

Retraction

Retracted: Analysis of Factors Influencing Stock Market Volatility Based on GARCH-MIDAS Model

Complexity

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] D. Ma, T. Yang, L. Liu, and Y. He, "Analysis of Factors Influencing Stock Market Volatility Based on GARCH-MIDAS Model," *Complexity*, vol. 2022, Article ID 6176451, 10 pages, 2022.

Research Article

Analysis of Factors Influencing Stock Market Volatility Based on GARCH-MIDAS Model

Dan Ma ¹, Tianxing Yang ², Liping Liu ^{3,4} and Yi He ⁵

¹School of Statistics, Southwestern University of Finance and Economics, Chengdu 610071, China

²School of Big Data Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China

³School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China

⁴Guizhou Key Laboratory of Big Data Statistical Analysis, Guiyang 550025, China

⁵School of Accountancy, Guizhou University of Commerce, 550025 Guiyang, China

Correspondence should be addressed to Liping Liu; llping@mail.gufe.edu.cn

Received 31 October 2021; Revised 8 December 2021; Accepted 16 December 2021; Published 17 January 2022

Academic Editor: Wei Wang

Copyright © 2022 Dan Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper further extends the existing GARCH-MIDAS model to deal with the effect of microstructure noise in mixed frequency data. This paper has two highlights. First, according to the estimation of the long-term volatility components of the GARCH-MIDAS model, rAVGRV is adopted to substitute for the RV estimator. rAVGRV uses the rich data sources in tick-by-tick data and significantly corrects the impact of the microstructure noise on volatility estimation. Second, besides introducing macroeconomic variables (i.e., macroeconomic consistency index (MCI), deposits in financial institutions (DFI), industrial value-added (IVA), and M2), Chinese Economic Policy Uncertainty (CEPU) index and Infectious Disease Equity Market Volatility Tracker (EMV) are introduced in the long-run volatility component of the GARCH-MIDAS model. As indicated by the results of this paper, the rAVGRV-based GARCH-MIDAS is slightly better than the RV model-based GARCH-MIDAS. In addition to the common macroeconomic variables significantly impacting stock market volatility, CEPU also substantially impacts stock market volatility. Nevertheless, the effect of EMV on the stock market is insignificant.

1. Introduction

Traditional econometric models have been extensively used to analyze macroeconomic and financial consistent sampling frequency data. On the whole, the research methods using such data consist of VAR-type models, GARCH-type models, cointegration tests, and Granger causality tests. Most of the mentioned studies complied with low-frequency data models to examine the correlation between macroeconomics and stock market volatility. Over the past few years, among the studies on modeling problems of variables at different sampling frequencies, the Mixed Data Sampling (MIDAS) proposed by Ghysels et al. [1] has aroused the biggest attention. Such a model can develop a linear correlation between high-frequency explanatory variables and low-frequency explanatory variables, and it has been extensively applied in studies on macroeconomics, stock

market, and crude oil futures for its ability to fully draw upon available information. Based on the MIDAS regression model, Engel et al. [2] developed a GARCH-MIDAS model, decomposing volatility into long-term and short-term components. Their model is adopted to study the correlation between stock market volatility and macroeconomic variables. Subsequently, Asgharian et al. [3] examined the effect of U.S. macroeconomic variables on stock market volatility by adopting the GARCH-MIDAS model.

The reason why this model outperforms the conventional GARCH-class models is that it can decompose the total conditional variance of the conventional GARCH model into two parts, that is, short-term volatility at a high frequency captured by a GARCH process and long-term volatility at a low frequency. To calculate the sum of squares of intraday yield data, Andersen et al. [4] proposed the GARCH-MIDAS model with a long-run component based

on realized volatility (RV). For the RV estimator, most scholars exploited return data with a 5-minute sampling frequency to determine high-frequency realized volatility (Wang and Ghysels [5]; Conrad and Kleen [6]). Though intraday high-frequency data involves sufficient data information and can increase the estimation efficiency of stock volatility, it is difficult to estimate due to considerable data. Moreover, when high-frequency data are used to estimate the stock market volatility, prices are sampled at finer intervals, and microstructure issues turn out to be more pronounced.

RV proposed by Andersen et al. [4] is justified based on the assumption of a continuous stochastic process to meet the challenge from market microstructure noise in practical applications (Ait-Sahalia et al. [7]). Zhang et al. [8] proposed realized volatility through subsample averaging (rAVGRV) that exploited the abundant sources in tick-by-tick data; to a great extent, it could correct the effect of the microstructure noise on volatility estimation. As indicated by Liu et al. [9], rAVGRV is a more theoretically and empirically reliable estimator than RV.

When predicting financial market volatility, macroeconomic indicators are important (Andersen et al. [4]; Conrad and Loch [10]; Dorion [11]). The GARCH-MIDAS model has been the most popular model adopted to investigate the correlations between aggregate financial volatility and macroeconomic or financial variables (Conrad et al. [12]; Conrad et al. [13]; Pan et al. [14]; Su et al. [15]; Conrad and Kleen [6]; Opschoor et al. [16]; Dominicy and Vander Elst [17]; Lindblad [18]; Amendola et al. [19]; Conrad et al. [12]; and Borup and Jakobsen [20]).

The study is different from existing studies, and the long-run volatility component of the GARCH-MIDAS model is impacted by realized volatility and other explanatory variables. The explanatory variables here included the macroeconomic variables, that is, macroeconomic consistency index (MCI), deposits in financial institutions (DFI), industrial value-added (IVA), and M2, as well as Chinese Economic Policy Uncertainty (CEPU) index and Infectious Disease Equity Market Volatility Tracker (EMV). The reason for selecting CEPU and EMV variables is twofold. On the one hand, although China's stock market has been leaping forward over the past two decades, it is still emerging. It is not sufficiently mature to require the government to stabilize it by releasing and implementing necessary policies. The government's policies are overly frequent, and the constant modifications in policies increase internal and external uncertainties, thereby increasing stock market volatility. On the other hand, the coronavirus (COVID-19) outbreak in December 2019 has significantly affected global macroeconomy and financial markets. Intuitively, stock market reacts to such a pandemic more promptly and directly than other sectors in economic and financial system. Accordingly, the two mentioned variables should be included in this paper.

The paper further extends the existing studies, and the highlights focus on two aspects. (1) In the estimation of the long-term volatility components of the GARCH-MIDAS model, rAVGRV is used to replace the RV estimator.

rAVGRV uses the rich sources in tick-by-tick data and to a great extent corrects the effect of the microstructure noise on volatility estimation. Accordingly, the rAVGRV-based GARCH-MIDAS model should be able to characterize the volatility of the stock market more effectively. As a matter of fact, the study by Liu et al. [9] confirmed that rAVGRV exhibited a better performance than RV. (2) Besides introducing macroeconomic variables MCI, IVA, DFI, and M2, CEPU and EMV were also introduced in the long-run volatility component of the GARCH-MIDAS model. The Chinese government's policies are too frequent, and the constant modifications in policies increase internal and external uncertainties. Moreover, COVID-19 has imposed great burden on global macroeconomy and financial markets. For this reason, CEPU and EMV should be introduced.

The rest of the study is organized as follows. The second section elucidates the GARCH-MIDAS model. The third section refers to an empirical study that explores the estimation, forecasting the GARCH-MIDAS model built in the study at several levels. The fourth section presents the application of the model to the portfolio. The fifth section is the robustness analysis of this paper. The last section concludes the present study.

2. GARCH-MIDAS Model

In accordance with Campbell [21], the correlation between the variations of unanticipated and expected returns in the stock market can be set below:

$$\begin{aligned} r_{i,t} - E_{i-1,t}(r_{i,t}) \\ = (E_{i,t} - E_{i-1,t}) \sum_{j=0}^{\infty} \rho^j \Delta d_{i+j,t} - (E_{i,t} - E_{i-1,t}) \sum_{j=1}^{\infty} \rho^j \Delta d_{i+j,t}, \end{aligned} \quad (1)$$

where $r_{i,t}$ denotes the logarithmic stock return on day i of month t ; $d_{i,t}$ expresses the logarithmic dividend on day i of month t ; $d_{i,t}$ represents the discount factor; $E_{i-1,t}(\cdot)$ denotes the conditional expectation for a given set of information $I_{i-1,t}$ up to moment $i-1$.

Engle and Rangel [22] argued that unanticipated returns can be determined based on future cash flows or expected returns:

$$r_{i,t} - E_{i-1,t}(r_{i,t}) = \sqrt{\tau_t} g_{i,t} \varepsilon_{i,t}, \quad (2)$$

where the volatility consists of at least two components, and the volatility of stock returns falls into short-term $g_{i,t}$ and long-term components, where $g_{i,t}$ represents the volatility on day i of month t , and τ_t denotes the volatility at month t . Moreover, it is assumed that the random perturbation term $\varepsilon_{i,t}$ follows with the conditional standard normal distribution, that is, $\varepsilon_{i,t}|I_{i-1,t} \sim N(0, 1)$.

Thus, the conditional variance of stock returns is written as

$$\sigma_{i,t}^2 = E\left[(r_{i,t} - E_{i-1,t}(r_{i,t}))^2\right] = \tau_t g_{i,t}. \quad (3)$$

Assume that $E_{i-1,t}(r_{i,t}) = u$, so equation (2) can be written as

$$r_{i,t} = \mu + \sqrt{\tau_t} g_{i,t} \varepsilon_{i,t}. \quad (4)$$

For the short-term volatility component, it follows a mean-reverting unit-variance GJR-GARCH (1, 1) process:

$$g_{i,t} = (1 - \alpha - \beta - 0.5\gamma) + \left(\alpha + \gamma I_{\{r_{i-1,t} < 0\}} \right) \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t}, \quad (5)$$

where $I_{\{\cdot\}}$ is an indicator function, which means that the function takes a value of one if the condition is satisfied and zero otherwise. The short-run parameters are subject to $\alpha > 0$; $\beta \geq 0$; $\gamma \geq 0$; $\alpha + \beta + \gamma/2 < 1$. Parameter γ contains the information of asymmetry.

The long-term volatility component with a single explanatory variable takes the following form:

$$\log(\tau_t) = m + \theta \sum_{j=1}^K \varnothing_j(\omega_1, \omega_2) RV_{t-j}, \quad (6)$$

$$RV_t = \sum_{i=1}^{N_i} r_{i,t}^2,$$

where K denotes the number of periods over which the volatility is smoothed. If t represents a day, RV_t denotes daily realized volatility; if the sampling frequency of intraday high-frequency data is 5 min, the value of N_i is 48; if t represents a month, the monthly realized volatility is written as

$$RV_t = \frac{1}{22} \sum_{i=1}^{22} RV_{i,t}. \quad (7)$$

Compared with daily return data, intraday high-frequency data containing rich data information and realized volatility estimation based on high-frequency data can significantly increase the estimation efficiency of volatility, whereas the effect of market microstructural noise on realized volatility cannot be ignored. When noise is present, the estimator RV is biased, and applying it to the GARCH-MIDAS model will adversely affect the estimation of this model. To address the mentioned problem, this paper also considered applying the RV via sub-sample averaging (rAVGRV) proposed by Zhang et al. [8] to the GARCH-MIDAS model to substitute for the RV estimator. Thus, the single-factor GARCH-MIDAS model is expressed as

$$\log(\tau_t) = m + \theta \sum_{j=1}^K \varnothing_j(\omega_1, \omega_2) rAVGRV_{t-j}. \quad (8)$$

The rAVGRV estimator can effectively eliminate the effect of noise. rAVGRV is defined as follows.

Assume that, in period t , there are N equispaced returns $r_{i,t}$ and Δ is set to equal alignPeriod. For $i \geq \Delta$, the sub-sampled Δ -period return is defined as

$$\tilde{r}_{i,t} = \sum_{k=0}^{\Delta-1} r_{i-k,t}. \quad (9)$$

It is defined that $N^*(j) = N/\Delta$ if $j = 0$; otherwise, $N^*(j) = N/\Delta - 1$. The j -th component of the rAVGRV estimator is expressed by

$$RV_t^j = \sum_{i=1}^{N^*(j)} \tilde{r}_{j+i-\Delta}^2. \quad (10)$$

Take the average across the different RV_t^j , $j = 0, \dots, \Delta - 1$, and the rAVGRV estimator is defined.

When Y_t is the MIDAS term, the long-run component of the GARCH-MIDAS model is

$$\log(\tau_t) = m + \theta \sum_{j=1}^K \varnothing_j(\omega_1, \omega_2) Y_{t-j}, \quad (11)$$

where Y denotes the macroeconomic variable.

In the GARCH-MIDAS model expressed in equations (8) and (11), $\varnothing_j(\omega_1, \omega_2)$ is obtained from the weight function proposed by Ghysels et al. [1], and the equation is expressed as

$$\varnothing_j(\omega_1, \omega_2) = \frac{(j/K)^{\omega_1-1} (1-j/K)^{\omega_2-1}}{\sum_{i=1}^K (i/K)^{\omega_1-1} (1-i/K)^{\omega_2-1}}. \quad (12)$$

To ensure that the weights of the lagged variables are in a decaying form, $\omega_1 = 1$ is generally fixed. Thus, equation (12) can be defined as

$$\varnothing_j(\omega_2) = \frac{(1-j/K)^{\omega_2-1}}{\sum_{i=1}^K (1-i/K)^{\omega_2-1}}. \quad (13)$$

The single-factor GARCH-MIDAS model presented in the previous section considers only the rAVGRV volatility estimator or macroeconomic variable in the MIDAS term. However, numerous studies have shown that both realized volatility and macroeconomic variables have a significant impact on stock market volatility. With Y denoting the macroeconomic variable, as inspired by Engle et al. [2], equation (4) can be modified as

$$r_{i,t} = \mu + \sqrt{\tau(Y_t)} g_{i,t} \varepsilon_{i,t}. \quad (14)$$

As a result, the long-run volatility component expressed in equations (8) and (11) can be rewritten as

$$\log(\tau_t) = m + \theta_1 \sum_{j=1}^K \varnothing_j(\omega_{1,R}, \omega_{2,R}) rAVGRV_{t-j} + \theta_2 \sum_{j=1}^K \varnothing'_j(\omega_1, \omega_2) Y_{t-j}. \quad (15)$$

Equations (14) and (15) represent multifactor GARCH-MIDAS model.

3. Empirical Analysis

3.1. Data

3.1.1. Stock Market Data. This paper considers daily log-returns on the SSE Composite Index, calculated as $r_{i,t} = 100 * (\ln(p_{i,t}) - \ln(p_{i,t-1}))$, for the 2006:M1 to 2021:M6 period. To assess the volatility forecasts, this paper employed daily realized variances $RV_{i,t}$ and rAVGRV $_{i,t}$, where $RV_{i,t}$ is calculated from 5 min intraday log-returns. The data can be obtained from Wind database.

3.1.2. Explanatory Variables. Explanatory variables consist of macroeconomic consistency index (MCI), deposits in financial institutions (DFI), industrial value-added (IVA), and M2, as well as economic policy uncertainty index (EPUI) and Infectious Disease Equity Market Volatility Tracker (EMV). They are monthly data.

The monthly data of the Chinese Economic Policy Uncertainty (CEPU) index built by Huang et al. [23] are used. The index using 10 mainland Chinese newspapers can capture a wide range of uncertainty timely [24]. The Infectious Disease Equity Market Volatility Tracker (EMV) was built by Baker et al. [25]. Note that this paper aims to investigate whether and how infectious disease pandemic can affect the stock market volatility from a long-term perspective, instead of focusing on a single public health emergency, so the data from January 2006 to June 2021 are selected.

Table 1 reports the descriptive statistics of these time series. EMV is found with much larger standard deviation than those of stock indices. All the series have significant autocorrelation up to 10th lag, and they are not normally distributed.

The entire sample falls into two parts (i.e., estimation and forecast), in which the length of the estimation interval is from January 2006 to December 2020 (total 3647 days). The size of the forecast interval is from January 2021 to June 2021 (total 118 days). Both the daily closing rate data and the intraday high-frequency data are obtained from the RESSET database. Notably, when forecasting the volatility of SSE Composite Index, this paper uses a one-step forward rolling time window method. In other words, the first estimation interval $t = 1, 2, \dots, 3647$ is adopted to estimate the parameters of the GARCH-MIDAS model to determine the volatility value of SSE Composite Index, which is used as the volatility prediction value on day 3648. By keeping the length of the estimation interval constant, the estimated sample interval is shifted back one day, and the second estimation interval is $t = 2, 3, \dots, 3648$, in which the parameters of the GARCH-MIDAS model are estimated again, and the volatility of the 3649th day is predicted. Next, the volatility prediction of the 118th day is conducted.

3.2. In-Sample Performance

3.2.1. Analysis Based on Single-Factor GARCH-MIDAS Model. In the estimation of the GARCH-MIDAS model, the choice of weights w and lags K is of high significance. For the choice of weights, this paper follows the study by Engle et al. [2], in which the first weight is taken, and the second weight is chosen during the estimation of the model to ensure that the weights decrease with the increase in the number of lags.

K is the number of lags in MIDAS; since we use monthly data in the MIDAS equation, the lag order K can be taken as 12 according to Engle et al. [2].

The single-factor GARCH-MIDAS model considers only the rAVGRV (RV) estimator or macroeconomic variable in the MIDAS term. The estimation results of single-factor GARCH-MIDAS model are listed in Table 2.

From Table 2, the following conclusions are drawn: (1) besides macroeconomic variable MCI, macroeconomic variables IVA, M2, and DFI are significant, thereby demonstrating that they significantly impact the volatility of the stock market. (2) Chinese Economic Policy Uncertainty (CEPU) index significantly impacts stock market volatility. The government's policies are overly frequent, and the constant changes in policies increase internal and external uncertainties, thereby increasing stock market volatility. (3) Infectious Disease Equity Market Volatility Tracker (EMV) does not significantly impact the stock market, probably because timely actions by the Chinese authorities can reduce the volatility of their stock market, as also verified by Ali et al. [26] in the recent COVID-19 pandemic. (4) The coefficients θ corresponding to RV and rAVGRV are significant and are taken as positive values, which demonstrates that RV and rAVGRV can significantly improve the volatility of the Chinese stock market. Moreover, the loss functions MSE and QLIKE values of the GARCH-MIDAS (rAVGRV) model are smaller, which demonstrates that the model can be made better by using the rAVGRV estimator instead of the RV estimator in the GARCH-MIDAS model.

3.2.2. Analysis Based on Multifactor GARCH-MIDAS Model.

The multifactor GARCH-MIDAS model built with equations (14) and (15) is estimated using data within the sample interval, and the estimation results are listed in Table 3.

According to Table 3, (1) for all multifactor GARCH-MIDAS models, the rAVGRV estimator still significantly improves the Chinese stock market. (2) Consistent with the results of the single-factor GARCH-MIDAS model shown in Table 2, macroeconomic variables IVA, M2, and DFI significantly impact the volatility of the stock market. Chinese Economic Policy Uncertainty (CEPU) index significantly impacts stock market volatility. Infectious Disease Equity Market Volatility Tracker (EMV) insignificantly impacts the stock market.

Figure 1 illustrates the long-term components of stock market volatility of the GARCH-MIDAS model incorporating significant macroeconomic variables and CEPU, basically complying with the overall trend of the total conditional variance. Thus, the GARCH-MIDAS model incorporating macroeconomic variables and CEPU is suggested to have high goodness of fit.

TABLE 1: Descriptive statistics.

	RV _t	rAVGRV _t	MCI	IVA	M2	DFI	CEPU	EMV
Frequency	Monthly	Monthly	Monthly	Monthly	Monthly	Monthly	Monthly	Monthly
Mean	2.294	1.999	1.992	0.101	5.974	5.975	2.284	-0.294
Std	0.217	0.074	0.018	0.046	0.257	0.255	0.369	0.491
Skewness	2.582***	2.043***	-0.935***	0.532***	-0.384***	-0.355***	0.194***	2.259***
Kurtosis	11.113***	6.930***	4.310***	2.185***	1.894**	1.887**	2.210	8.708***
JB stat	693.668***	241.078***	39.153***	13.477***	13.593***	13.085***	5.810*	397.347***
Q (5)	232.73***	226.1***	517.89***	627.17***	814.5***	814.36***	556.5***	398.03***
Q (10)	322.64***	324.34***	623.98***	1045.6***	1493***	1494.6***	917.41***	486.07***

Notes: the Jarque-Bera statistic test for the null hypothesis of normality in sample returns distribution. Q (n) is the Ljung-Box statistics of the return series for up to nth order serial correlation. ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively.

TABLE 2: Estimation results of single-factor GARCH-MIDAS (K = 12).

	α	β	γ	m	θ	w_2	MSE	QLIKE
GARCH-MIDAS (RV)	0.0457*** (≤0.001)	0.951*** (≤0.001)	0.006 (0.606)	2.031*** (≤0.001)	0.434** (0.049)	1.328*** (0.002)	42.217	1.632
GARCH-MIDAS (rAVGRV)	0.045*** (≤0.001)	0.951*** (≤0.001)	0.005 0.642	2.013*** (≤0.001)	0.456* (0.084)	1.558*** (0.005)	41.659	1.601
GARCH-MIDAS (MCI)	0.065 (0.128)	0.933*** (0.966)	0.003 (0.966)	2.061 (0.341)	-16.070 (0.944)	7.553 (0.973)	42.197	1.628
GARCH-MIDAS (IVA)	0.062*** (≤0.001)	0.933*** (≤0.001)	0.008 (0.655)	2.147*** (≤0.001)	9.062*** (≤0.001)	4.965*** (≤0.001)	42.087	1.628
GARCH-MIDAS (M2)	0.064** (0.024)	0.933*** (≤0.001)	0.005 (0.782)	2.152*** (0.001)	1.296* (0.095)	2.312 (0.997)	42.189	1.629
GARCH-MIDAS (DFI)	0.063*** (≤0.001)	0.933*** (≤0.001)	0.005 (0.766)	2.095*** (≤0.001)	-1.598*** (≤0.001)	1.928*** (≤0.001)	42.190	1.629
GARCH-MIDAS (CEPU)	0.060*** (≤0.001)	0.935*** (≤0.001)	0.007 (0.667)	2.096*** (≤0.001)	-3.480** (0.019)	2.359*** (≤0.001)	42.206	1.626
GARCH-MIDAS (EMV)	0.064*** (≤0.001)	0.933*** (≤0.001)	0.005 (0.766)	2.081*** (≤0.001)	0.438 (0.448)	2.801*** (≤0.001)	42.235	1.629

Notes: the bracketed numbers are the p value of the estimations. ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively.

3.3. *Forecast Comparisons.* To assess the predictive performance exhibited by different models, the following loss functions are employed in the study:

$$\begin{aligned}
\text{MSE} &= \frac{1}{N} \sum_{t=1}^N (h_t - \hat{h}_t)^2, \\
\text{MAE} &= \frac{1}{N} \sum_{t=1}^N |h_t - \hat{h}_t|, \\
\text{MSD} &= \frac{1}{N} \sum_{t=1}^N \left(\sqrt{h_t} - \sqrt{\hat{h}_t} \right)^2, \\
\text{MAD} &= \frac{1}{N} \sum_{t=1}^N \left| \sqrt{h_t} - \sqrt{\hat{h}_t} \right|, \\
\text{QLIKE} &= \frac{1}{N} \sum_{t=1}^N \left| \frac{h_t}{\hat{h}_t} - \log \left(\frac{h_t}{\hat{h}_t} \right) \right|.
\end{aligned} \tag{16}$$

N in the loss function represents the length of the prediction interval, with $N=118$ days. h_t and \hat{h}_t denote the actual and predicted values of stock market volatility, respectively. Since the actual value of stock market volatility is

unobservable, as suggested by Pan et al. [27], an estimate of RV based on the 5 min frequency was used instead of h_t . A minor loss function indicates higher accuracy and better out-of-sample predictive power of the model. To verify whether the differences between the different prediction models are significant, the MCS proposed by Hansen et al. [28] is introduced for testing. The first step of the MCS test takes $M = M_0$, M_0 denotes the candidate model, and the significant level is set to α . If the null hypothesis is rejected, the worse-performing prediction model will be eliminated. The process continues till there is no more rejection of the null hypothesis to obtain the set of surviving models, which will be recorded as \hat{M}_α^* . The model contained in \hat{M}_α^* refers to the optimal prediction model at the $1 - \alpha$ confidence level. A condition for a model belonging to M is that its p value of the MCS test exceeds the significant level. In other words, the larger the p value of the prediction model is, the stronger the model's predictive power will be. Table 4 lists the results of the MCS tests based on different models.

The benchmark p value of the MCS test is set to 0.1. Given the principle of the MCS test, if the corresponding p value of the model is less than 0.10, the out-of-sample predictive ability of the model will be poor and will be rejected in the MCS test process. A larger p value reveals that the out-of-sample predictive ability of the model is better. As

TABLE 3: Estimation results of multifactor GARCH-MIDAS model ($K=12$).

	α	β	m	θ_1	θ_2	ω_1	ω_2	LLF
GARCH-MIDAS (rAVGRV + MCI)	0.070*** (≤ 0.001)	0.909*** (≤ 0.001)	-0.017 (0.998)	0.342*** (≤ 0.001)	0.142 (0.971)	1.858*** (≤ 0.001)	4.433 (0.917)	-5902.481
GARCH-MIDAS (rAVGRV + IVA)	0.069*** (0.002)	0.898*** (≤ 0.001)	-0.585* (0.083)	0.325*** (0.001)	7.403*** (0.006)	2.017*** (0.001)	1.001 (0.920)	-5894.395
GARCH-MIDAS (rAVGRV + M2)	0.068*** (≤ 0.001)	0.897*** (≤ 0.001)	6.035* (0.086)	0.306*** (≤ 0.001)	-0.972* (0.093)	2.292*** (≤ 0.001)	2.037 (0.386)	-5896.277
GARCH-MIDAS (rAVGRV + DFI)	0.067*** (≤ 0.001)	0.898*** (≤ 0.001)	7.458** (0.032)	0.289*** (≤ 0.001)	-1.207** (0.035)	2.182*** (≤ 0.001)	1.608 (0.354)	-5895.538
GARCH-MIDAS (rAVGRV + CEPU)	0.067*** (≤ 0.001)	0.906*** (≤ 0.001)	1.872*** (0.001)	0.332*** (≤ 0.001)	-0.724*** (0.003)	1.941*** (≤ 0.001)	38.808 (0.248)	-5894.433
GARCH-MIDAS (rAVGRV + EMV)	0.069*** (≤ 0.001)	0.909*** (≤ 0.001)	0.282 (0.528)	0.341*** (≤ 0.001)	0.084 (0.748)	1.891*** (≤ 0.001)	4.151 (0.763)	-5902.443

Notes: LLF indicates maximum likelihood function value. The bracketed numbers are the p value of the estimations. ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively.

indicated by the above Table, the p value of the GARCH-MIDAS model based on the rAVGRV statistic is also slightly larger than the p value of the GARCH-MIDAS model based on the RV statistic, and the ranking of the model is higher after a two-by-two comparison. Thus, the results above demonstrate that the GARCH-MIDAS (rAVGRV) model can be better than the GARCH-MIDAS (RV) model to some extent, since the rAVGRV statistic removes the effect of noise in the estimation, and the estimated realized volatility can be more accurate.

4. Application in the Portfolio

To verify the effectiveness of various types of volatility forecasting models in practice, they can be applied to a portfolio. It is assumed that the investor invests his money in equities and risk-free assets, respectively. In a standard mean-variance portfolio, the optimal weighting of an investor's investment in a stock is determined a priori based on the predicted variance. A volatility timing strategy popular in forecasting literature (Campbell and Thompson [29]; Ferreira and Santa-Clara [30]; Neely et al. [31]) is adopted in this paper. To be specific, at the end of day t , the investor calculates the optimal weight of the stock index according to the following equation for the next day $t+1$:

$$w_t = \frac{1}{\delta} \frac{\widehat{R}_{t+1} - R_{f,t}}{\widehat{h}_{t+1}}. \quad (17)$$

In the above equation, δ denotes the risk aversion coefficient, \widehat{R}_{t+1} represents the predicted value of stock returns that exceed the risk-free rate $R_{f,t}$, and here this paper selected the benchmark bank 1-year time deposit rate in place of the risk-free rate. \widehat{h}_{t+1} expresses the predicted value of stock market volatility. The weight of an investor's investment in equities is expressed as w_t , and the remainder weight $1 - w_t$ is assigned to the risk-free asset. Certainly, the optimal weight of stock is affected by the value of risk coefficient δ . For robustness check, four different δ 's of 5, 10, 15, and 20 are adopted.

Then the return of the portfolio is expressed as

$$R_{p,t+1} = w_t R_{t+1} + R_{f,t}. \quad (18)$$

To assess the portfolio performance, the measure of certainty equivalent return (CER) is adopted as follows:

$$\text{CER}_p = \mu_p - \frac{1}{2} \delta \sigma_p^2, \quad (19)$$

where μ_p and σ_p^2 denote the mean and variance of the portfolio returns, respectively. The CER values of the portfolios by using different volatility models are listed in the tables below.

Tables 5 and 6 list the annualized percentage values. (1) The economic value corresponding to the GARCH-MIDAS model significantly exceeds that of the GARCH model, so the GARCH-MIDAS model can have high performance in the portfolio, regardless of the risk aversion coefficient. (2) GARCH-MIDAS (rAVGRV) model is slightly better than GARCH-MIDAS (RV) model, and the application of the GARCH-MIDAS (rAVGRV) model to a portfolio can create a higher economic value.

5. Robustness Checks

To verify whether it is better to use rAVGRV instead of the RV estimator in the GARCH-MIDAS model, the GARCH-MIDAS-X model (Amendola et al. [24]; Engle and Patton [32]) is applied for further analysis. GARCH-MIDAS-X models are built for MCI, IVA, DFI, CEPU, EWV, and M2, respectively. RV or rAVGRV is included as a daily lagged variable in the short-run component (the so-called “-X” term). In this paper, the SSE Composite Index data from January 2006 to December 2020 are still used. The estimation results of the GARCH-MIDAS-X model are listed in Table 7.

As indicated by the results in Table 7, (1) for all GARCH-MIDAS-X models, the corresponding loss functions MSE and QLIKE are significantly smaller when the X term is the rAVGRV estimator, which demonstrates that the GARCH-MIDAS-X model built based on rAVGRV is better. (2) According to the parameter term z , when the X term is the rAVGRV estimator, it significantly impacts the Chinese stock market in most cases.

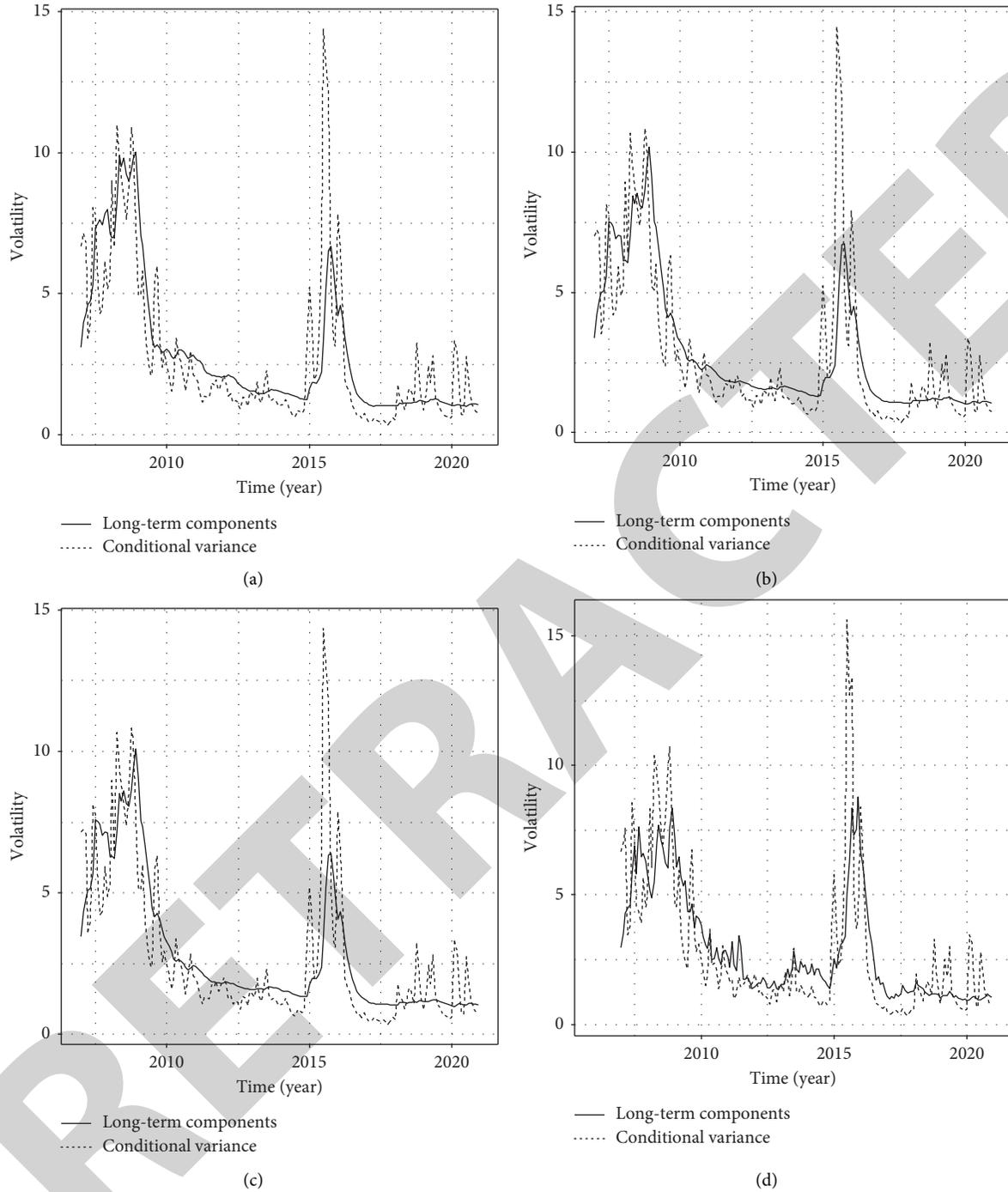


FIGURE 1: Fitting of conditional variance and long-run components of volatility for the multifactor GARCH-MIDAS model incorporating macroeconomic variables and CEPU. (a) $rAVGRV + IVA$. (b) $rAVGRV + M2$. (c) $rAVGRV + DFI$. (d) $rAVGRV + CEPU$.

To test the robustness of the research results in the previous section, CSI 300 index is also used as a proxy variable for the Chinese stock market. The selected data estimation interval remains from January 2006 to December 2020. Moreover, the estimation results are listed in Table 8.

According to Table 8, (1) the coefficients θ_1 corresponding to $rAVGRV$ are significant and are taken as a positive value, so $rAVGRV$ estimator can exert a significantly positive effect on the volatility of the Chinese stock market. (2) Variables IVA, M2, DFI, and CEPU still

TABLE 4: The MCS tests of GARCH-MIDAS model.

Loss function	MSE	MAE	MSD	MAD	QLIKE	Rank_M
GARCH-MIDAS (rAVGRV)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV)	0.845	0.793	0.845	0.762	0.813	2
GARCH-MIDAS (rAVGRV + MCI)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV + MCI)	0.743	0.803	0.790	0.746	0.802	2
GARCH-MIDAS (rAVGRV + IVA)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV + IVA)	0.751	0.698	0.821	0.792	0.880	2
GARCH-MIDAS (rAVGRV + M2)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV + M2)	0.797	0.813	0.796	0.808	0.779	2
GARCH-MIDAS (rAVGRV + DFI)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV + DFI)	0.862	0.787	0.794	0.838	0.745	2
GARCH-MIDAS (rAVGRV + CEPU)	1.000	1.000	1.000	1.000	1.000	1
GARCH-MIDAS (RV + CEPU)	0.812	0.799	0.852	0.884	0.802	2

Notes: numbers in the table indicate the p values of the MCS test based on different loss functions. Rank_M indicates the ranking of the model.

TABLE 5: Investment performance of different models based on RV.

δ	GARCH	GARCH-MIDAS (RV)	GARCH-MIDAS (RV + MCI)	GARCH-MIDAS (RV + IVA)	GARCH-MIDAS (RV + M2)	GARCH-MIDAS (RV + DFI)	GARCH-MIDAS (RV + CEPU)
5	1.545	1.670	1.801	1.790	1.840	1.805	1.796
10	1.329	1.503	1.664	1.595	1.699	1.686	1.599
15	1.260	1.332	1.493	1.320	1.582	1.480	1.436
20	1.065	1.252	1.325	1.295	1.399	1.373	1.311

TABLE 6: Investment performance of different models based on rAVGRV.

δ	GARCH-MIDAS (rAVGRV)	GARCH-MIDAS (rAVGRV + MCI)	GARCH-MIDAS (rAVGRV + IVA)	GARCH-MIDAS (rAVGRV + M2)	GARCH-MIDAS (rAVGRV + DFI)	GARCH-MIDAS (rAVGRV + CEPU)
5	1.789	1.880	1.822	1.900	1.873	1.815
10	1.585	1.662	1.718	1.826	1.800	1.770
15	1.330	1.465	1.604	1.684	1.612	1.592
20	1.112	1.340	1.482	1.526	1.493	1.391

TABLE 7: Estimation results of GARCH-MIDAS-X ($K=12$).

	X	β	z	m	θ	w_2	MSE	QLIKE
GARCH-MIDAS-X (MCI)	RV	0.800**	0.081	1.072***	-4.976***	2.263	52.488	2.02
	rAVGRV	0.928***	0.024**	1.033***	-20.493***	1.618	46.380	1.913
GARCH-MIDAS-X (IVA)	RV	0.808***	0.078	1.076***	7.569	9.130***	52.435	2.021
	rAVGRV	0.926***	0.027**	1.040***	8.172***	9.013***	46.800	1.931
GARCH-MIDAS-X (DFI)	RV	0.798***	0.081*	1.071***	0.264**	1.967	52.363	2.02
	rAVGRV	0.925***	0.025*	1.037***	0.825***	2.088	46.322	1.913
GARCH-MIDAS-X (CEPU)	RV	0.801***	0.08	1.0740***	-0.13	8.928***	52.361	2.021
	rAVGRV	0.930***	0.021	1.055***	-0.893**	5.674	45.488	1.878
GARCH-MIDAS-X (EMV)	RV	0.801***	0.08	1.068***	0.574	2.025**	52.189	2.021
	rAVGRV	0.930***	0.004*	1.462***	-0.632*	1.001**	42.907	1.689
GARCH-MIDAS-X (M2)	RV	0.797***	0.082	1.071***	1.710***	2.287***	52.759	2.024
	rAVGRV	0.927***	0.024**	0.950***	16.540*	18.302**	46.004	1.896

Notes: ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively. X represents RV or rAVGRV. z represents the coefficients corresponding to X term. Other parameters are consistent with Table 2.

TABLE 8: Estimation results of multifactor GARCH-MIDAS model based on CSI 300 ($K = 12$).

	α	β	γ	m	θ_1	θ_2	ω_1	ω_2	LLF
GARCH-MIDAS (rAVGRV + MCI)	0.065*** (≤ 0.001)	0.914*** (≤ 0.001)	0.016 (0.321)	-0.700 (0.650)	0.302*** (0.001)	0.562 (0.393)	1.827*** (≤ 0.001)	1.003 (0.782)	-6201.200
GARCH-MIDAS (rAVGRV + IVA)	0.064*** (0.001)	0.911*** (≤ 0.001)	0.015 (0.360)	-0.321 (0.251)	0.276*** (≤ 0.001)	7.175*** (0.004)	1.838*** (0.001)	6.663** (0.037)	-6193.656
GARCH-MIDAS (rAVGRV + M2)	0.061*** (≤ 0.001)	0.909*** (≤ 0.001)	0.018 (0.265)	9.231** (0.026)	0.217** (0.014)	-1.461** (0.030)	2.059*** (0.002)	2.576 (0.665)	-6194.839
GARCH-MIDAS (rAVGRV + DFI)	0.061*** (≤ 0.001)	0.908*** (≤ 0.001)	0.018 (0.261)	8.742** (0.031)	0.228*** (0.008)	-1.382** (0.036)	2.096*** (0.001)	4.214 (0.481)	-6194.75
GARCH-MIDAS (rAVGRV + CEPU)	0.061*** (≤ 0.001)	0.915*** (≤ 0.001)	0.017*** (0.006)	2.200*** (≤ 0.001)	0.277*** (≤ 0.001)	-0.780*** (0.001)	1.868*** (≤ 0.001)	56.843*** (0.002)	-6191.956
GARCH-MIDAS (rAVGRV + EMV)	0.064*** (0.001)	0.917*** (≤ 0.001)	0.015 (0.389)	0.534 (0.484)	0.282*** (0.003)	0.073 (0.803)	1.710** (0.048)	3.458 (0.684)	-6201.44

Notes: LLF indicates maximum likelihood function value. The bracketed numbers are the p value of the estimations. ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively.

significantly impact the volatility of the stock market, and the impact of EMV on the stock market remains insignificant. In brief, the conclusions drawn from Table 8 comply with Table 3. Thus, the findings of this paper are verified to be robust.

6. Conclusion

We further extend the existing GARCH-MIDAS model. This paper has two highlights. First, the rAVGRV estimator considering noise effects is adopted to estimate the long-term volatility components of the GARCH-MIDAS model. Second, in the GARCH-MIDAS model, the Infectious Disease Equity Market Volatility Tracker (EMV) and Chinese Economic Policy Uncertainty (CEPU) index are introduced besides macroeconomic variables to more comprehensively analyze the factors of Chinese stock market volatility based on the research in the study. Moreover, the following conclusions are drawn:

The GARCH-MIDAS (rAVGRV) model is slightly better than the GARCH-MIDAS (RV) model, since the effect of noise on the stock market in high-frequency data cannot be ignored. rAVGRV statistic removes the effect of noise in the estimation. As a result, the estimated realized volatility can be more accurate.

In single-factor GARCH-MIDAS model, the coefficients θ corresponding to RV and rAVGRV are significant and are taken as positive values, which demonstrates that RV and rAVGRV significantly improve the volatility of the Chinese stock market.

For all GARCH-MIDAS models, macroeconomic variables IVA, M2, and DFI significantly impact stock market volatility. Likewise, Chinese Economic Policy Uncertainty (CEPU) index impacts stock market volatility significantly, the government's policies are overly frequent, and the constant changes in policies cause more internal and external uncertainties, which increases stock market volatility. Besides, Infectious Disease Equity Market Volatility Tracker (EMV) insignificantly impacts the stock market, since timely actions by the Chinese authorities can reduce the volatility of their stock market, which is also verified by Amendola et al. [24] in the recent COVID-19 pandemic.

Data Availability

The stock data used in this article can be obtained from the Wind database. The macroeconomic consistency index (MCI), industrial value-added (IVA), M2, and deposits of financial institutions (DFI) can be obtained from the official website of the People's Bank of China (<https://www.pbc.gov.cn/diaochatongjisi/116219/index.html>) or the Oriental Fortune website (<https://data.eastmoney.com/cjsj/xfzxx.html>). The Chinese Economic Policy Uncertainty (CEPU) index (<https://economicpolicyuncertaintyinchina.weebly.com/>) was constructed by Huang et al. [23]. The Infectious Disease Equity Market Volatility Tracker (EMV) (http://www.policyuncertainty.com/infectious_EMV.html) was constructed by Baker et al. [25]. To save space, we will not show all the data in this article, but they can be provided upon request.

Conflicts of Interest

The authors solemnly declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was funded by the major project of the National Social Science Foundation of China, "Research on the path and measurement of digital empowerment of China's global value chain" (grant number 21&ZD149).

References

- [1] E. Ghysels, A. Sinko, and R. Valkanov, "MIDAS regressions: further results and new directions," *Econometric Reviews*, vol. 26, no. 1, pp. 53–90, 2007.
- [2] R. F. Engle, E. Ghysels, and B. Sohn, "Stock market volatility and macroeconomic fundamentals," *The Review of Economics and Statistics*, vol. 95, no. 3, pp. 776–797, 2013.
- [3] H. Asgharian, A. J. Hou, and F. Javed, "The importance of the macroeconomic variables in forecasting stock return variance: a GARCH-MIDAS approach," *Journal of Forecasting*, vol. 32, pp. 600–612, 2013.

- [4] T. G. Andersen, T. Bollerslev, and F. X. Diebold, "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.
- [5] F. Wang and E. Ghysels, "Econometric analysis of volatility component models," *Econometric Theory*, vol. 31, pp. 362–393, 2015.
- [6] C. Conrad and O. Kleen, "Two are better than one: volatility forecasting using multiplicative component GARCH-MIDAS models," *Journal of Applied Econometrics*, vol. 35, no. 1, pp. 19–45, 2019.
- [7] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang, "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, vol. 18, no. 2, pp. 351–416, 2005.
- [8] L. Zhang, P. A. Mykland, and Y. Ait-Sahalia, "A tale of two time scales: determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, vol. 100, pp. 1394–1411, 2005.
- [9] L. Y. Liu, A. J. Patton, and K. Sheppard, "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes," *Journal of Econometrics*, vol. 187, pp. 293–311, 2015.
- [10] C. Conrad and K. Loch, "Anticipating long-term stock market volatility," *Journal of Applied Econometrics*, vol. 30, pp. 1090–1114, 2015.
- [11] C. Dorion, "Option valuation with macro-finance variables," *Journal of Financial and Quantitative Analysis*, vol. 51, pp. 1359–1389, 2016.
- [12] C. Conrad, A. Custovic, and E. Ghysels, "Long- and short-term cryptocurrency volatility components: a GARCH-MIDAS analysis," *Journal of Risk and Financial Management*, vol. 11, p. 23, 2018.
- [13] C. Conrad, K. Loch, and D. Rittler, "On the macroeconomic determinants of long-term volatilities and correlations in U.S. Stock and crude oil markets," *Journal of Empirical Finance*, vol. 29, pp. 26–40, 2014.
- [14] Z. Pan, Y. Wang, C. Wu, and L. Yin, "Oil price volatility and macroeconomic fundamentals: a regime switching GARCH-MIDAS model," *Journal of Empirical Finance*, vol. 43, pp. 130–142, 2017.
- [15] Z. Su, T. Fang, and L. Yin, "The role of news-based implied volatility among US financial markets," *Economics Letters*, vol. 157, pp. 24–27, 2017.
- [16] A. Opschoor, D. V. Dijk, and M. V. D. Wel, "Predicting volatility and correlations with financial conditions indexes," *Journal of Empirical Finance*, vol. 29, pp. 435–447, 2014.
- [17] Y. Dominicy and H. Vander Elst, "Macro-driven VAR forecasts: from very high to very low-frequency data," 2015, <https://ssrn.com/abstract=2701256>.
- [18] A. Lindblad, *Sentiment Indicators and Macroeconomic Data as Drivers for Low-Frequency Stock Market Volatility*, Helsinki Center of Economic Research, Helsinki, Finland, 2017.
- [19] A. Amendola, V. Candila, and A. Scognamillo, "On the influence of US monetary policy on crude oil price volatility," *Empirical Economics*, vol. 52, pp. 155–178, 2017.
- [20] D. Borup and J. S. Jakobsen, "Capturing volatility persistence: a dynamically complete realized egarchmidas model," *Quantitative Finance*, vol. 19, no. 11, pp. 1839–1855, 2019.
- [21] J. Y. Campbell, "A variance decomposition for stock returns," *Economic Journal*, vol. 101, no. 405, pp. 157–179, 1991.
- [22] R. F. Engle and J. G. Rangel, "The spline-GARCH model for low-frequency volatility and its global macroeconomic causes," *Review of Financial Studies*, vol. 21, no. 3, pp. 1187–1222, 2008.
- [23] Y. Huang, P. Luk, and B. M. Fleisher, "Measuring economic policy uncertainty in China," *China Economic Review*, vol. 59, 2020.
- [24] A. Amendola, V. Candila, and G. M. Gallo, "Choosing the frequency of volatility components within the Double Asymmetric GARCH-MIDAS-X model," *Econometrics and Statistics*, vol. 20, pp. 12–28, 2021.
- [25] S. Baker, B. Nicholas, J. D. Steven, K. Kyle, S. Marco, and V. I. Tasaneeya, "The unprecedented stock market reaction to COVID-19," *Covid Economics: Vetted and Real-Time Papers*, vol. 1, no. 3, 2020.
- [26] M. Ali, N. Alam, and S. A. R. Rizvi, "Coronavirus (COVID-19) - an epidemic or pandemic for financial markets," *Journal of Behavioral and Experimental Finance*, vol. 27, Article ID 100341, 2020.
- [27] Z. Pan, R. Bu, L. Liu, and Y. Wang, "Macroeconomic fundamentals, jump dynamics and expected volatility," *Quantitative Finance*, vol. 20, no. 8, pp. 1345–1371, 2020.
- [28] P. R. Hansen, A. Lunde, and J. M. Nason, "The model confidence set," *Econometrica*, vol. 79, pp. 453–497, 2011.
- [29] J. Y. Campbell and S. B. Thompson, "Predicting excess stock returns out of sample: can anything beat the historical average?" *Review of Financial Studies*, vol. 21, no. 4, pp. 1509–1531, 2008.
- [30] M. A. Ferreira and P. Santa-Clara, "Forecasting stock market returns: the sum of the parts is more than the whole," *Journal of Financial Economics*, vol. 100, no. 3, pp. 514–537, 2011.
- [31] C. J. Neely, D. E. Rapach, J. Tu, and G. Zhou, "Forecasting the equity risk premium: the role of technical indicators," *Management Science*, vol. 60, no. 7, pp. 1772–1791, 2014.
- [32] R. F. Engle and A. J. Patton, "What good is a volatility model?" *Quantitative Finance*, vol. 1, pp. 237–245, 2001.