Journal of Econometrics 249 (2025) 105959

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Multiplicative factor model for volatility

Yi Ding [a], Robert Engle [b], Yingying Li [c,*], Xinghua Zheng [d]

[a] Faculty of Business Administration, University of Macau, Taipa, Macau, China
[b] Stern School of Business, New York University, 44 West Fourth Street, Suite 9-62, NY, United States of America
[c] Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[d] Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

JEL classification:
C13
C51
C53
C55
C58
G17

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

ABSTRACT

Facilitated with high-frequency observations, we introduce a remarkably parsimonious one-factor volatility model that offers a novel perspective for comprehending daily volatilities of a large number of stocks. Specifically, we propose a multiplicative volatility factor (MVF) model, where stock daily variance is represented by a common variance factor and a multiplicative idiosyncratic component. We demonstrate compelling empirical evidence supporting our model and provide statistical properties for two simple estimation methods. The MVF model reflects important properties of volatilities, applies to both individual stocks and portfolios, can be easily estimated, and leads to exceptional predictive performance in both US stocks and global equity indices.

1. Introduction

Volatilities, intrinsically linked with macro- and microeconomics, play a central role in investments, asset pricing, risk management, and monetary policies. The past few decades have witnessed tremendous progress in modeling time-varying volatilities. Pioneer works 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 variance (RV), which is the squared realized volatility, is a consistent estimator of the integrated variance (IV) as the sampling frequency increases when microstructure noise can be safely ignored; 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 variance (TSRV, Zhang et al., 2005), multi-scale realized variance (Zhang, 2006), pre-averaging approach (PAV, Jacod et al., 2009 and Jacod et al., 2019), and quasi-maximum likelihood estimator (Xiu, 2010). Jump-robust variance 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 et al. (2017) develop bootstrap methods for inference on integrated variance. Li et al. (2019) propose an efficient multi-scale jackknife estimator for integrated variance functionals. Empirically, Liu et al. (2015) find that the simple 5-min RVs achieve high estimation accuracies. RV-based models have been further studied for variance prediction, including fractionally-integrated Gaussian vector autoregression for log RV (Andersen et al., 2003) and heterogeneous AR model (HAR, Corsi, 2009).

* Corresponding author.
E-mail addresses: yiding@um.edu.mo (Y. Ding), rengle@stern.nyu.edu (R. Engle), yyli@ust.hk (Y. Li), xhzheng@ust.hk (X. Zheng).

https://doi.org/10.1016/j.jeconom.2025.105959
Received 29 April 2024; Received in revised form 22 January 2025; Accepted 25 January 2025
Available online 21 February 2025
0304-4076/

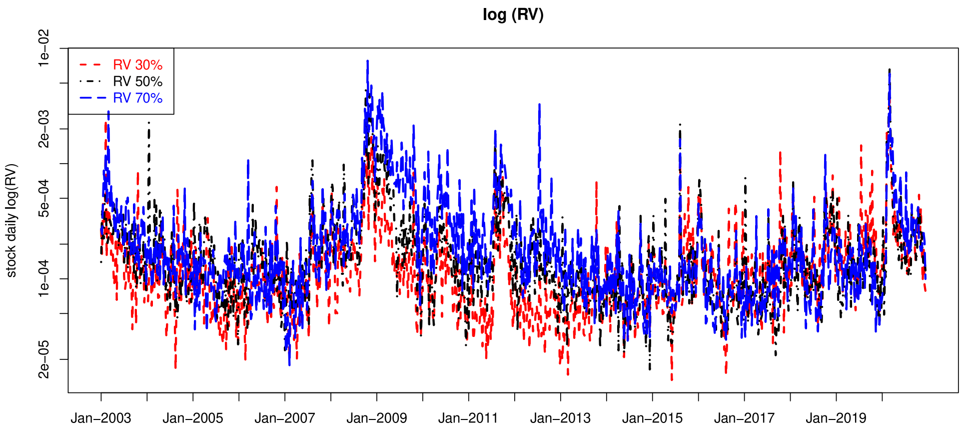

Fig. 1. Time series plots (log scale) of three representative S&P500 Index constituent stocks' RVs (RV 30%, RV 50%, RV 70%) based on 5-min intraday returns from 2003 to 2020 with mean RVs falling on the 30%, 50%, 70% quantiles of all mean RVs.

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-min 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 Fig. 1.

Fig. 1 shows clearly that the stock RVs co-move. Such a co-movement feature in volatilities has been documented. For example, Engle et al. (1990) examine exchange markets and find evidence of volatility spillover in different markets. Calvet et al. (2006) also find a strong correlation among volatilities in exchange rates through a multifrequency volatility decomposition approach. Volatility co-movement in equity markets are studied in Susmel and Engle (1994). Kelly et al. (2013) analyze the impact of firm networks and firm size dispersion on the co-movement of firm volatilities. Herskovic et al. (2016) show that there is strong co-movement in idiosyncratic volatilities of US firms and investigate the pricing implication of a common idiosyncratic volatility factor. Bollerslev et al. (2018) and Engle and Martin (2019) study volatility co-movements in the global multiple asset classes.

Volatility co-movement has been used in volatility modeling and forecasting. For example, elliptical models provide one way to capture such co-movements and have been applied to estimate high-dimensional integrated covolatility matrices (Zheng and Li, 2011, Xia and Zheng, 2018 and Ding and Zheng, 2025). Luciani and Veredas (2015) propose a factor model in log volatilities and estimate the factors using principal components in log realized volatilities. Barigozzi and Hallin (2017) study a generalized dynamic factor model for volatilities in log-linear form. Asai and McAleer (2015) predict co-volatilities based on a multivariate factor model of returns, which implies a factor model for variances in linear form. Bollerslev et al. (2018) propose a global volatility factor from multiple asset classes and use it as a predictor of volatilities within time-series models in linear form. Barigozzi and Hallin (2020) discuss a factor model for log variance and the model estimation. Kapadia et al. (2023) identify a common factor volatility component in volatilities of traded equity factors and show its usefulness in forecasting log volatilities of equity factors.

In this paper, we propose a remarkably parsimonious high-dimensional volatility model for the cross-section of stock variances, which offers a novel perspective for comprehending the volatilities in the financial market and leads to exceptional predictive performance when exercised on a large number of stocks. Specifically, we propose a single factor model in multiplicative form, which we name *multiplicative volatility factor* (MVF) model. The MVF model captures the co-movements of volatilities for hundreds or thousands of stocks simultaneously. In our proposed MVF model, the variance is represented by a multiplicative common factor and an idiosyncratic variance exposure:

$$V_{it} = \mathrm{V}_t^F\, \xi_{it}, \quad 1 \le i \le N, \tag{1.1}$$

where V_t^F is the multiplicative latent factor, and ξ_{it} is a multiplicative idiosyncratic component that is independent with V_t^F. We refer to ξ_{it} as the idiosyncratic variance exposure (iE).

The MVF model (1.1) has several desirable properties. First, it reflects important volatility properties such as non-negativity and heavy-tailedness. Second and third, it is notably parsimonious, which makes it easy to estimate and utilize in volatility forecasting. Fourth, our MVF model not only applies to individual stocks but also to portfolios. Lastly, it greatly enhances the volatility forecasting accuracy when applied to a large number of stocks. We explain these properties in more detail below.

First, the multiplicative form helps to capture the nonnegativity and heavy-tailedness in volatilities, overcoming limitations from models in the usual linear form. Examples of multiplicative models include the lognormal stochastic volatility model (Hull and White, 1987), EGARCH (Nelson, 1991), and realized-GARCH with log-linear specification (Hansen et al., 2012). Multiplicative models are also used when describing low-frequency components in volatilities; see, for example, the Spline-GARCH model (Engle and Rangel, 2008) and the multiplicative GARCH model (Engle and Sokalska, 2012). These studies focus on univariate modeling of volatilities. Different from these studies, our proposed MVF model uses a multiplicative factor to capture the common movement of volatilities in high dimensions. Connor et al. (2006) study a common multiplicative component in idiosyncratic volatilities based on analysis of monthly returns. Our analysis, on the other hand, is based on daily volatility measures obtained using high-frequency data. Moreover, our study uncovers the presence of a common multiplicative component in volatilities of both return factors and idiosyncratic returns.

Second, the MVF model (1.1) is remarkably parsimonious. The model is considerably simpler than those induced by return factor models; see Section 4.1.2 for a list of such models; and Section 2.3 for discussions on its relationship with return-factor models. The MVF model is also connected to the log-linear single factor model:

$$\log(V_{it}) = a_i' + b_i' \mathrm{V}_t^F + \varepsilon_{it}'. \tag{1.2}$$

There have been studies that focus on factor models for log variances; see, e.g., Barigozzi and Hallin (2017, 2020). The MVF model (1.1) can be rewritten as a factor model for logarithms of variances with all exposure coefficients being one. That is,

$$\log(V_{it}) = \log(\mathrm{V}_t^F) + \log(\xi_{it}). \tag{1.3}$$

Compared with model (1.2), the MVF model is more parsimonious in terms of the number of parameters that need to be estimated. In addition, unlike model (1.2), the MVF model (1.1) or (1.3) is readily applicable to portfolios, which we elaborate in the fourth point below.

Third, our MVF model is easy to estimate. We develop two estimators of the common latent factor V_t^F. The first estimator is the common realized variance (CRV), that is, the cross-sectional average of the stock realized variances, $\mathrm{CRV}_t = \sum_{i=1}^N RV_{it}/N$. The second estimator is the first principal component based on principal component analysis (PCA) on the stock realized variances. Under the condition that ξ_{it} are cross-sectionally weakly dependent, we show that CRV_t is consistent in estimating V_t^F. The consistency of CRV_t is established under a high-dimensional and high-frequency setting where both the number of stocks and the sampling frequency are large. About the estimator based on PCA, we establish conditions for its consistency, which requires stronger moment assumptions and also a large time span. Throughout our asymptotic analysis, we take the errors of RV in estimating the volatilities into consideration. Compared with the estimator based on PCA, CRV_t has the advantage that it is more interpretable, easier in terms of estimation, and enjoys theoretical consistency under weaker regularity conditions. Therefore, we recommend to estimate the latent factor with CRV_t. The idiosyncratic variance exposure is estimated using the idiosyncratic realized variance exposure (iRE), $\hat{\xi}_{it} = RV_{it}/\mathrm{CRV}_t$. For volatility forecasting, note that $RV_{it} = \mathrm{CRV}_t \times \hat{\xi}_{it}$, and the two components can be modeled separately by, for example, a log HAR model (Corsi, 2009). The stock variance forecast is then the product of the forecasts of the two components.

Fourth, the MVF model (1.1) naturally applies to stocks and portfolios, which is an advantage over the models in the usual log form like (1.2). Specifically, when evaluating portfolio risk, the variance of a portfolio involves a linear combination of underlying stock variances. Under the MVF model, the common volatility factor in individual stocks is also the volatility factor in portfolios.

Finally, our proposed MVF model substantially enhances volatility forecasting performance when exercised on a large number of stocks. In our empirical analysis, we apply our MVF model to the S&P 500 Index constituent stocks and find that our approach statistically and economically significantly outperforms the benchmark models in more than 90% of the stocks that we evaluate. Beyond the US market, we find that the MVF model also applies to the global market and outperforms in the volatility forecasting based on a parallel analysis using daily realized variances of 31 global equity indices.

The rest of this paper is organized as follows. In Section 2, we discuss the evidence of the MVF model. Section 3 gives the theoretical results. 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. Additional theoretical results and the proofs are collected in the Supplementary Materials (Ding et al., 2025).

2. Empirical evidence

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-min 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 et al., 2001). 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-min intraday log-returns per day. About return factors, we consider the Fama–French three-factor model (Fama and French, 1993) and use 5-min returns of the market, the small-minus-big (SmB) and the high-minus-low (HmL) portfolios.[1]

Following Bollerslev and Todorov (2011), Aït-Sahalia et al. (2013) and Li et al. (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 $R_{i;t[j]}^{tr} = R_{i;t[j]} \mathbf{1}_{\{|R_{i;t[j]}| \le v_{it}\}}$, $1 \le j \le 77$, 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. We can as well include market microstructure noise in our study. The proofs will be technically more involved, but the main message remains the same.

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

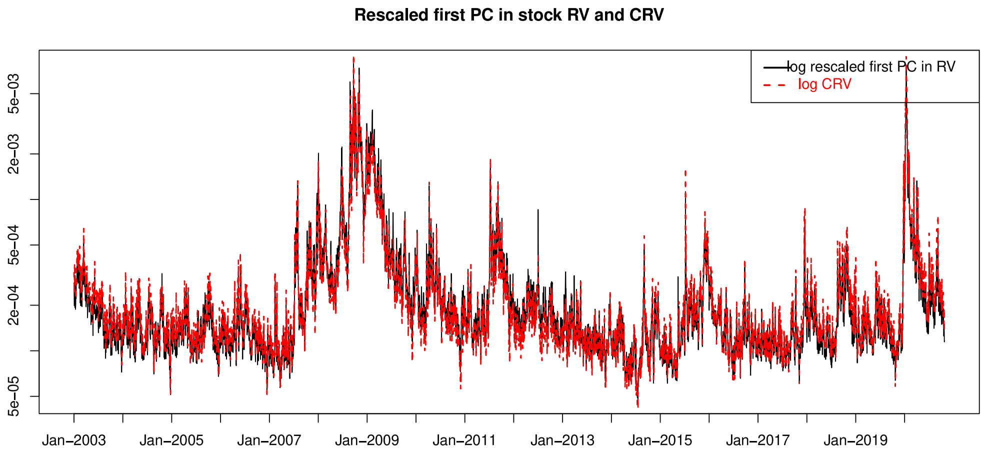

Fig. 2. Time series plots of common realized variance (CRV) and rescaled first PC in stock RVs. The plots are drawn on log scale to improve visibility.

2.2. Volatility factor

Co-movement in stock volatilities is natural given that returns admit a factor structure. In fact, the factor structure in returns will induce a factor structure in variance. In Section 2.3.3, we show that the first PC in stock RVs explains 60% of the total variations. Interestingly, we find that all return factors' volatilities are highly correlated with each other and with the first PC, suggesting that both return factor variances and idiosyncratic variances are driven by a single variance factor.

In order to construct the common factor in stock variances, by Theorem 2 below, we can use the first PC in stock RVs. Alternatively, by Theorem 1 below, we can use the common realized variance (CRV), that is, the cross-sectional average of stock realized variances:

$$\mathrm{CRV}_t = \frac{1}{N}\sum_{i=1}^{N} RV_{it}.$$

In Fig. 2, 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 largely coincides with CRV with a high correlation of 0.979. Compared with the first PC, the CRV is more interpretable and easier to estimate. Theoretically, as we will show in Section 3 below, compared with the CRV, using the PC to estimate the common factor will result in an additional error term of order $\sqrt{(\log N)/T}$. In addition, the consistency of PCA requires stronger assumptions on the moment and the time span T. Therefore, we recommend to estimate the latent factor with the CRV.

Remark 1. Besides CRV, one may also consider VIX (the CBOE Market Volatility Index, transformed to daily variance) as a possible candidate for the volatility factor. 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. There has also been a rapid growth in the market for ultra short-maturity options; see, Bandi et al. (2023). Short-term option implied volatilities (IV) can be informative in short horizon. It could be beneficial to incorporate short-term IVs in predicting stock volatilities.

Remark 2. Another potential candidate for the common factor is the common idiosyncratic realized variance (CiRV). We investigate the factor structure on idiosyncratic volatility and find results consistent with Herskovic et al. (2016). Detailed analysis about idiosyncratic volatilities is given in Section 2.3.2. CRV and CiRV have a high correlation of 0.980. In terms of volatility forecasting, the MVF model with CiRV as the factor estimator underperforms the MVF model that uses CRV as the factor estimator.

To examine how well CRV can capture the cross-sectional dependence among stock volatilities, we compute the pairwise correlations among the idiosyncratic realized variance exposures: $\hat{\xi}_{it} = RV_{it}/\mathrm{CRV}_t$. We compare these correlations with the correlations among the RVs. The results are shown in Fig. 3. We see that the RVs are highly correlated with each other. However, once the common factor is removed, the idiosyncratic exposures display much weaker correlations. The average pairwise correlation is 0.628 for the stock realized variances, while for the idiosyncratic exposures, it is 0.147. These findings confirm that the MVF model effectively captures the common component in variances.

2.3. Relationship with return factor models

2.3.1. Volatility factor or return factor volatility

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

$$V_{it} = \beta^2_{i,Mkt} V_{Mkt,t} + \beta^2_{i,HmL} V_{HmL,t} + \beta^2_{i,SmB} V_{SmB,t} + V_{U_i,t} + \text{covariance terms},$$

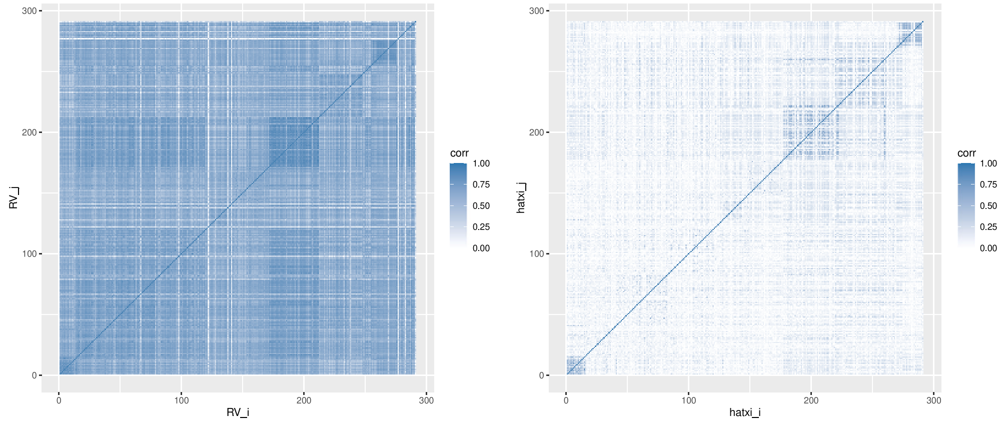

Fig. 3. Heatmap of pairwise correlations among the RVs (left) and among the idiosyncratic realized variance exposures under the MVF model (right). Stocks are sorted into industrial sectors.

Table 1
Regression summary of (2.1) for different stocks. We report the 25% quantile (Q1), median, mean and the 75% quantile (Q3) of the t-statistics of $\text{CRV}_{\varepsilon,t}$, the percentage (Sig.Pct.(%)) and the number (Sig.No.) of stocks in which $\text{CRV}_{\varepsilon,t}$ is significant at the significance level of 0.05.

t-stat($\text{CRV}_{\varepsilon,t}$)				Sig.No.	Sig.Pct. (%)
Q1	Median	Mean	Q3	($t > 2$)	($t > 2$)
6.42	14.05	15.17	22.30	273	93.8

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. The formulation shows that the return factor variances ($V_{Mkt,t}$, $V_{HmL,t}$, $V_{SmB,t}$) are potential factors for the stock variance. One natural question is, can return factors completely explain the volatility factor structure?

To investigate whether CRV has additional explanation power of stock volatilities than the return factor variances, we perform the following analysis. First, we regress CRV_t against the return factor volatilities, $RV_{Mkt,t}$, $RV_{SmB,t}$ and $RV_{HmL,t}$, and observe that the R^2 is 0.91. We denote the residuals by $\text{CRV}_{\varepsilon,t}$. This component is the part of CRV_t that cannot be explained by the return factor volatilities. Then, for each stock i, we regress RV_{it} against the four variables, $RV_{Mkt,t}$, $RV_{SmB,t}$, $RV_{HmL,t}$, and $\text{CRV}_{\varepsilon,t}$:

$$RV_{it} = a_i + b_{i,Mkt} RV_{Mkt,t} + b_{i,SmB} RV_{SmB,t} + b_{i,HmL} RV_{HmL,t} + b_{i,\text{CRV}_\varepsilon} \text{CRV}_{\varepsilon,t} + u_{it}. \tag{2.1}$$

We summarize the t-statistics and the significance of $\text{CRV}_{\varepsilon,t}$ in Table 1. We find that $\text{CRV}_{\varepsilon,t}$ is significant for 93.8% of the stocks (273 out of 291 stocks) at the significance level of 0.05. This result demonstrates that $\text{CRV}_{\varepsilon,t}$ and hence CRV_t contains additional information over the return factor volatilities in explaining the stock volatilities.

In Section 4, we further compare the return factor variances and the CRV in terms of their forecasting performance. The comparison shows that CRV leads to substantially more accurate forecast than the return factor variances.

2.3.2. Factor structure in idiosyncratic variances

We analyze daily idiosyncratic realized variances constructed from 5-min returns of S&P 500 Index constituent stocks and the Fama–French three-factor model. Specifically, we regress the 5-min 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 et al. (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 variance (CiRV). Therefore, the cross-sectional average of idiosyncratic variance, or common idiosyncratic variance (CiV),[2] can be considered as the factor in idiosyncratic variance.

Note that the HmL and SmB factors have non-zero beta exposure to the market, hence their volatilities are correlated with the market volatility as shown in Table 2 below. Such a correlation does not impact the analysis above about idiosyncratic variance. To check the robustness of our findings, we also used statistical factors, which are uncorrelated by construction, to conduct the analysis. The results are similar.[3]

[2] In Herskovic et al. (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.

Table 2
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

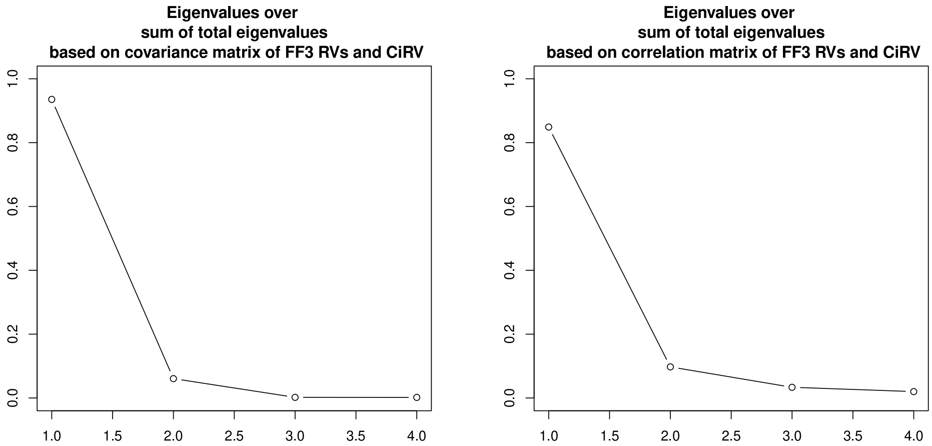

Fig. 4. Eigenvalue ratios of the sample covariance matrix (left panel) and sample correlation matrix (right panel) of FF3 factors' RVs and CiRV.

Our findings of common factor structure in idiosyncratic variances are consistent with several theoretical and empirical evidence in the literature, e.g., Barigozzi and Hallin (2016) and Renault et al. (2023). The existence of factors in idiosyncratic variances has significant implications in the estimation of return factor models. For example, in the presence of idiosyncratic volatility factors, the asymptotic homoskedasticity assumption, which is necessary for consistent latent factor estimations in short panel with large cross-sectional dimension, will fail; see Bai (2003) and Fortin et al. (2023a,b).

2.3.3. PCA analysis of factor structure in stock variances

In Section 2.3.2, we find a single factor in idiosyncratic variance, CiV. In addition, there are three return factor volatilities. Therefore, altogether, there are four potential factors in stock variances. An interesting question arises: 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 Fig. 4. 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 Fig. 5.

Fig. 5 shows that a high proportion (60%) of the total variation in stock RVs is 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 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 2.

Table 2 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 et al. (2016), which show a high correlation between the spot market factor volatilities and idiosyncratic volatilities of sector portfolios. Kapadia et al. (2023) 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.

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

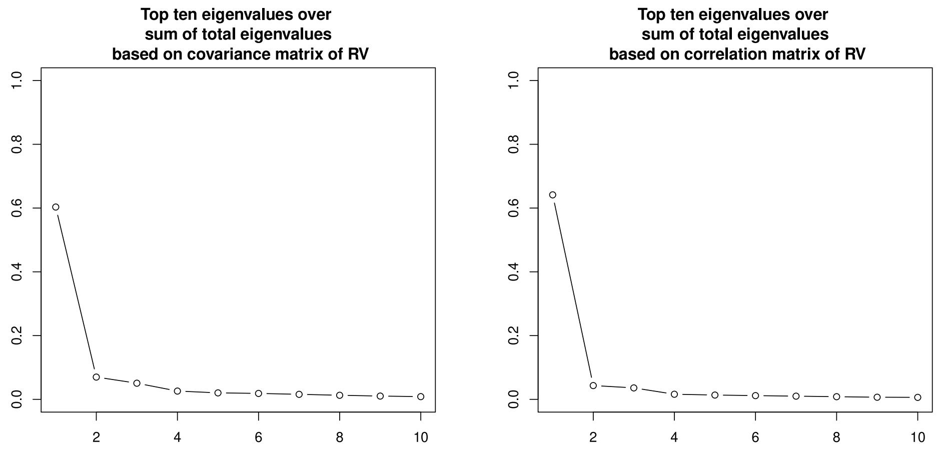

Fig. 5. Top ten eigenvalue ratios of the sample covariance matrix (left panel) and sample correlation matrix (right panel) of stock RVs.

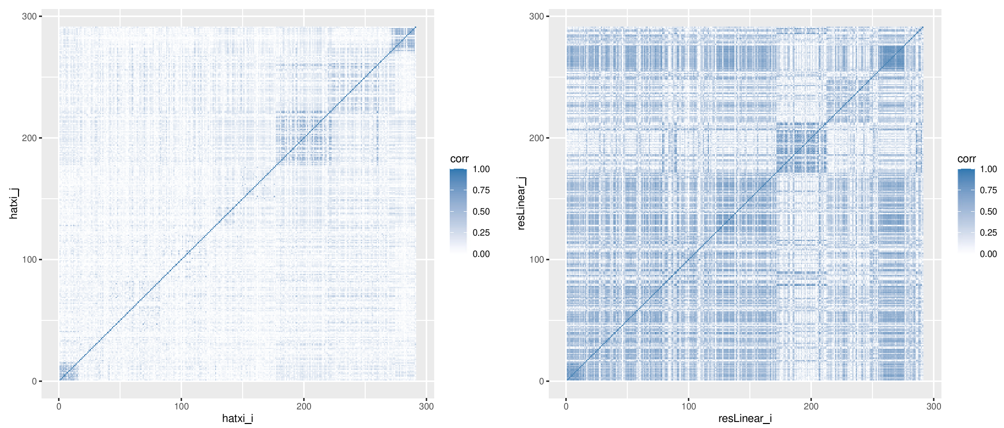

Fig. 6. Heatmap of pairwise correlations among the idiosyncratic exposures under the MVF model (left) and among the residuals under additive linear model (2.2) (right).

2.4. *Volatility factor model*

2.4.1. *MVF model*

As introduced in Section 1, we propose a multiplicative volatility factor (MVF) model in Eq. (1.1), in which the variance is represented by a multiplicative common factor and an idiosyncratic variance exposure.

In the following, we discuss the advantages of our MVF model over models of two alternative forms.

2.4.2. *Multiplicative or additive*

Our MVF model (1.1) takes a multiplicative form, which provides an advantage in capturing important volatility properties such as non-negativity and heavy-tailedness. These properties are not easily achieved by models in linear form. We compare the MVF model (1.1) with the following additive linear factor model:

$$V_{it} = a_i + b_i \mathrm{V}_t^F + \varepsilon_{it}, \quad 1 \le i \le N. \tag{2.2}$$

Again, we use the CRV_t as the proxy of the common factor and estimate model (2.2) using the S&P 500 Index constituent stock RVs. We compute the pairwise correlations among the residuals $\hat{\varepsilon}_{it}$, and compare them with the correlations among the idiosyncratic exposure of the proposed MVF model. The results are collected in Fig. 6. We see that the pairwise correlations of the residuals under the proposed MVF model is weaker than that under the additive model. The average absolute correlation among the idiosyncratic exposure is 0.147, compared to 0.309 for the residuals under model (2.2). The results suggest that the MVF model is more effective in capturing the common component in variances.

2.4.3. MVF or log-linear model

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 (1.2). Our MVF model (1.1) is closely related to (1.2). However, there are subtle but important differences between them.

Comparing the MVF model with the single factor model for log variance, we see that model (1.2) is not easily applicable for portfolio analysis. To be more specific, note that model (1.2) is equivalent to

$$V_{it} = (\mathrm{V}_t^F)^{b_i'} \; e^{a_i' + \epsilon_{it}'},$$

where the coefficient b_i' becomes the exponent. This makes it challenging to interpret. Moreover, one natural choice for estimating the factor V_t^F in (1.2) is the common log realized variance (ClogRV), namely, the cross-sectional average of the log realized variances. One then estimates the coefficients b' in the log-linear model (1.2) by regressing log(RV) over ClogRV. The estimated b' vary from stock to stock. As a result, model (1.2) cannot hold for portfolios.

In addition, compared with model (1.2), our proposed MVF model is more parsimonious because it has the same format as (1.2) but with the slope term constrained to be one, that is, expression (1.3). Note that if all $b_i' = 1$, then model (1.2) becomes our MVF model (1.1). The difference between the two is the choice of factor. 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 largely equivalent to the log-linear single factor model with the slope term constrained to be one.

3. Theoretical results

3.1. Setup and assumptions

Suppose that there are N assets under study. We consider the following continuous-time model for the log price processes $(\boldsymbol{Y}_t)$:

$$d\boldsymbol{Y}_t = \boldsymbol{\alpha}_t dt + \boldsymbol{\zeta}_s d\boldsymbol{B}_s, \quad t \in [0, T], \tag{3.1}$$

where $\boldsymbol{\zeta}_t$ is an $N \times N$ dimensional spot co-volatility process, $\boldsymbol{\alpha}_t$ is the drift process, and $(\boldsymbol{B}_t)$ is a standard N-dimensional Brownian motion. We allow the existence of factors in the log price processes.

For any matrix $A = (a_{ij})$, we denote its entrywise norm by $\|A\|_{\max} := \max_{i,j} |a_{ij}|$; the spectrum norm is denoted by $\|\mathbf{A}\|_2 := \max_{\|\mathbf{x}\|_2 \le 1} \|\mathbf{A}\mathbf{x}\|_2$, where $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$; and the minimum singular value and the maximum singular value are denoted by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. In the following, $c, C, \ldots$, etc, denote generic constants which do not depend on T, N, Δ_n, and can vary from place to place.

We make the following assumptions.

Assumption 1. The process (ζ_t) is càdlàg. The spot covariance matrix, $\boldsymbol{\Psi}_t = \zeta_t \zeta_t^T$, satisfies that $\boldsymbol{\Psi}_t$ and $\boldsymbol{\Psi}_{t-}$ are positive definite for all t, a.s. In addition, $\sup_{s \ge 0} \|\boldsymbol{\alpha}_s\|_{\max} \le C$ for some constant $C > 0$.

For each day $t = 1, \ldots, T$, we denote the integrated variances of the N stocks as $V_t = (V_{1t}, \ldots, V_{Nt})^T$, where

$$V_{it} = \int_{t-1}^{t} \boldsymbol{\Psi}_{s,ii} ds. \tag{3.2}$$

Assumption 2. The integrated variances $(V_{it})_{1 \le i \le N}$ are stationary, and $\max_{1 \le i \le N} \sup_{t \in \mathbb{N}} E\left((\sup_{t-1 \le s < t} \boldsymbol{\Psi}_{s,ii})^M \right) \le c$ for some positive constants c and $M > 0$.

Assumption 3. Model (1.1) holds for $(V_{it})_{1 \le i \le N}$. In addition, for some $c, C > 0$, $E(V_{it}) > c$ for all $i \ge 1$, and $c < \min(\mathrm{Var}(\mathrm{V}_t^F), E(\mathrm{V}_t^F)) \le \max(\mathrm{Var}(\mathrm{V}_t^F), E(\mathrm{V}_t^F)) < C$, and $\|\boldsymbol{\Sigma}_\xi\|_2 = O(N^\delta)$ for some $\delta \in [0, 1)$, where $\boldsymbol{\Sigma}_\xi = \mathrm{Cov}(\xi_t)$ for $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})^T$.

Note that Assumption 3 allows the existence of weak factors in (ξ_{it}).

3.2. Common factor estimation

3.2.1. Consistency of common realized variance

We present the theoretical results for the estimation of the common factor. Suppose that we observe log-returns of stocks and factors at sampling interval Δ_n. For each $t = 1, \ldots, T$ and $j = 1, \ldots, n := [1/\Delta_n]$, we write the log-returns of stocks as $\boldsymbol{R}_{t[j]}$:

$$\boldsymbol{R}_{t[j]} = \boldsymbol{Y}_{t-1+\Delta_n j} - \boldsymbol{Y}_{t-1+\Delta_n (j-1)} =: (R_{1,t[j]}, \ldots, R_{N,t[j]})^T.$$

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.

The following results show the consistency of using $(\mathrm{CRV}_t = \sum_{i=1}^N RV_{it}/N)$ to estimate (V_t^F) up to a scale factor.

Theorem 1. *Under Assumptions 1–3, if $\Delta_n \to 0$, $\log(N)/|\log \Delta_n| = O(1)$, and the M in Assumption 2 is bigger than 2, then for any t,*

$$|CRV_t - V_t^F\,\overline{E(\xi)}| = O_p\left(\sqrt{\Delta_n} + \frac{1}{N^{(1-\delta)/2}}\right),$$

where $\overline{E(\xi)} = \sum_{i=1}^N E(\xi_{it})/N$.

Remark 3. The conditions $\Delta_n \to 0$ and $\log(N)/|\log \Delta_n| = O(1)$ imply that $(1/\Delta_n) \gg N^{\varepsilon}$ for some arbitrarily small $\varepsilon > 0$. The conditions guarantee that $RV_{it} - V_{it}$, after suitable truncation, is uniformly of order $\sqrt{\Delta_n}$ under the ℓ_2 norm; see (B.49)–(B.54) in Appendix B of the Supplementary Materials (Ding et al., 2025) for details.

Using CRV_t, we can estimate the idiosyncratic variance exposure with the idiosyncratic realized variance exposure (iRE), $\hat{\xi}_{it} = RV_{it}/\mathrm{CRV}_t$. The next result shows the consistency of $\hat{\xi}_{it}$ in estimating ξ_{it} up to a scale factor.

Proposition 1. *Under the assumptions of Theorem 1, for all $i \le N$ and any t,*

$$\left|\hat{\xi}_{it} - \frac{1}{\overline{E(\xi)}}\xi_{it}\right| = O_p\left(\sqrt{\Delta_n} + \frac{1}{N^{(1-\delta)/2}}\right).$$

Another potential estimator of the common factor is ExpClogRV, that is, the exponential of the common log RVs defined as:

$$\mathrm{ExpClogRV}_t = \exp\left(\frac{1}{N}\sum_{i=1}^N \log(RV_{it})\right).$$

We present the consistency of $\mathrm{ExpClogRV}_t$ in Proposition 2 of Appendix A of the Supplementary Materials (Ding et al., 2025). The convergence rate of $\mathrm{ExpClogRV}_t$ is the same as that of CRV_t. However, the consistency of ExpClogRV requires a lower bound for both the integrated variances and RV's, imposing stronger conditions on V_{it} and the sampling interval Δ_n.

3.2.2. Consistency of principal component

A common practice to estimate latent factors is to use principal component analysis. In this subsection, we give the results about consistency of using principal component in estimating the common factor in the MVF model.

Denote the ith largest eigenvalue and the corresponding eigenvector of the sample covariance matrix of the realized variances, $\hat{\boldsymbol{\Sigma}}_{RV}$, by $\hat{\lambda}_{RV_i}$ and $\hat{\boldsymbol{\phi}}_{RV_i}$, respectively, where

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

with $RV_t = (RV_{1t}, \dots, RV_{Nt})^T$ and $\overline{RV} = \sum_{t=1}^T RV_t/T$. The estimated factor using the first principal component (PC) is

$$\widehat{\mathrm{V}_t^{F,PC}} = \frac{1}{\sqrt{N}}\hat{\boldsymbol{\phi}}_{RV_1}^T RV_t.$$

The next theorem gives the error bound of using $(\widehat{\mathrm{V}_t^{F,PC}})$ to estimate (V_t^F) up to a scale factor.

Theorem 2. *Suppose that Assumptions 1–3 hold, and $\Delta_n \to 0$, $T \to \infty$, $\log(T)/|\log \Delta_n| = O(1)$, and $N = O(T^{\gamma})$ for some $\gamma > 0$, and the M in Assumption 2 is bigger than $4(1+2\gamma)$. In addition, 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 $\{\boldsymbol{\Psi}_t : -\infty \le t \le 0\}$ and $\{\boldsymbol{\Psi}_t : \chi \le t \le \infty\}$, respectively, satisfy that $\rho(\chi) \le c_1\exp(-c_2\chi)$ for some constants $c_1, c_2 > 0$ and any positive integer χ. Then for any $1 \le t \le T$,*

$$\left|\widehat{V_t^{F,PC}} - V_t^F\frac{\|E(\xi)\|_2}{\sqrt{N}}\right| = O_p\left(\sqrt{\Delta_n} + \sqrt{\frac{\log N}{T}} + \frac{1}{N^{(1-\delta)/2}}\right).$$

Theorem 2 guarantees that, up to a scale factor, the latent common factor in the MVF model can be consistently estimated by conducting PCA on the stock RV. Comparing Theorems 1 and 2, we see that the PC has an additional error term of order $\sqrt{(\log N)/T}$ than the CRV. This additional error term is due to the error in estimating the principal eigenvector. The consistency of the PC also requires stronger conditions on the moment and the time span T. Similar to the consistency of ExpClogRV, one can show that PCA consistency holds when it is applied to log(RV)'s, and it has the same convergence rate as PCA on RVs. However, PCA on log(RV)'s also requires stronger conditions on V_{it} and Δ_n. Besides the sample covariance matrix, robust estimators such as the Huber estimator (Fan et al., 2018) also apply and need less a stringent condition on T.

Remark 4. Under the high-dimensional setting, PCA is consistent in identifying factor structure in returns; see, for example, Bai and Ng (2002), Bai (2003), Fan et al. (2013), Aït-Sahalia and Xiu (2017), and Ding et al. (2021). In this paper, we focus on identifying the factor structure in variances, which are not observable and have to be estimated. Estimated variances are subject to estimation errors, which accumulate as the dimension increases. Proving PCA consistency based on estimated variances requires non-trivial arguments. Our theoretical results reveal important insights and give explicit conditions for the PCA consistency of using high-frequency data

Table 3
Summary statistics of the lag-1 auto-correlations of $(\hat{\xi}_{it})$ across different stocks. The reported values are the 25% quantile (Q1), median, mean and the 75% quantile (Q3) of the auto-correlations across the stocks under consideration.

	Q1	Median	Mean	Q3
$\mathrm{corr}(\hat{\xi}_{it}, \hat{\xi}_{i\,t-1})$	0.536	0.596	0.616	0.690

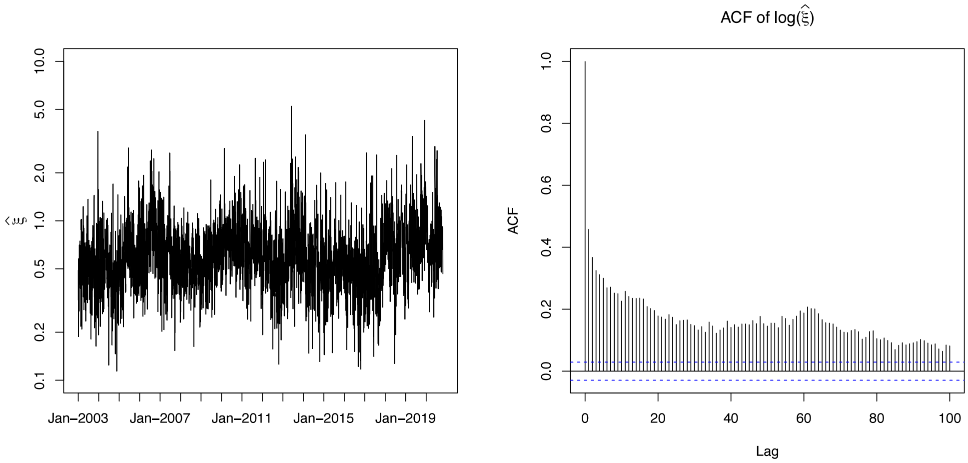

Fig. 7. Time series plot of idiosyncratic volatility exposure and its autocorrelations for a typical stock.

to identify latent factor structure in unobserved integrated variances. The term $\sqrt{\Delta_n}$ in Theorems 1 and 2 comes from estimating V_{it} with RV_{it}. If microstructure noise is present in the data, then the term $\sqrt{\Delta_n}$ will be $\Delta_n^{1/6}$ if one uses TSRV (Zhang et al., 2005) or $\Delta_n^{1/4}$ for MSRV (Zhang, 2006), PAV (Jacod et al., 2009 and Jacod et al., 2019) or QMLE (Xiu, 2010).

When a factor structure exists in idiosyncratic variances, theoretical results in Appendix A.1 of the Supplementary Materials (Ding et al., 2025) show the consistency of conducting PCA on realized idiosyncratic variances in identifying factor structure in integrated idiosyncratic variances.

4. Volatility forecasting

In this section, we utilize the proposed MVF model (1.1) for volatility forecasting.

4.1. Forecasting models

4.1.1. Our approach: MVF model

We use CRV_t as the estimator[5] of the latent factor in our proposed MVF model (1.1). The idiosyncratic realized variance exposure (iRE) The idiosyncratic realized variance exposure (iRE) is $\hat{\xi}_{it} = RV_{it}/\mathrm{CRV}_t$ for $i = 1, \ldots, N$. In Section 2, we show that the MVF model, in particular, the CRV captures the common movement in the stock volatilities well. The common realized variance inherits the long-memory dependence from the stock variances, so we predict CRV_t with a log HAR model.

The next question is how to model the idiosyncratic volatility exposure. To answer this question, we analyze the time-series dependence in the stock idiosyncratic realized variance exposures. In Table 3, we summarize the lag-1 auto-correlations of $(\hat{\xi}_{it})$ for different stocks. We see that the idiosyncratic exposures exhibit moderate auto-correlations. Furthermore, we find that the serial dependence appears to be long-memory; see Fig. 7 for the time series plot of the idiosyncratic realized variance exposures and its acf of a typical stock. Motivated by these observations, we model $\hat{\xi}_{it}$ as a log HAR as well.

To sum up, we model the $\log(\mathrm{CRV}_t)$ and $\log(\hat{\xi}_{it})$ separately using the HAR model (Eq. (6) in Corsi, 2009). The parameters in the log HAR model and the models in benchmark approaches presented in the next subsection are estimated with a 252-day rolling window. We denote the forecasts of CRV and $\hat{\xi}_i$ on day $t+1$ as $\widehat{\mathrm{CRV}}_{t+1}$ and $\hat{\hat{\xi}}_{i\,t+1}$, respectively. Then, the forecast of the variance of stock i on day $t+1$ is

$$\hat{V}_{i\,t+1} = \widehat{\mathrm{CRV}}_{t+1} \times \hat{\hat{\xi}}_{i\,t+1}, \quad i = 1, \ldots, N.$$

We denote the predictions by our proposed MVF model as $\mathrm{CRV}_{\mathrm{logHAR}} \times \mathrm{iRE}_{\mathrm{logHAR}}$.

[5] We also evaluated the performance of the MVF model by replacing CRV with the first PC in stock RVs as the variance factor estimator. This is a natural candidate as established in Theorem 2. However, potentially due to greater estimation error in estimating the principal eigenvector, PC underperforms CRV. We also observe that when using the market realized variance as the factor estimator, the forecasting results are significantly worse than the model with CRV.

In the above model, we fit log HAR to $\hat{\xi}_{it}$ stock-by-stock. Bollerslev et al. (2018) show that pooled fitting may help with prediction. However, we find that a pooled model for $\hat{\xi}_{it}$ does not improve the prediction performance. This could be due to that stocks have diverse dynamics in their $\hat{\xi}_{it}$'s, possibly because after extracting the common component, the dynamics of the remaining volatilities are different from one another.

4.1.2. Benchmark models

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

Cross-Sectional Benchmark Models

BM1: 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 $(\boldsymbol{U}_{t[j]})_{1\le j\le n}$ based on 5-min returns using a 252-day rolling window under the Fama–French three-factor model or a statistical factor model.[6] Given the empirical evidence in Kong and Liu (2018) and Kong et al. (2023), which showed that the factor space and factor loadings may change over time, we also tried shorter moving windows like one week to estimate β's. However, the results appear to be worse.

As to volatility forecasting, the idiosyncratic variance (iV) is predicted with a log HAR model. To forecast the factor covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{F}_{t+1}}$, BM1 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 et al., 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 by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{t+1}}$. Then, the forecast of the stock variance is

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

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 $\mathrm{FF3_{rG}{+}iV_{logHAR}}$ and the prediction under the statistical factor model as $\mathrm{StatsF_{rG}{+}iV_{logHAR}}$.

BM2: Return Factor Model + Common Idiosyncratic Volatility Modeling

This approach utilizes the return factor model as well as the factor structure in idiosyncratic variances. Same as the BM1 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 (Herskovic et al., 2016):

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

On each day t, $c_{0;i}$ and $c_{1;i}$ 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 log HAR model to predict the CiV factor. The residual of the factor model for idiosyncratic variance, ε_{it}, is modeled by AR(1), $\varepsilon_{it} = \phi_i\varepsilon_{i\,t-1} + u_{it}$.[7] We predict the idiosyncratic variance of stock i on day $t+1$ with $\widehat{iV}_{i\,t+1} = \hat{c}_{0;i(t)} + \hat{c}_{1;i(t)}\widehat{CiV}_{t+1} + \hat{\varepsilon}_{i\,t+1}$. The forecast of the stock variance is

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

We denote the prediction of BM2 under FF3 model as $\mathrm{FF3_{rG}{+}iV_{CiV}}$ and the prediction of BM3 under the statistical factor model as $\mathrm{StatsF_{rG}{+}iV_{CiV}}$.

BM3: Linear Volatility Model

This approach uses linear additive model (2.2) to predict stock variance with CRV_t being the factor. We predict CRV_t with the log HAR model and ε_{it} with AR(1).[8] We denote the predictions by this model as $\mathrm{Linear{:}CRV_{logHAR}}$.

BM4: Log-Linear Volatility Model

This approach uses log-linear additive model (1.2) to predict stock variance with $\mathrm{ClogRV}_t = \sum_{i=1}^{N}\log(RV_{it})/N$ being the factor. We predict the ClogRV_t and ε'_{it} with the HAR model. We denote the predictions by this model as $\mathrm{Log\text{-}Linear{:}ClogRV_{HAR}}$.

Time-series Benchmark Model

BM5: Individual Volatility Modeling

This approach predicts each stock volatility using a log HAR model.[9] Specifically, for each stock i, we fit $\log(RV_{it})$ with a HAR model. 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 $\mathrm{Idv_{logHAR}}$.

6 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.

7 Predicting ε's to be 0 or using HAR models leads to worse performance than AR(1), and the results are omitted.

8 Again, predicting ε's to be 0 or using HAR models leads to worse performance than AR(1), and the results are omitted.

9 We also evaluate the forecasting performance of the standard HAR model by fitting a HAR model directly on variance. We find that the log HAR model outperforms.

Table 4
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 consideration 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 consideration.

Forecasting models	Q1	Median	Mean	Q3
Cross-sectional benchmark models				
BM1: $\mathrm{FF3}_{\mathrm{rG}}$+$\mathrm{iV}_{\mathrm{logHAR}}$	0.127	0.136	0.142	0.150
BM1: $\mathrm{StatsF}_{\mathrm{rG}}$+$\mathrm{iV}_{\mathrm{logHAR}}$	0.129	0.137	0.144	0.150
BM2: $\mathrm{FF3}_{\mathrm{rG}}$+$\mathrm{iV}_{\mathrm{CiV}}$	0.134	0.143	0.159	0.165
BM2: $\mathrm{StatsF}_{\mathrm{rG}}$+$\mathrm{iV}_{\mathrm{CiV}}$	0.136	0.145	0.161	0.166
BM3: Linear:$\mathrm{CRV}_{\mathrm{logHAR}}$	0.206	0.226	0.240	0.255
BM4: Log-Linear:$\mathrm{ClogRV}_{\mathrm{HAR}}$	0.153	0.168	0.176	0.185
Time-series benchmark model				
BM5: $\mathrm{Idv}_{\mathrm{logHAR}}$	0.126	0.135	0.140	0.149
MVF model				
$\mathrm{CRV}_{\mathrm{logHAR}} \times \mathrm{iRE}_{\mathrm{logHAR}}$	**0.120**	**0.129**	**0.135**	**0.142**

4.2. Evaluation metrics

We evaluate the performance of different models in forecasting daily continuous variance. We use the same S&P 500 Index constituent 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]:

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

where $\widehat{V}_{it}$ is stock i's forecast variance on day t, RV_{it} is the realized variance, and L is the total length of the forecasting period. A smaller Q-like indicates a better volatility prediction. We use the same formulation of Q-like loss as (Bollerslev et al., 2016) so that the Q-likes for different stocks are standardized by the stocks' variances.

4.3. Forecasting results

In Table 4, 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, $\mathrm{CRV}_{\mathrm{logHAR}}\times\mathrm{iRE}_{\mathrm{logHAR}}$, outperforms the benchmark models in that its Q-likes have the lowest mean, median, 25%- and 75%- quantiles.

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:

$$\text{Out.perf.} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{QLIKE_{MVF,i}<QLIKE_{bm,i}\}}, \tag{4.1}$$

where $N = 291$ is the total number of stocks under consideration, 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 $e_{m,t} = \log(\widehat{V}_{m,t}/RV_t) + RV_t/\widehat{V}_{m,t} - 1$ with m equals 1 or 2 is 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:

$$\text{Sig.Out.perf.} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{p_i<\alpha\}}, \tag{4.2}$$

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 MVF model over the benchmark models and the significant outperformance proportion (Sig.Out.perf.) are reported in Table 5. The results show that the MVF model, $\mathrm{CRV}_{\mathrm{logHAR}} \times \mathrm{iRE}_{\mathrm{logHAR}}$,

[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 consideration.

Table 5
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 (Out.perf.(%)) or significantly outperforms (Sig.Out.perf. (%)) the benchmark models.

MVF Model: $CRV_{logHAR} \times iRE_{logHAR}$	Outperformance proportion	
vs. Benchmark models	Out.perf. (%)	Sig.Out.perf. (%)
BM1: $FF3_{rG}$+iV_{logHAR}	96.9	86.6
BM1: $StatsF_{rG}$+iV_{logHAR}	97.9	93.1
BM2: $FF3_{rG}$+iV_{CiV}	99.3	94.5
BM2: $StatsF_{rG}$+iV_{CiV}	99.0	96.6
BM3: Linear:CRV_{logHAR}	100.0	99.6
BM4: Log-Linear:$ClogRV_{HAR}$	99.7	98.6
BM5: Idv_{logHAR}	99.3	93.5

Table 6
Summary statistics of the utilities 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 consideration 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 utilities (in percentage) across the stocks under consideration.

Forecasting models	Q1	Median	Mean	Q3
Cross-sectional benchmark models				
BM1:$FF3_{rG}$+iV_{logHAR}	3.682	3.718	3.704	3.739
BM1:$StatsF_{rG}$+iV_{logHAR}	3.686	3.718	3.702	3.736
BM2:$FF3_{rG}$+iV_{CiV}	3.657	3.699	3.660	3.725
BM2:$StatsF_{rG}$+iV_{CiV}	3.648	3.702	3.658	3.723
BM3:Linear:CRV_{logHAR}	3.556	3.605	3.587	3.638
BM4:Log-Linear:$ClogRV_{HAR}$	3.653	3.690	3.676	3.715
Time-series benchmark model				
BM5: Idv_{logHAR}	3.683	3.719	3.705	3.740
MVF model				
$CRV_{logHAR} \times iRE_{logHAR}$	**3.696**	**3.730**	**3.718**	**3.752**

yields a more accurate prediction than the benchmark models for almost all the stocks. Moreover, the outperformance of our MVF model is statistically significant for a very high percentage of stocks under consideration.

The prediction results in Tables 4 and 5 have several implications. First, compared with models induced by return factor models, BM1: $FF3_{rG}$+iV_{logHAR}, BM1: $StatsF_{rG}$+iV_{logHAR}, BM2: $FF3_{rG}$+iV_{CiV} and $StatsF_{rG}$+iV_{CiV}, our model is more parsimonious and generates more accurate forecasts. This demonstrates that it is beneficial to work with volatilities directly than utilizing factor models for returns. Furthermore, compared with direct volatility factor models of alternative forms, the linear additive model (BM3: Linear:CRV_{logHAR}) and log-linear additive model (BM4: Log-Linear:$ClogRV_{HAR}$), our model is simpler and produces better forecasts. Finally, the outperformance of our proposed MVF model compared with the model that only uses the individual stock's time-series information, BM5: Idv_{logHAR}, highlights the benefit of utilizing cross-sectional information in individual stock volatility prediction.

4.4. Economic significance

To assess the economic gain of our proposed model in volatility prediction, we consider a utility-based framework. Specifically, we use the utility measure in (Bollerslev et al., 2018):

$$U = \frac{1}{L}\sum_{t=1}^{L}\left(8\%\sqrt{\frac{RV_{t+1}}{\hat{V}_{t+1}}} - 4\%\frac{RV_{t+1}}{\hat{V}_{t+1}}\right).$$

We summarize the utilities of various methods in Table 6. We see that our MVF model generates the highest utilities in mean, median, 25%- and 75%- quantiles.

We further compare the utilities stock by stock and test the significance in their differences between our MVF model and the benchmark models. We then compute the proportion of stocks for which our MVF model generates higher utilities than the benchmark models.

The outperformance proportion (Out.perf.) of our proposed MVF model over the benchmark models and the significant outperformance proportion (Sig.Out.perf.) in terms of the utility measure are reported in Table 7. We also report the proportion of outperformance with the difference greater than one basis point, at which level the improvement can be considered economically significant (Bollerslev et al., 2018). The results show that the MVF model generates higher utilities than the benchmark models for almost all stocks. Moreover, the outperformance of our MVF model is both economically and statistically significant for the vast majority of stocks under consideration.

Table 7
Outperformance proportion of the proposed MVF model over the benchmark models among the S&P Index constituent stocks in terms of the utility measure during January 2004–December 2020. The values are the percentages of stocks for which the MVF model outperforms (Out.perf.(%)), significantly outperforms (Sig.Out.perf. (%)) the benchmark models, or outperforms with difference greater than 0.01% (Diff ≥ 1bp (%)).

MVF Model: $CRV_{logHAR} \times iRE_{logHAR}$	Outperformance proportion		
vs. Benchmark models	Out.perf. (%)	Sig.Out.perf. (%)	Diff ≥ 1bp (%)
BM1: $FF3_{rG}$+iV_{logHAR}	93.8	75.6	61.2
BM1: $StatsF_{rG}$+iV_{logHAR}	96.2	77.7	64.6
BM2: $FF3_{rG}$+iV_{CiV}	98.3	88.3	92.8
BM2: $StatsF_{rG}$+iV_{CiV}	99.0	90.4	94.8
BM3: Linear:CRV_{logHAR}	99.3	98.3	99.0
BM4: Log-Linear:$ClogRV_{HAR}$	98.3	92.8	96.2
BM5: Idv_{logHAR}	99.0	90.7	72.2

Table 8
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 consideration.

Forecasting models	Q1	Median	Mean	Q3
BM3: Linear:CRV_{logHAR}	0.246	0.353	0.327	0.384
BM4: Log-Linear:$ClogRV_{HAR}$	0.194	0.239	0.227	0.262
BM5: Idv_{logHAR}	0.152	0.185	0.189	0.222
MVF: $CRV_{logHAR} \times iRE_{logHAR}$	**0.146**	**0.183**	**0.183**	**0.209**

Table 9
Summary statistics of the utilities 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 utilities across the indices under consideration.

Forecasting models	Q1	Median	Mean	Q3
BM3: Linear:CRV_{logHAR}	3.340	3.423	3.439	3.568
BM4: Log-Linear:$ClogRV_{HAR}$	3.497	3.560	3.567	3.640
BM5: Idv_{logHAR}	3.501	3.595	3.588	3.679
MVF: $CRV_{logHAR} \times iRE_{logHAR}$	3.550	3.607	3.607	3.697

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 using 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 estimators of number of factors (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 correlation of 0.988 with the first PC and can again 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 holds for the volatilities in the global market.[12]

We next evaluate volatility forecasting for the global indices. We forecast the volatility based on our proposed MVF model, $CRV_{logHAR} \times iRE_{logHAR}$, and compare the results with the benchmark models, BM3:Linear:CRV_{logHAR}, BM4:Log-Linear:$ClogRV_{HAR}$ and BM5:Idv_{logHAR}.[13] In Table 8, we summarize the Q-likes of these two models in forecasting the volatilities of the global equity indices. The results of utilities are summarized in Table 9.

Table 8 shows that our MVF model outperforms the benchmark models in that its Q-likes have the lowest mean, median, 25%- and 75%- quantiles. We also find that our MVF model outperforms for the majority of the indices we study. In Table 9, we see that the MVF also achieves higher utilities than the benchmark model. The results confirm the advantages of our MVF model in global market volatility forecasting.

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.

13 We use the RV of the Oxford-Man Institute's "realized library" to conduct the analysis. Due to lack of high-frequency data, we cannot implement BM1 and BM2.

6. Conclusion and discussion

We propose a multiplicative volatility factor (MVF) model for the cross-section of stock variances. The MVF model comprises a common variance factor and a multiplicative idiosyncratic component. The common factor is approximated by the cross-sectional average of stock variances. The MVF model is remarkably parsimonious, providing valuable insights into the co-movement of volatilities in the financial market. Furthermore, it leads to substantial improvement in volatility forecasting for both the US stocks and global equity indices.

Our proposed MVF model achieves significant dimension reduction and provides an effective distributional description for modeling volatility across a large cross-section of assets. Our framework can directly benefit related areas including volatility trading, risk management, large covariance matrix modeling and forecasting, and portfolio allocation. For example, thanks to the superior performance of our model in predicting individual volatilities, for large covariance matrix modeling and forecasting, one can put more focus on the modeling of the correlation matrix. Existing ideas from the constant conditional correlation model (Bollerslev, 1990), the dynamic conditional correlation model (Engle, 2002) and the factor DCC model (Engle, 2009) can come into the picture.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the Editor, Michael Jansson, the anonymous Associate Editor, and two anonymous referees for their constructive suggestions. We also thank the participants of SoFiE 2019 Annual Conference, and various seminars. Research was partially supported by National Science Foundation of China grants NSFC71922902 and NSFC72101226, Research Grant Council of Hong Kong SAR grants GRF16503419, GRF16304521, GRF16504223, SRFS2324-6H02 and T31-604/18-N, Science and Technology Development Fund of Macau SAR grant FDCT0057/2024/ITP2, and University of Macau grants SRG2023-00001-FBA, MYRG-GRG2024-00072-FBA and MYRG-GRG2023-00119-FBA.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.105959.

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81 (3), 1203–1227.
Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: Disentangling sources of bias at high frequency. J. Financ. Econ. 109 (1), 224–249.
Aït-Sahalia, Y., Mancini, L., 2008. Out of sample forecasts of quadratic variation. J. Econometrics 147 (1), 17–33.
Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. J. Econometrics 201 (2), 384–399.
Aleti, S., 2022. The High-Frequency Factor Zoo. Working paper.
Andersen, T.G., Benzoni, L., 2010. Do bonds span volatility risk in the US treasury market? A specification test for affine term structure models. J. Financ. 65 (2), 603–653.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Asai, M., McAleer, M., 2015. Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance. J. Econometrics 189 (2), 251–262.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–171.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.
Bandi, F.M., Fusari, N., Renò, R., 2023. 0dte Option Pricing. ESSEC Business School Research Paper, (2023-03).
Barigozzi, M., Hallin, M., 2016. Generalized dynamic factor models and volatilities: Recovering the market volatility shocks. Econom. J. 19 (1), C33–C60.
Barigozzi, M., Hallin, M., 2017. Generalized dynamic factor models and volatilities: Estimation and forecasting. J. Econometrics 201 (2), 307–321.
Barigozzi, M., Hallin, M., 2020. Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals. J. Econometrics 216 (1), 4–34.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2), 253–280.
Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. 2 (1), 1–37.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 (3), 307–327.
Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. Rev. Econ. Stat. 498–505.
Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: Modeling and managing volatility. Rev. Financ. Stud. 31 (7), 2729–2773.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. J. Econometrics 192 (1), 1–18.
Bollerslev, T., Todorov, V., 2011. Estimation of jump tails. Econometrica 79 (6), 1727–1783.
Calvet, L.E., Fisher, A.J., Thompson, S.B., 2006. Volatility comovement: A multifrequency approach. J. Econometrics 131 (1–2), 179–215.
Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. Econometrica: J. Econ. Soc. 135–155.
Connor, G., Korajczyk, R.A., Linton, O., 2006. The common and specific components of dynamic volatility. J. Econometrics 132 (1), 231–255.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7 (2), 174–196.
Diebold, F.X., Marino, R.S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13 (3).
Ding, Y., Engle, R., Li, Y., Zheng, X., 2025. Supplement to "Multiplicative factor model for volatility".
Ding, Y., Li, Y., Zheng, X., 2021. High dimensional minimum variance portfolio estimation under statistical factor models. J. Econometrics 222 (1), 502–515.
Ding, Y., Zheng, X., 2025. Sub-gaussian estimation of the scatter matrix in ultra-high dimensional elliptical factor models with $2+\varepsilon$th moment.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 391–407.
Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econom. Statist. 20 (3), 339–350.
Engle, R.F., 2009. High dimension dynamic correlations. In: The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford University Press, pp. 122–148.
Engle, R.F., Ito, T., Lin, W.-L., 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 58 (3), 525–542.
Engle, R., Martin, S., 2019. Measuring and hedging geopolitical risks.
Engle, R.F., Rangel, J.G., 2008. The spline-garch model for low-frequency volatility and its global macroeconomic causes. Rev. Financ. Stud. 21 (3), 1187–1222.
Engle, R.F., Sokalska, M.E., 2012. Forecasting intraday volatility in the US equity market. Multiplicative component garch. J. Financ. Econ. 10 (1), 54–83.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 603–680, With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva.
Fan, J., Liu, H., Wang, W., 2018. Large covariance estimation through elliptical factor models. Ann. Stat. 46 (4), 1383.
Fortin, A.-P., Gagliardini, P., Scaillet, O., 2023a. Eigenvalue tests for the number of latent factors in short panels. J. Financ. Econ. nbad024.
Fortin, A.-P., Gagliardini, P., Scaillet, O., 2023b. Latent factor analysis in short panels. Swiss Financ. Inst. Res. Pap. 23–44.
Gençay, R., Dacorogna, M., Muller, U.A., Pictet, O., Olsen, R., 2001. An Introduction To High-Frequency Finance. Elsevier.
Gonçalves, S., Meddahi, N., 2009. Bootstrapping realized volatility. Econometrica 77 (1), 283–306.
Hansen, P.R., Huang, Z., Shek, H.H., 2012. Realized garch: A joint model for returns and realized measures of volatility. J. Appl. Econometrics 27 (6), 877–906.
Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. J. Financ. Econ. 119 (2), 249–283.
Hounyo, U., Gonçalves, S., Meddahi, N., 2017. Bootstrapping pre-averaged realized volatility under market microstructure noise. Econometric Theory 33 (4), 791–838.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. J. Financ. 42 (2), 281–300.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.
Jacod, J., Li, Y., Zheng, X., 2019. Estimating the integrated volatility with tick observations. J. Econometrics 208 (1), 80–100.
Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for Stochastic differential equations. Ann. Probab. 26 (1), 267–307.
Kapadia, N., Linn, M., Paye, B., 2023. One vol to rule them all: Common volatility dynamics in factor returns. J. Financ. Quant. Anal. 1–48.
Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2013. Firm Volatility in Granular Networks. Technical report, National Bureau of Economic Research.
Kong, X.-B., Lin, J.-G., Liu, C., Liu, G.-Y., 2023. Discrepancy between global and local principal component analysis on large-panel high-frequency data. J. Amer. Statist. Assoc. 118 (542), 1333–1344.
Kong, X.-B., Liu, C., 2018. Testing against constant factor loading matrix with large panel high-frequency data. J. Econometrics 204 (2), 301–319.
Li, J., Liu, Y., Xiu, D., 2019. Efficient estimation of integrated volatility functionals via multiscale jackknife. Ann. Statist. 47 (1), 156–176.
Li, J., Todorov, V., Tauchen, G., 2016. Inference theory for volatility functional dependencies. J. Econometrics 193 (1), 17–34.
Li, J., Todorov, V., Tauchen, G., 2017. Jump regressions. Econometrica 85 (1), 173–195.
Li, J., Xiu, D., 2016. Generalized method of integrated moments for high-frequency data. Econometrica 84 (4), 1613–1633.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econometrics 187 (1), 293–311.
Luciani, M., Veredas, D., 2015. Estimating and forecasting large panels of volatilities with approximate dynamic factor models. J. Forecast. 34 (3), 163–176.
Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. 36 (2), 270–296.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica: J. Econ. Soc. 347–370.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometrics 160 (1), 246–256.
Renault, E., Van Der Heijden, T., Werker, B.J., 2023. Arbitrage pricing theory for idiosyncratic variance factors. J. Financ. Econ. 21 (5), 1403–1442.
Susmel, R., Engle, R.F., 1994. Hourly volatility spillovers between international equity markets. J. Int. Money Financ. 13 (1), 3–25.
Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. Time Ser. Anal.: Theory Pr. 1, 203–226.
Xia, N., Zheng, X., 2018. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. Ann. Statist. 46 (2), 500–525.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12 (6), 1019–1043.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100 (472), 1394–1411.
Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann. Statist. 39 (6), 3121–3151.