big|\int_t^{t+1} \boldsymbol\Phi_{s,jk} ds\Big|^M\bigg) \le E\bigg(\Big(\int_t^{t+1} \big|\boldsymbol\Phi_{s,jk}\big| ds\Big)^M\bigg) \le E\bigg(\int_t^{t+1} \big|\boldsymbol\Phi_{s,jk}\big|^M ds\bigg) \le C. \tag{B.11}$$

Under Assumption 1 that $0 < c_1 < \lambda_{\min}\Big(E(\int_0^1 \boldsymbol\Phi_s d_s)\Big) \le \lambda_{\max}\Big(E(\int_0^1 \boldsymbol\Phi_s d_s)\Big) < C_1$, by Lemma 2(i) and Weyl's Theorem, we have, for all large T

$$P\bigg(\frac{c_1}{2} < \lambda_{\min}\Big(\frac{1}{T}\int_0^T \boldsymbol\Phi_s d_s\Big) < \lambda_{\max}\Big(\frac{1}{T}\int_0^T \boldsymbol\Phi_s d_s\Big) < 2C_1\bigg) \ge 1 - \frac{CK^2}{T}. \tag{B.12}$$

Applying Lemma 1 to $\boldsymbol X_{tT}/\sqrt{T}$ with $x = \sqrt{T}$ and $\delta = 1/2$, and by Bonferroni's inequality, under Assumption 4, we have, for some constants $C_1, C_2 > 0$,

$$P\bigg(\Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} \boldsymbol F_{t[j]}\boldsymbol F_{t[j]}^T - \frac{1}{T}\int_0^T \boldsymbol\Phi_s ds\Big\|_{\max} > C_1\sqrt{\Delta_n}\bigg) < \frac{C_2K^2}{T^{M/2}} \le \frac{C_2K^2}{T^2}. \tag{B.13}$$

where the last inequality holds because $M > 4$.

By Assumption 1 that $\sup_{s\ge 0} \|\boldsymbol h_s\|_{\max} = O(1)$, we have

$$\max_{1\le k\le K}\bigg|\Big(\frac{1}{T}\int_0^T \boldsymbol h_s ds\Big)_k\bigg| \le C.$$

By Assumption 4 and the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned}&\max_{1\le k\le K} E\bigg(\Big(\int_0^T \boldsymbol\eta_s d\boldsymbol W_s/\sqrt{T}\Big)_k^{2M}\bigg)\\ &\le \max_{1\le k\le K} E\bigg(\Big(\frac{1}{T}\int_0^T \boldsymbol\Phi_{s,kk} ds\Big)^M\bigg) \le \max_{1\le k\le K} E\bigg(\Big(\frac{1}{T}\int_0^T \boldsymbol\Phi_{s,kk}^M ds\Big)\bigg) \le C,\end{aligned}$$

where the second inequality holds by Jensen's inequality and that $M > 4$. By

Markov's inequality, we have, for large T,

$$P\left(\left\|\frac{1}{T}\int_0^T \boldsymbol{\eta}_s d\boldsymbol{W}_s\right\|_{\max} > C\right) \leq \frac{CK}{T^2}. \tag{B.14}$$

Note that $\overline{\boldsymbol{F}} = (\int_0^T \boldsymbol{h}_s ds + \int_0^T \boldsymbol{\eta}_s d\boldsymbol{W}_s)/(T[1/\Delta_n])$. For large T, we have

$$P\Big(\max_{1\leq k\leq K}(|\overline{F}_k|) > C\Delta_n\Big) \leq \frac{CK}{T^2}. \tag{B.15}$$

Therefore, by the inequality that $\|A\|_2 \leq \mathrm{tr}(A)$ for any nonnegative definite matrix A, we have

$$P\Big(T\cdot[1/\Delta_n]\cdot\|\overline{\boldsymbol{F}}\,\overline{\boldsymbol{F}}^T\|_2 > CKT\Delta_n\Big) \leq \frac{CK^2}{T^2}. \tag{B.16}$$

Note that $\Delta_n = o(1)$. By Weyl's Theorem, (B.12), (B.13) and (B.16), we get

$$\begin{aligned}
&P\left(\lambda_{\min}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})^T\Big) < \frac{1}{2}\lambda_{\min}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\right)\\
&= P\left(\lambda_{\min}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{F}_{t[j]}\,\boldsymbol{F}_{t[j]}^T - T\cdot[1/\Delta_n]\cdot\overline{\boldsymbol{F}}\,\overline{\boldsymbol{F}}^T\Big) < \frac{1}{2}\lambda_{\min}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\right) \qquad \text{(B.17)}\\
&\leq \frac{CK^2}{T^2},
\end{aligned}$$

and

$$\begin{aligned}
&P\left(\lambda_{\max}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})(\boldsymbol{F}_{t[j]}-\overline{\boldsymbol{F}})^T\Big) > 2\lambda_{\max}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\right)\\
&= P\left(\lambda_{\max}\Big(\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\boldsymbol{F}_{t[j]}\,\boldsymbol{F}_{t[j]}^T - T\cdot[1/\Delta_n]\cdot\overline{\boldsymbol{F}}\,\overline{\boldsymbol{F}}^T\Big) > 2\lambda_{\min}\Big(\int_0^T \boldsymbol{\Phi}_s d_s\Big)\right) \qquad \text{(B.18)}\\
&\leq \frac{CK^2}{T^2}.
\end{aligned}$$

The assumptions that $\log T/|\log\Delta_n| = O(1)$ and $N = O(T^\gamma)$ imply that there exists $\delta_0 \in (0,1)$ such that $N^{1/(2\gamma)} = o((T/\Delta_n)^{\frac{1}{2(1-\delta_0)}})$. Applying Lemma 1 to $\boldsymbol{X}_{tT}/\sqrt{T}$ and $\boldsymbol{Z}_{tT}/\sqrt{T}$ with $x = N^{1/(2\gamma)}$ and $\delta = \delta_0$, and using Bonferroni's inequality again, we obtain under Assumption 4 that

$$P\left(\max_{1\le k\le K,1\le i\le N}\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}F_{i,t[j]}U_{i,t[j]}\Big|>C\sqrt{\Delta_n N^{1/\gamma}}\right)\le\frac{CKN}{N^{M\delta_0/(2\gamma)}}\le\frac{CK}{N}, \tag{B.19}$$

where the last inequality holds by the assumption that $M>4(1+2\gamma)$ and we can choose δ_0 close to one such that $M\delta_0/(2\gamma)>2$. By Assumption 4, the Burkholder-Davis-Gundy inequality and Jensen's inequality, we have

$$\begin{aligned}&\max_{1\le i\le N,1\le t\le T}E\left(\Big(\int_{t-1}^{t}\boldsymbol{\zeta}_s d\boldsymbol{B}_s\Big)_i^{2M}\right)\\ &\le\max_{1\le i\le N,1\le t\le T}E\left(\Big(\int_{t-1}^{t}\boldsymbol{\Theta}_{s,ii}ds\Big)^{M}\right)\le\max_{1\le i\le N,1\le t\le T}E\left(\Big(\int_{t-1}^{t}\boldsymbol{\Theta}_{s,ii}^{M}ds\Big)\right)\le C.\end{aligned}$$

By Lemma 2(ii), under the assumption that $M>4(1+2\gamma)$, we have

$$P\left(\Big\|\int_0^T\boldsymbol{\zeta}_s d\boldsymbol{B}_s\Big\|_{\max}>C\sqrt{T\log N}\right)\le C\Big(\frac{1}{N}+\frac{1}{T}\Big).$$

Noting that $\overline{U}_i=(\int_0^T\boldsymbol{\zeta}_s d\boldsymbol{B}_s)/(T[1/\Delta_n])$, we get

$$P\left(\max_{1\le i\le N}|\overline{U}_i|>C\Delta_n\sqrt{\frac{\log N}{T}}\right)\le C\Big(\frac{1}{N}+\frac{1}{T}\Big). \tag{B.20}$$

Combining (B.15) and (B.20) yields, for large T,

$$P\left(\max_{1\le k\le K,1\le i\le N}\Big|T\cdot[1/\Delta_n]\cdot\overline{F}_k\overline{U}_i\Big|>\frac{C}{2}\Delta_n\sqrt{T\log N}\right)\le C\left(\frac{1}{N}+\frac{K}{T}\right). \tag{B.21}$$

By (B.19), (B.21) and that $\Delta_n = o(1)$,

$$\begin{aligned}
&P\left(\max_{1\le k\le K, 1\le i\le N}\Big|\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(F_{k,t[j]}-\overline{F}_i)(U_{i,t[j]}-\overline{U}_i)\Big| > C\sqrt{\Delta_n T N^{1/\gamma}}\right)\\
=&P\left(\max_{1\le k\le K, 1\le i\le N}\Big|\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}F_{k,t[j]}U_{i,t[j]} - T\cdot[1/\Delta_n]\cdot\overline{F}_k\overline{U}_i\Big| > C\sqrt{\Delta_n T N^{1/\gamma}}\right)\\
\le&P\left(\max_{1\le k\le K, 1\le i\le N}\Big|\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}F_{k,t[j]}U_{i,t[j]}\Big| > \frac{C}{2}\sqrt{\Delta_n T N^{1/\gamma}}\right)\\
&+P\left(\max_{1\le k\le K, 1\le i\le N}\Big|T\cdot[1/\Delta_n]\cdot\overline{F}_k\overline{U}_i\Big| > \frac{C}{2}\sqrt{\Delta_n T N^{1/\gamma}}\right)\\
\le&CK\left(\frac{1}{N}+\frac{1}{T}\right).
\end{aligned} \tag{B.22}$$

Combining (B.12), (B.17), (B.18) and (B.22), we get that $P(A) \ge 1 - O(1/N + 1/T)$. The desired bound (B.9) follows.

As to (B.10), note that $\widehat{\boldsymbol{\alpha}_n} - \overline{\boldsymbol{\alpha}_n} = (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\overline{\boldsymbol{F}} + \overline{\boldsymbol{U}}$. Hence,

$$\|\widehat{\boldsymbol{\alpha}_n} - \overline{\boldsymbol{\alpha}}_n\|_{\max} \le \|(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\overline{\boldsymbol{F}}\|_{\max} + \|\overline{\boldsymbol{U}}\|_{\max} \le \max_{1\le i\le N}\|\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\|_2\cdot\|\overline{\boldsymbol{F}}\|_2 + \|\overline{\boldsymbol{U}}\|_{\max}.$$

The bound (B.10) follows from (B.9), (B.15) and (B.20). □

Lemma 4. *Under the assumptions of Theorem 2, for some $0 < \varepsilon < M/4 - 1 - 2\gamma$ and $C_0 > 0$, $RV_{\widehat{\boldsymbol{U}}}$ defined in (2.8) satisfies*

$$P\left(\max_{1\le i\le N}\frac{1}{T}\left|\sum_{t=1}^{T}RV_{\widehat{\boldsymbol{U}};it} - \sum_{t=1}^{T}V_{\boldsymbol{U};it}\right| > C_0\sqrt{\Delta_n}\right) \le C_0\left(\frac{1}{N}+\frac{1}{T}\right), \tag{B.23}$$

and

$$P\left(\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}V_{\boldsymbol{U};jt} - \frac{1}{T}\sum_{t=1}^{T}RV_{\widehat{\boldsymbol{U}};it}RV_{\widehat{\boldsymbol{U}};jt}\right| > C_0\sqrt{\Delta_n}\right) \le C_0\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right). \tag{B.24}$$

Proof: Recall that $RV_{\widehat{\boldsymbol{U}};it} = \sum_{j=1}^{[1/\Delta_n]}\widehat{U}_{i,t[j]}^2$. We write $RV_{\boldsymbol{U};it} = \sum_{j=1}^{[1/\Delta_n]}U_{i,t[j]}^2$.

About (B.23), we consider the following event

$$B = \left\{ \max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} (\widehat{U}_{i,t[j]} - U_{i,t[j]})^2 \le CK^2\Delta_n\Big(1+\frac{N^{1/\gamma}}{T}\Big) \right\}$$
$$\bigcap \left\{ \max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T} RV_{\boldsymbol{U};it} - \frac{1}{T}\sum_{t=1}^{T} V_{\boldsymbol{U}_i;t} \right| \le C\sqrt{\frac{\Delta_n N^{1/\gamma}}{T}} \right\}$$
$$\bigcap \left\{ \max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T} V_{\boldsymbol{U}_i;t} \right| \le C \right\} \quad \text{for some } C>0\,.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T} RV_{\widehat{\boldsymbol{U}};it} - \frac{1}{T}\sum_{t=1}^{T} RV_{\boldsymbol{U};it} \right| \\
&\le \max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} (\widehat{U}_{i,t[j]} - U_{i,t[j]})^2 \right| + 2\max_{1\le i,s\le N} \left| \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} (\widehat{U}_{i,t[j]} - U_{i,t[j]})U_{s,t[j]} \right| \\
&\le \max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} (\widehat{U}_{i,t[j]} - U_{i,t[j]})^2 \right| \\
&\quad + 2\sqrt{\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]} (\widehat{U}_{i,t[j]} - U_{i,t[j]})^2} \cdot \sqrt{\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T} RV_{\boldsymbol{U};it}}\,.
\end{aligned}$$

Under event B, by the triangle inequality,

$$\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T} RV_{\boldsymbol{U};it} \le C\left(1+\sqrt{\frac{\Delta_n N^{1/\gamma}}{T}}\right) < 2C,$$

where the last inequality holds by the assumptions that $N^{1/\gamma} = O(T)$ and $\Delta_n = o(1)$. Therefore, under event B,

$$\max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T} RV_{\widehat{\boldsymbol{U}};it} - \frac{1}{T}\sum_{t=1}^{T} RV_{\boldsymbol{U};it} \right| \le C'K\sqrt{\Delta_n}\,.$$

It follows from the triangle inequality that

$$\begin{aligned}
&\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}RV_{\widehat{\boldsymbol{U}};it}-\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}\right| \\
\le&\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}RV_{\widehat{\boldsymbol{U}};it}-\frac{1}{T}\sum_{t=1}^{T}RV_{\boldsymbol{U};it}\right|+\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}RV_{\boldsymbol{U};it}-\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}\right| \\
\le& C^{''}K\sqrt{\Delta_n}\,.
\end{aligned}$$

It remains to show that $P(B)\ge 1-O(K/N+K^2/T)$. The assumptions that $N=O(T^\gamma)$, $\log T/|\log\Delta_n|=O(1)$, and $M>4(1+2\gamma)$ imply that there exists δ_0 such that $M\delta_0/(2\gamma)>2$ and $N^{1/(2\gamma)}=o\big((T/\Delta_n)^{\frac{1}{2(1-\delta_0)}}\big)$. Applying Lemma 1(ii) to $\boldsymbol{Z}_{tT}/\sqrt{T}$ with $x=N^{1/(2\gamma)}$ and $\delta=\delta_0$, and using Bonferroni's inequality, under Assumption 4, we have, for some $C>0$,

$$P\left(\max_{1\le i\le N}\Big|\frac{1}{T}\sum_{t=1}^{T}RV_{\boldsymbol{U};it}-\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}\Big|>\sqrt{\frac{\Delta_nN^{1/\gamma}}{T}}\right)\le\frac{C_2N}{N^{M\delta_0/(2\gamma)}}\le\frac{C}{N}. \tag{B.25}$$

By (2.7) and the inequality that $(a+b)^2\le 2a^2+2b^2$, for each $1\le i\le N$, $1\le t\le T$ and $1\le j\le[1/\Delta_n]$,

$$\begin{aligned}
(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2&\le 2(\widehat{\alpha_{ni}}-\alpha_{n;t[j]i})^2+2\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\|\boldsymbol{F}_{t[j]}\|_2^2\\
&\le 4(\widehat{\alpha_{ni}}-\overline{\alpha_{ni}})^2+4(\overline{\alpha_{ni}}-\alpha_{n;t[j]i})^2+2\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\|\boldsymbol{F}_{t[j]}\|_2^2.
\end{aligned}$$

By the assumption that $\sup_{s\ge0}\|\boldsymbol{\alpha}_s\|_{\max}=O(1)$, we have

$$\begin{aligned}
&\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\\
\le&2\max_{1\le i\le N}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\|\boldsymbol{F}_{t[j]}\|_2^2\right)+4[1/\Delta_n]\cdot\|\widehat{\boldsymbol{\alpha}_n}-\overline{\boldsymbol{\alpha}_n}\|_{\max}^2+C\Delta_n.
\end{aligned} \tag{B.26}$$

By (B.12) and (B.13), for some $C' > 0$, when T is large, we have

$$P\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}\|\boldsymbol{F}_{t[j]}\|_2^2 > KC'\right) \leq \frac{C'K^2}{T}. \tag{B.27}$$

Combining (B.9), (B.10), (B.26) and (B.27) yields, for some $C > 0$,

$$P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]} - U_{i,t[j]})^2 > CK^2\Delta_n\left(1+\frac{N^{1/\gamma}}{T}\right)\right) \leq C\left(\frac{K}{N}+\frac{K^2}{T}\right). \tag{B.28}$$

Under Assumptions 1 and 4, by Lemma 2(ii),

$$P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it} - E(V_{\boldsymbol{U};i})\right| \geq C\sqrt{\frac{\log N}{T}}\right) \leq C\left(\frac{1}{N}+\frac{1}{T}\right). \tag{B.29}$$

Assumption 4 implies $\max_{1\leq i\leq N} E(V_{\boldsymbol{U};i}^4) = O(1)$. By the assumption that $N = O(T^\gamma)$, we have for some $C > 0$,

$$P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it} > C\right) \leq C\left(\frac{1}{N}+\frac{1}{T}\right). \tag{B.30}$$

Combining (B.25), (B.28) and (B.30) yields $P(B) \geq 1 - O(1/N + 1/T)$. The desired bound (B.23) follows.

As to (B.24), by the triangle inequality and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}V_{\boldsymbol{U};jt}-\frac{1}{T}\sum_{t=1}^{T}RV_{\widehat{\boldsymbol{U}};it}RV_{\widehat{\boldsymbol{U}};jt}\right|\\
\le&\max_{1\le i,j\le N}\left|\frac{2}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}(V_{\boldsymbol{U};jt}-RV_{\widehat{\boldsymbol{U}};jt})\right|\\
&+\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}(V_{\boldsymbol{U};it}-RV_{\widehat{\boldsymbol{U}};it})(V_{\boldsymbol{U};jt}-RV_{\widehat{\boldsymbol{U}};jt})\right| \qquad \text{(B.31)}\\
\le&2\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}V_{\boldsymbol{U};it}^2}\cdot\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{\boldsymbol{U};it}-RV_{\widehat{\boldsymbol{U}};it})^2}\\
&+\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{\boldsymbol{U};it}-RV_{\widehat{\boldsymbol{U}};it})^2.
\end{aligned}$$

Under Assumption 3, by Lemma 2(ii), we have, for some $\varepsilon < M/4-1-2\gamma$,

$$P\left(\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{U;it}V_{U;jt}-E(V_{U;i}V_{U;j})\right|>C\sqrt{\frac{\log N}{T}}\right)\le C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right). \qquad \text{(B.32)}$$

Assumption 4 implies that $\max_{1\le i\le N}E(V_{U;i}^M)=O(1)$. By Lemma 2(ii) and $M>4(1+2\gamma)$,

$$P\left(\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{U;it}^2-E(V_{U;i}^2)\right|\ge C\sqrt{\frac{\log N}{T}}\right)\le C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right).$$

This implies that

$$P\left(\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}V_{U;it}^2>C\right)\le C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right). \qquad \text{(B.33)}$$

By Lemma 1, for any $1\le t\le T$, $1\le i\le N$, $0<\delta<1$, there exist constants $C, C_1>0$ such that for all $x\le C\Delta_n^{-\frac{1}{2(1-\delta)}}$,

$$P\left(\left|RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it}\right|>\sqrt{\Delta_n}x\right)\le\frac{C_1}{x^{M\delta}}. \qquad \text{(B.34)}$$

We define $(RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it})_{tr}=(RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it})\cdot\mathbf{1}_{\left\{|RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it}|\le\sqrt{\Delta_n}T^{(1+\gamma+\varepsilon)/(M\delta)}\right\}}$, where

$\varepsilon > 0$ and $0 < \delta < 1$ are to be determined. The assumptions $\log T/|\log \Delta_n| = O(1)$ and $M > 4(1+2\gamma)$ imply that, for some ε sufficiently small, we have

$$\Delta_n^{1/\varepsilon} T = o(1) \text{ and } M > 4(1+\gamma+\varepsilon).$$

Set δ satisfying $M\delta > 4(1+\gamma+\varepsilon)$, and $\delta/(1-\delta) > 2(1+\gamma+\varepsilon)/(M\varepsilon)$. Then

$$T^{(1+\gamma+\varepsilon)/(M\delta)} = o\big(\Delta_n^{-1/(2(1-\delta))}\big).$$

Hence, by (B.34), for any $1 \le t \le T$, $1 \le i \le N$ and all $x > 0$,

$$P\left(\frac{(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}}{\Delta_n} > x\right) \le \frac{C_1}{x^{M\delta/2}}.$$

This implies that, for all $0 < y < M\delta/2$,

$$\max_{1\le i\le N, 1\le t\le T} E\left(\left(\frac{(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}}{\Delta_n}\right)^y\right) = O(1).$$

In particular, by setting $y = 1$,

$$\max_{1\le i\le N, 1\le t\le T} E\left(\frac{(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}}{\Delta_n}\right) < C. \tag{B.35}$$

Applying Lemma 2(ii) to $(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}/\Delta_n - E\Big((RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}/\Delta_n\Big)$, and by $M\delta > 4(1+\gamma)$, we have that

$$P\left(\left|\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}}{\Delta_n} - E\Big(\frac{(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr}}{\Delta_n}\Big)\right)\right| > C\sqrt{\frac{\log N}{T}}\right)$$
$$\le C\left(\frac{1}{N} + \frac{1}{T^{\varepsilon}}\right). \tag{B.36}$$

By (B.35), (B.36) and the triangle inequality, we get

$$P\left(\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}(RV_{\boldsymbol{U};it} - V_{\boldsymbol{U};it})^2_{tr} > C\Delta_n\right) \le C\left(\frac{1}{N} + \frac{1}{T^{\varepsilon}}\right). \tag{B.37}$$

Note also that by (B.34) and Bonferroni's inequality, we have

$$
\begin{aligned}
&P\left(\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}(RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it})_{tr}^2-\frac{1}{T}\sum_{t=1}^{T}(RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it})^2\right|>0\right)\\
&\le\sum_{i=1}^{N}\sum_{t=1}^{T}P\left(|RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it}|>\sqrt{\Delta_n}T^{(1+\gamma+\varepsilon)/(M\delta)}\right)\le\frac{C_1N}{T^{\gamma+\varepsilon}}=O\left(\frac{1}{T^{\varepsilon}}\right),
\end{aligned}
\tag{B.38}
$$

where the last inequality holds by the assumption that $N=O(T^{\gamma})$. Combining (B.37) and (B.38) yields, for some constant $C>0$,

$$
P\left(\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(RV_{\boldsymbol{U};it}-V_{\boldsymbol{U};it})^2>C\Delta_n\right)\le C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right).
\tag{B.39}
$$

Moreover, by the inequality $(a+b)^2\le 2a^2+2b^2$, we have

$$
\begin{aligned}
&\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(RV_{\widehat{\boldsymbol{U}};it}-RV_{\boldsymbol{U};it})^2\\
&=\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2+2\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})U_{i,t[j]}\right)^2\\
&\le\max_{1\le i\le N}\frac{2}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\right)^2\\
&\quad+\max_{1\le i\le N}\frac{8}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})U_{i,t[j]}\right)^2.
\end{aligned}
\tag{B.40}
$$

Applying the Cauchy-Schwarz inequality repeatedly yields

$$
\begin{aligned}
&\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})U_{i,t[j]}\right)^2\\
&\le\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\right)\cdot\left(\sum_{j=1}^{[1/\Delta_n]}U_{i,t[j]}^2\right)\\
&\le\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\right)^2}\cdot\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}RV_{\boldsymbol{U};it}^2}.
\end{aligned}
\tag{B.41}
$$

By (B.33) and (B.39), we have, for some constant $C > 0$,

$$P\left(\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}RV_{U;it}^2 > C\right)\le C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right). \tag{B.42}$$

Similarly, we have

$$\begin{aligned}
&\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\Big)^2}\\
\le&\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{[1/\Delta_n]}2\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\|\boldsymbol{F}_{t[j]}\|_2^2+\sum_{j=1}^{[1/\Delta_n]}2|\widehat{\alpha_{ni}}-\alpha_{n;t[j]i}|^2\right)^2}\\
\le&\max_{1\le i\le N}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\sqrt{\frac{8}{T}\sum_{t=1}^{T}\Big(\sum_{k=1}^{K}\sum_{j=1}^{[1/\Delta_n]}F_{k,t[j]}^2\Big)^2}\\
&+\max_{1\le i\le N}\sqrt{\frac{8}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{[1/\Delta_n]}|\widehat{\alpha_{ni}}-\alpha_{n;t[j]i}|^2\Big)^2}\\
\le&\max_{1\le i\le N}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_2^2\cdot\sqrt{\frac{8K}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}RV_{F;kt}^2}\\
&+8[1/\Delta_n]\cdot\|\widehat{\boldsymbol{\alpha}_n}-\overline{\boldsymbol{\alpha}_n}\|_{\max}^2+8\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{[1/\Delta_n]}|\overline{\alpha_{ni}}-\alpha_{n;t[j]i}|^2\Big)^2},
\end{aligned}$$

where $RV_{F;kt}=\sum_{j=1}^{[1/\Delta_n]}F_{k,t[j]}^2$. By the assumption that $\sup_{s\ge 0}\|\boldsymbol{\alpha}_s\|_{\max}=O(1)$, we have

$$\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{[1/\Delta_n]}|\overline{\alpha_{ni}}-\alpha_{n;t[j]i}|^2\Big)^2}\le C\Delta_n. \tag{B.43}$$

Moreover, by Lemma 2(i), similar to the proof of (B.33), (B.39) and (B.42), one can show that, for some $C > 0$,

$$P\Big(\max_{1\le k\le K}\frac{1}{T}\sum_{t=1}^{T}RV_{F;kt}^2 > C\Big)\le\frac{C}{T^{\varepsilon}}. \tag{B.44}$$

Under the assumption that $N = O(T^{\gamma})$, combining (B.43), (B.44) and Lemma 3 yields

$$P\left(\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{[1/\Delta_n]}(\widehat{U}_{i,t[j]}-U_{i,t[j]})^2\Big)^2}>CK^2\Delta_n\right)\le C\left(\frac{K}{N}+\frac{K^2}{T}\right). \tag{B.45}$$

By the triangle inequality, combining (B.39), (B.40), (B.41), (B.42) and (B.45) yields

$$P\left(\max_{1\le i\le N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(RV_{\widehat{\boldsymbol{U}};it}-V_{\boldsymbol{U};it})^2}>CK\sqrt{\Delta_n}\right)\le C\left(\frac{K}{N}+\frac{1}{T^{\varepsilon}}\right). \tag{B.46}$$

By (B.31), (B.33) and (B.46), the desired bound (B.24) follows. □

Proof of Theorem 1:

Applying the same argument as the proof of (B.25) to the stock RV, under the assumptions of Theorem 1, one can show that, for some $C>0$,

$$P\left(\max_{1\le i\le N}\frac{1}{T}\left|\sum_{t=1}^{T}RV_{it}-\sum_{t=1}^{T}V_{it}\right|>\sqrt{\frac{\Delta_n N^{1/\gamma}}{T}}\right)\le\frac{C}{N}. \tag{B.47}$$

By the triangle inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{it}V_{jt}-\frac{1}{T}\sum_{t=1}^{T}RV_{it}RV_{jt}\right|\\
&\le\max_{1\le i,j\le N}\left|\frac{2}{T}\sum_{t=1}^{T}V_{it}(V_{jt}-RV_{jt})\right|+\max_{1\le i,j\le N}\left|\frac{1}{T}\sum_{t=1}^{T}(V_{it}-RV_{it})(V_{jt}-RV_{jt})\right|\\
&\le 2\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}V_{it}^2}\cdot\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{it}-RV_{it})^2}+\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{it}-RV_{it})^2.
\end{aligned}$$

In addition, under the assumptions of Theorem 1, (B.29), (B.32), (B.33) and (B.39) hold by replacing $RV_{\boldsymbol{U};it}$ with RV_{it}, and $V_{\boldsymbol{U};it}$ with V_{it}. It follows that

$$P\left(\max_{1\leq i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}V_{it}V_{jt}-\frac{1}{T}\sum_{t=1}^{T}RV_{it}RV_{jt}\right|>C\sqrt{\Delta_n}\right)\leq C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right),\ \text{and} \tag{B.48}$$

$$P\left(\|\widehat{\boldsymbol{\Sigma}}_V-\boldsymbol{\Sigma}_V\|_{\max}>C\sqrt{\frac{\log N}{T}}\right)\leq C\left(\frac{1}{N}+\frac{1}{T^{\varepsilon}}\right), \tag{B.49}$$

where $\widehat{\boldsymbol{\Sigma}}_V$ is the sample covariance matrix of V_{it}.

The bound (2.4) follows from (B.47), (B.48) and (B.49). The bounds (2.5) and (2.6) follow from (2.4), Assumption 3, Weyl's Theorem and the $\sin\theta$ Theorem (Davis and Kahan (1970)), which asserts that, for $i\leq q$,

$$\|\widehat{\boldsymbol{\xi}}_{RV_i}-\boldsymbol{\xi}_{V_i}\|\leq\frac{\sqrt{2}\|\widehat{\boldsymbol{\Sigma}}_V-\widehat{\boldsymbol{\Sigma}}_{RV}\|_2}{\min\left(|\widehat{\lambda}_{RV_{i-1}}-\lambda_{V_i}|,|\widehat{\lambda}_{RV_{i+1}}-\lambda_{V_i}|\right)}\,.$$

□

Proof of Theorem 2:

By (B.29) and (B.32), we have

$$\|\widehat{\boldsymbol{\Sigma}}_{V_{\boldsymbol{U}}}-\boldsymbol{\Sigma}_{V_{\boldsymbol{U}}}\|_{\max}=O_p\left(\sqrt{\frac{\log N}{T}}\right), \tag{B.50}$$

where $\widehat{\boldsymbol{\Sigma}}_{V_{\boldsymbol{U}}}$ is the sample covariance matrix of $V_{\boldsymbol{U}}$. By Lemma 4, we have

$$\|\widehat{\boldsymbol{\Sigma}}_{V_{\boldsymbol{U}}}-\widehat{\boldsymbol{\Sigma}}_{RV_{\widehat{\boldsymbol{U}}}}\|_{\max}=O_p\left(\sqrt{\Delta_n}\right). \tag{B.51}$$

Combining (B.50) and (B.51) yields the desired bound (2.9).

The bound (2.10) follows from (B.51), Weyl's Theorem and Assumption 5. The bound (2.11) follows from (2.9) and the $\sin\theta$ Theorem. □

Proof of Proposition 1:

Define

$$\widehat{b}_{i,V} = \frac{\sum_{t=1}^{T}(CV_t - \overline{CV})(V_{it} - \overline{V}_i)}{\sum_{t=1}^{T}(CV_t - \overline{CV})^2}, \qquad \widehat{a}_{i,V} = \overline{V}_i - \widehat{b}_{i,V}\overline{CV}, \tag{B.52}$$

where $\overline{CV} = \sum_{t=1}^{T} CV_t/T$, and $\overline{V}_i = \sum_{t=1}^{T} V_{it}/T$. We have,

$$\widehat{b}_{i,V} - b_i = \Big(\sum_{t=1}^{T}(CV_t - \overline{CV})^2\Big)^{-1} \cdot \Big(\sum_{t=1}^{T}(CV_t - \overline{CV})(\varepsilon_{it} - \overline{\varepsilon}_i)\Big),$$
$$|\widehat{a}_{i,V} - a_i| \leq |\widehat{b}_{i,V} - b_i| \cdot \overline{CV} + |\overline{\varepsilon}_i|,$$

where $\overline{\varepsilon}_i = \sum_{t=1}^{T} \varepsilon_{it}/T$.

Under the assumptions of Theorem 1, $\max_{1\leq i\leq N} E(V_{it}^M) = O(1)$, $E(CV_t^M) = O(1)$, and $\max_{1\leq i\leq N}\Big(|b_{\xi,i}|, |a_{\xi,i}|\Big) = O(1)$. Moreover, under Assumption 6 that $|\bar{b}_\xi| > c$, we have $\max_{1\leq i\leq N}|(b_{\xi,i}/\bar{b}_\xi)| = O(1)$. Hence,

$$\max_{1\leq i\leq N}|E(\varepsilon_{it}^M)| \leq M \cdot \max_{1\leq i\leq N} E(V_{it}^M) + M \cdot |b_{\xi,i}/\bar{b}_\xi|^M \cdot E(CV_t^M) = O(1).$$

By Lemma 2, we have $\sum_{t=1}^{T}(CV_t - \overline{CV})^2/T = \mathrm{Var}(CV_t) + o_p(1)$, and $\overline{CV} = O_p(1)$. By the assumption that $|\bar{b}_\xi| > c$, we have $\mathrm{Var}(CV_t) > \bar{b}_\xi^2 > 0$.

Recall that $CV_t = \bar{a}_\xi + \bar{b}_\xi\xi_t + \bar{\varepsilon}_{\xi,t}$, and $\varepsilon_{it} = \varepsilon_{\xi,it} - E(\varepsilon_{\xi,it}) - (b_{\xi,i}/\bar{b}_\xi)\big(\bar{\varepsilon}_{\xi,t} - E(\bar{\varepsilon}_{\xi,t})\big)$. By Assumption 6, we have

$$\begin{aligned}
&\max_{1\leq i\leq N}|E(CV_t \cdot \varepsilon_{it})| \\
=&\max_{1\leq i\leq N}\Big|E\Big(\big(\bar{\varepsilon}_{\xi,t} - E(\bar{\varepsilon}_{\xi,t})\big) \cdot \big(\varepsilon_{\xi,it} - E(\varepsilon_{\xi,it})\big) - (b_{\xi,i}/\bar{b}_\xi)\big(\bar{\varepsilon}_{\xi,t} - E(\bar{\varepsilon}_{\xi,t})\big)^2\Big)\Big| \\
\leq&\Big(\frac{1}{\sqrt{N}} + \frac{1}{N}\max_{1\leq i\leq N}|b_{\xi,i}/\bar{b}_\xi|\Big) \cdot \Big(\lambda_{\max}\big(\mathrm{Cov}(\boldsymbol{\varepsilon}_{\xi,t})\big)\Big) = O(1/\sqrt{N}).
\end{aligned}$$

By Lemma 2 and that $M > 4(1+2\gamma)$, we get

$$\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T} CV_t \cdot \varepsilon_{it} - E(CV_t \cdot \varepsilon_{it})\right| = O_p\left(\sqrt{\frac{\log N}{T}}\right),$$
$$|\overline{CV} - E(CV_t)| = O_p\left(\sqrt{\frac{\log T}{T}}\right), \quad \text{and} \max_{1\leq i\leq N}|\overline{\varepsilon}_i| = O_p\left(\sqrt{\frac{\log N}{T}}\right).$$

Combining the results above yields that

$$\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}(CV_t-\overline{CV})\cdot(\varepsilon_{it}-\overline{\varepsilon}_i)\right| = O_p\left(\sqrt{\frac{\log N}{T}}+\frac{1}{\sqrt{N}}\right).$$

It follows that

$$\max_{1\le i\le N}\left(|\widehat{b}_{i,V}-b_i|,|\widehat{a}_{i,V}-a_i|\right) = O_p\left(\sqrt{\frac{\log N}{T}}+\frac{1}{\sqrt{N}}\right). \tag{B.53}$$

Next, we bound the differences $\widehat{b}_{i,V}-\widehat{b}_i$ and $\widehat{a}_{i,V}-\widehat{a}_i$, where, recall that $\widehat{b}_i$ and $\widehat{a}_i$ are defined in (3.4). By (B.47),

$$|\overline{CV}-\overline{CRV}| \le \max_{1\le i\le N}|\overline{RV}_i-\overline{V}_i| = O_p\left(\sqrt{\frac{\Delta_n N^{1/\gamma}}{T}}\right). \tag{B.54}$$

By the triangle inequality and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}CRV_t^2-\frac{1}{T}\sum_{t=1}^{T}CV_t^2\right| \le & 2\sqrt{\frac{1}{T}\sum_{t=1}^{T}CV_t^2}\cdot\sqrt{\frac{1}{T}\sum_{t=1}^{T}(CV_t-CRV_t)^2} \\ & +\frac{1}{T}\sum_{t=1}^{T}(CV_t-CRV_t)^2,\end{aligned}$$

and

$$\begin{aligned}&\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}CV_t\cdot V_{it}-\frac{1}{T}\sum_{t=1}^{T}CRV_t\cdot RV_{it}\right| \\ &\le 2\sqrt{\frac{1}{T}\sum_{t=1}^{T}CV_t^2}\cdot\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{it}-RV_{it})^2} \\ &\quad+\sqrt{\frac{1}{T}\sum_{t=1}^{T}(CV_t-CRV_t)^2}\cdot\sqrt{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(V_{it}-RV_{it})^2}.\end{aligned}$$

By the Cauchy-Schwarz inequality again,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(CV_t - CRV_t)^2 =& \frac{1}{TN^2}\sum_{t=1}^{T}\Big(\sum_{i=1}^{N} V_{it} - \sum_{i=1}^{N} RV_{it}\Big)^2\\ \leq& \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}(V_{it} - RV_{it})^2 \leq \max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}(V_{it} - RV_{it})^2.\end{aligned}$$

Under Assumption 6, we have $\frac{1}{T}\sum_{t=1}^{T} CV_t^2 = O_p(1)$. Moreover, similar to (B.39), one can show that $\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}(RV_{it} - V_{it})^2 = O_p(\Delta_n)$. Combining the results above, we have that

$$\begin{aligned}\Big|\frac{1}{T}\sum_{t=1}^{T} CRV_t^2 - \frac{1}{T}\sum_{t=1}^{T} CV_t^2\Big| =& O_p(\sqrt{\Delta_n}), \quad \text{and}\\ \max_{1\leq i\leq N}\Big|\frac{1}{T}\sum_{t=1}^{T} CV_t\cdot V_{it} - \frac{1}{T}\sum_{t=1}^{T} CRV_t\cdot RV_{it}\Big| =& O_p(\sqrt{\Delta_n}).\end{aligned} \tag{B.55}$$

Combining (B.47), (B.52), (B.54) and (B.55) yields

$$\max_{1\leq i\leq N}|\widehat{b}_{i,V} - \widehat{b}_i| = O_p(\sqrt{\Delta_n}), \quad \text{and} \quad \max_{1\leq i\leq N}|\widehat{a}_{i,V} - \widehat{a}_i| = O_p(\sqrt{\Delta_n}). \tag{B.56}$$

The desired bound (3.5) follows from (B.53) and (B.56). □

References

Cai, T Tony, Jianchang Hu, Yingying Li, and Xinghua Zheng (2020): "High-dimensional minimum variance portfolio estimation based on high-frequency data," *Journal of Econometrics*, 214, 482–494.

Davis, Chandler and William Morton Kahan (1970): "The rotation of eigenvectors by a perturbation. III," *SIAM Journal on Numerical Analysis*, 7, 1–46.

Ding, Yi, Robert Engle, Yingying Li, and Xinghua Zheng (2022): "Factor modeling for volatility," .

Fan, Jianqing, Yingying Li, and Ke Yu (2012): "Vast volatility matrix estimation using high-frequency data for portfolio selection," *J. Amer. Statist. Assoc.*, 107, 412–428.

MERLEVÈDE, FLORENCE, MAGDA PELIGRAD, EMMANUEL RIO, ET AL. (2009): "Bernstein inequality and moderate deviations under strong mixing conditions," in *High dimensional probability V: the Luminy volume*, Institute of Mathematical Statistics, 273–292.