

# Forecasting Chinese stock market volatility under uncertainty

Wei Li\*

<sup>1</sup>School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

\*Corresponding author: LW97919@163.com

**Abstract.** As the trend of world economic integration intensifies, global uncertainty also has a certain impact on the stock markets of various countries. To explore the volatility of China's stock market under uncertainty, we use a mixed-frequency data model and introduce uncertainty variables, namely the Global Economic Policy Uncertainty Index (GEPU), the United States (US EPU) and China (China EPU) economic policy uncertainty indices, the implied volatility index (VIX) and the geopolitical risk index (GPR), to build an extended GARCH-MIDAS model to analyse the impact of uncertainty indices on Chinese stock market volatility. The empirical results show that, except for the US EPU index, all other uncertainty indices have some impact on the Chinese stock market, and the out-of-sample forecasting results indicate that the introduction of these variables improves the forecasting effect of the model, with China's economic policy uncertainty index showing significant advantages in forecasting both weekly and monthly volatilities.

**Keywords:** Chinese stock market; GARCH-MIDAS model; Uncertainty Index.

## 1. Introduction

The volatility of the stock market touches all aspects of life and accurate forecasting of stock market volatility plays a vital role in asset allocation, risk management and financial regulation. However, predicting stock market volatility is difficult because of its susceptibility to various uncertainties, such as interest rates, inflation, fiscal and policy uncertainties, which are all important sources of stock market volatility.

According to the existing literature, the study of volatility can be divided into two directions. The first is to innovate on existing models and strive to propose better extended models to improve the accuracy of volatility forecasting. For example, the generalized autoregressive conditional heteroskedasticity (GARCH) model by Bollerslev (1986)[1], the stochastic volatility (SV) model by Taylor (1986)[2] and the heterogeneous autoregressive HAR model constructed by Corsi (2009)[3] based on the heterogeneous market hypothesis.

Another direction in studying volatility is to find new factors or construct new variables in existing factors in the hope of extracting information that reflects volatility. Hossein Asgharian et al. (2013)[4] added seven macro variables such as inflation, short-term interest rates and unemployment to the GARCH-MIDAS model to compare these seven macroeconomic variables to compare the predictive effects of these seven macroeconomic variables. For example, Lei L et al. (2020)[5] argued that the prediction of the volatility of the Chinese stock market should not rely solely on information from the Chinese market, and that the components extracted from 28 international markets should more adequately reflect the volatility of the Chinese stock market, so the authors combined the HAR-RV model with principal component analysis and proposed the HAR-RV-PC model.

With the gradual opening of China's financial markets to the outside world, the Chinese stock market has become an integral part of the global financial market. Shocks in the international market will also inevitably increase the volatility and risk of the Chinese market. In fact, in addition to traditional indicators, recent studies have also mentioned uncertainty indices as an important factor affecting stock markets. Since Baker, Bloom and Davis (2016)[6] constructed the Global Economic Policy Uncertainty Index (GEPU), more and more scholars have been exploring the impact of economic uncertainty indices on stock market volatility, and Honghai Yu et al. (2018)[7] used the GARCH-MIDAS model to study the impact of the GEPU on the SSE index, arguing that GEPU plays an important role in predicting the volatility of the Chinese stock market. In addition to GEPU, there

is also the Global Geopolitical Risk Index (GPR), Mei Dexiang et al. (2020)[8] studied the impact of GPR on oil futures volatility, the authors concluded that short-term is has volatility has a positive relationship with GPR and that GPR helps to improve the accuracy of short-term oil futures volatility. The third is the implied volatility index. The implied volatility index (VIX) proposed by the CBOE Chicago Board Options Exchange is derived from option prices and includes not only historical volatility information, but also investors' expectations of future market conditions. In this paper, the above uncertainty indices will be used as variables added to the GARCH-MIDAS model to compare the impact of different uncertainty indices on Chinese stock market volatility and their predictive effects.

There are 2 main innovations in this paper. The first is to link the 4 uncertainty indices together and compare their usefulness in predicting stock markets; the second is to construct new global economic policy uncertainty index variables based on the uncertainty indices of 14 countries and regions, one simple average and one extracting their first principal component, to analyze whether the new variables will improve the accuracy of stock market volatility forecasts.

## 2. Models and methods

### 2.1 GARCH-MIDAS model

Drawing on Engle et al.'s study, this paper constructs a GARCH-MIDAS model incorporating exogenous variables to analyze the impact of economic uncertainty on stock market volatility. The stock market returns and volatilities are set as follows.

$$r_{it} = \mu_t + \sqrt{\tau_t g_{it}} \varepsilon_t, \varepsilon_t | \Phi_{i-1,t} \sim N(0,1) \quad (1)$$

$$\sigma_{it}^2 = \tau_t g_{i,t} \quad (2)$$

Where  $\tau_t$  is the long-term volatility component and  $g_{i,t}$  is the short-term volatility component. Short-term volatility obeys the GARCH(1,1) model. where  $\alpha > 0, \beta \geq 0, \alpha + \beta < 1$ . The long-term trend is influenced by the realised volatility (hereafter RV) and the economic uncertainty indicator X. The specific form of  $\tau_t$  is

$$\tau_t = m + \theta_1 \sum_{k=1}^K \phi_{1k}(\omega_{11}, \omega_{12}) RV_{t-k} \quad (3)$$

$$\tau_t = m + \theta_1 \sum_{k=1}^K \phi_{1k}(\omega_{11}, \omega_{12}) RV_{t-k} + \theta_2 \sum_{k=1}^K \phi_{2k}(\omega_{21}, \omega_{22}) X_{t-k} \quad (4)$$

$$RV = \sum_{i=1}^{N_t} r_{i,t}^2 \quad (5)$$

where K is the maximum lag order of the low-frequency variables, which is set to 6 in this paper by referring to Mei Dexiang et al. (2020), i.e. the lag period is chosen to be half a year. The coefficients  $\theta_1, \theta_2$  are the long-run impact coefficients of RV and X on volatility, respectively.  $\phi_k(\omega_1, \omega_2)$  are weighting functions for Beta lagged variables of the form :

$$\phi_k(\omega_1, \omega_2) = \frac{(k/K)^{\omega_1-1}(1-k/K)^{\omega_2-1}}{\sum_{j=1}^K (j/K)^{\omega_1-1}(1-j/K)^{\omega_2-1}} \quad (6)$$

### 2.2 Principal component analysis

Principal component analysis (PCA), as the name suggests, is the process of identifying the most dominant aspects in the data and replacing the raw data with the most dominant aspects in the data. Libing Fang et al. (2018)[9] used principal component analysis to extract three principal components of macro variables to forecast the US futures market, with the first and second principal components predicting significantly better results. Due to the complexity of the GARCH-MIDAS calculation, including EPU indices for multiple countries in one model may lead to convergence problems. Therefore, we use 2 variables at a time in the GARCH-MIDAS equation. In order to include information on the EPU of each country in the equation, the first principal component was extracted from the 17 country and regional EPU indices using principal component analysis.

$$\tau_t = m + \theta_1 \sum_{k=1}^K \phi_{1k}(\omega_{11}, \omega_{12}) RV_{t-k} + \theta_2 \sum_{k=1}^K \phi_{2k}(\omega_{21}, \omega_{22}) PC_{t-k} \quad (7)$$

### 2.3 DM test

The DM test studied by Diebold et al. (1995)[10] can determine whether the extended model mentioned above outperforms the baseline model in terms of prediction. The original hypothesis of the DM test is that the prediction error of the DM test is consistent with the baseline model. The DM statistic is expressed as follows.

$$DM_{model} = \frac{\frac{1}{T} \sum_{t=1}^T (L_{i,model,t} - L_{i,benchmark,t})}{Var(L_{i,model} - L_{i,benchmark})} \quad (8)$$

where T is the length of the out-of-sample prediction time,  $L_{i,model}$  is the sequence calculated by the comparison model based on the loss function, and  $L_{i,benchmark}$  is the loss sequence calculated by the benchmark model. If the DM statistic is negative, it means that the extended model has higher prediction accuracy than the benchmark model. There are two loss functions used in this paper, namely MAE (mean absolute error) and MSE (mean square error), and the loss function equations are as follows.

$$MAE = \frac{1}{T} \sum_{t=1}^T |\sigma_t^2 - \hat{\sigma}_t^2| \quad (9)$$

$$MSE = \frac{1}{T} \sum_{t=1}^T (\sigma_t^2 - \hat{\sigma}_t^2)^2 \quad (10)$$

where  $\sigma_t^2$  is the actual volatility and  $\hat{\sigma}_t^2$  is the model's volatility forecast.

### 2.4 MCS test

According to Hansen (2011)[11], the purpose of the MCS test is to select the models with good predictions from the original set of models  $M_0$  to form the set of models  $M^*$ . The set  $M^*$  can be expressed as

$$M^* \equiv \{u \in M_0 : E(d_{i,uv,t}) \leq 0 \text{ for all } v \in M_0\} \tag{11}$$

where  $d_{i,uv,t} \equiv L_{i,u,t} - L_{i,v,t}$ ,  $L_{i,u,t}$  and  $L_{i,v,t}$  are the loss values for models  $u$  and  $v$  at time  $t$  respectively for a given loss function. The MCS test is a simultaneous comparison of multiple models to determine the model with the significantly best predictive performance. There are a number of obvious advantages to the MCS test, firstly, the MCS test can reduce the effect of outliers in the data by using bootstrap methods, the second advantage is that there is no need to specify a baseline model, and thirdly, it does not value filter out a best model and the return is a collection of models. The original assumption of this test is that any two candidate models in the current model set have equal predictive power. There are various forms of test statistics, and in this paper the range statistic is chosen.

$$T_R = \max_{u,v \in M} \frac{|\bar{d}_{uv}|}{\sqrt{\widehat{var}(\bar{d}_{uv})}} \tag{12}$$

In this paper, we refer to the relevant literature and set the confidence level at 0.5. If the p-value of the MCS test is greater than 0.5 then the model is a surviving model and has a higher predictive power compared to other models.

### 3. Data analysis

The index that reflects the overall condition of the Chinese stock market better is the Shanghai Securities Composite Index(SSEC). The return data and intra-day high frequency data of the SSE index are chosen as the raw data. The time span is from January 4, 2002 to May 27, 2021. The log returns are obtained by taking first order differences on the stock prices, which are simpler to calculate through logarithmic transformation and the log returns have better statistical properties than the general returns.

The four uncertainty indices are the Global Economic Policy Uncertainty Index (GEPU), the Global Geopolitical Risk Index (GPR), the US Economic Policy Uncertainty Index (US EPU) and the Implied Volatility Index (VIX) proposed by the Chicago Stock Exchange in the U.S. The EPU and GPR indices are available from <http://www.policyuncertainty.com/> available.

**Table 1.** Descriptive statistics

	SSEC	GEPU	US EPU	China_EPU	VIX	GPR
Obs.	4514	233	233	233	233	233
Freq.	Day	Month	Month	Month	Month	Month
Mean	0.000174	138.89	123.92	244.26	19.51	108.87
Std. dev.	0.015	70.60	46.78	232.23	8.69	68.11
Skew.	-0.40	1.35	1.47	1.56	2.18	2.93
Kurt.	4.83	1.77	3.18	4.46	6.31	12.92
JB	4515.85***	100.96***	181.62***	116.09***	571.79***	1954.56***
ADF	-28.033***	-3.349*	-2.542	-3.729*	-3.86***	-3.249**
Q(5)	22.03***	828.26***	589.21***	929.68***	473.79***	339.69***
Q <sup>2</sup> (5)	551.64***	170.09***	721.99***	842.85***	334.67***	170.09***

As can be seen from Table 1, the mean of the SSEC returns is much smaller than the standard deviation and the stock market is less volatile compared to the uncertainty index. Of interest is that all stock market returns show a left bias, while all other factors show a right bias. With the exception of GEPU, all other data have a kurtosis of over 3, showing a spiky post-tailed distribution. For the JB test, it can be seen that all time series reject the normal distribution, and then looking at the smoothness test, the original hypothesis of non-smoothness is rejected for all series except US EPU. The LB statistic can be seen that the original hypothesis of non-autocorrelation is rejected for all series and all are autocorrelated.

## 4. Empirical analysis

### 4.1 In-sample estimation results

**Table 2.** GARCH-MIDAS model estimation results

	$\mu$	$\alpha$	$\beta$	$\theta_0$	$\theta_1$	$\omega_0$	$\omega_1$	m
RV	0.015	0.0609*	0.9290*	-		1.9986**		1.0597**
	6	**	**	0.0119***		*		*
RV+GEPU	0.017	0.0677*	0.9041*	0.0087***	-	1.0008**	6.4444	0.7977
	7	**	**		0.0034**			
					*			
RV+US EPU	0.007	0.0517*	0.9252*	0.0109***	-0.0021	1.0006**	2.9416	0.6388*
	1	**	**			*		
RV+China EPU	0.017	0.0604*	0.9289*	-	-	2.0081**	44.668	1.3848**
	5	**	**	0.0113***	0.0015**	*	6	*
					*			
RV+GPR	0.020	0.0563*	0.9367*	-0.0113	-0.0060**	3.6511	1.0007	2.0730**
	8	**	**					
RV+VIX	0.018	0.0865*	0.9030*	0.0003	0.0362**	2.8108	1.0004	0.5139
	3	**	**		*			
RV+PC1	0.008	0.0753*	0.9123*	0.0088***	-	1.0009**	5.4709	0.6329**
	0	**	**		0.1073**			
					*			

The results of the GARCH-MIDAS in-sample estimation are shown in Table 2. Firstly, it can be seen that the key coefficients  $\alpha$  and  $\beta$  are all significant and the sum of the two coefficients is essentially close to one, showing significant volatility and short-term agglomeration. For the coefficients  $\theta_0$ , all but two models, RV+GPR and RV+VIX, are significant, indicating that each month's realized volatility has a significant impact on the long-term volatility of our stock market. The coefficients reflect the impact of uncertainty indices on the volatility of the SSE. As can be seen from the table, all variables have a significant impact on our stock market, except for the US Economic Policy Uncertainty Index (US EPU). Below we then compare the relationship between long term volatility and conditional variance to determine the model's fit. Figure 3 shows that, except for the US EPU index and the geopolitical risk index, the total and long-term volatility trends estimated for the other variables are essentially the same, showing their relatively good fit, but it is also clear from this that different uncertainty indices have different effects on long-term volatility. What investors are most concerned about is future volatility, how much return they can have in the future and roughly what the risk is, so we need to make out-of-sample forecasts of this and analyze whether these factors can really improve the accuracy of volatility forecasts.

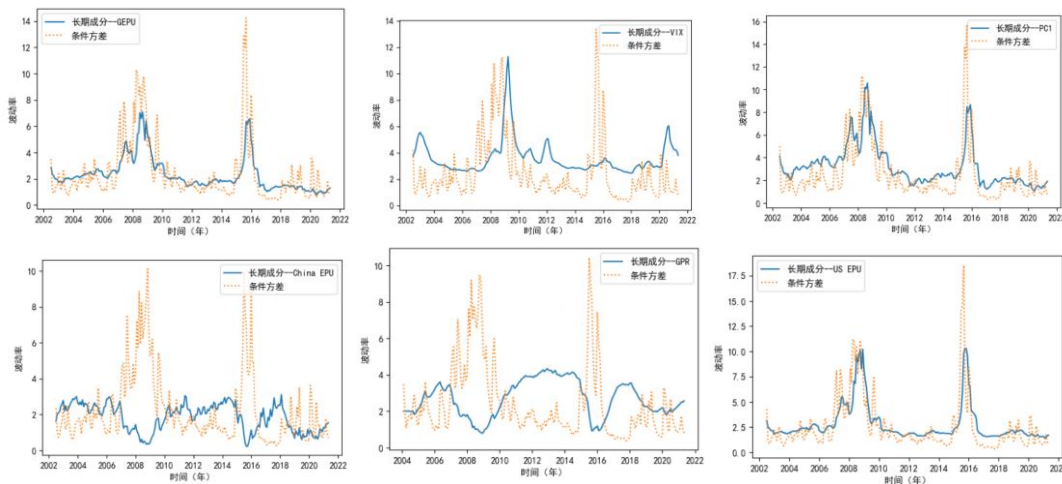


Fig. 1 Total and long-term volatility estimated by different models

### 4.2 Out-of-sample predictions

In the following we explore the ability of the GARCH-MIDAS model to predict volatility. The methodology uses a rolling window forecast with an estimation window of January 2002 to December 2017 and an out-of-sample forecast period of January 2018 to May 2021. Yaojie Zhang et al. (2020)[12] then construct a simple average of 22 international market volatilities through a HAR-Mean model, and the results show that the model can successfully predict the volatility of global equity markets. In this section, in addition to using the above six model forecasts and then constructing an index of global economic policy uncertainty calculated by simple averaging, a GARCH-MIDAS-RV-Mean model is proposed to investigate whether simple averaging can significantly improve the forecasting effectiveness of the model when compared with the benchmark model. DM tests are done on the out-of-sample forecasts of the eight models and the results are as follows.

Table 3. Out-of-sample predictive DM tests

	MAE		MSE	
	DM statistic	p-value	DM statistic	p-value
RV+GEPU	<b>-10.9008</b>	<b>0.0000</b>	<b>-9.2614</b>	<b>0.0000</b>
RV+US EPU	12.4665	0.0000	8.2823	0.0000
RV+China EPU	<b>-10.4987</b>	<b>0.0000</b>	-0.5706	0.5684
RV+VIX	<b>-3.5217</b>	<b>0.0005</b>	<b>-5.0591</b>	<b>0.0000</b>
RV+PC1	<b>-3.7749</b>	<b>0.0000</b>	<b>-5.5924</b>	<b>0.0000</b>
RV+Mean	<b>-3.9069</b>	<b>0.0001</b>	<b>-5.6303</b>	<b>0.0000</b>
RV+GPR	-1.1212	0.2625	<b>-2.8965</b>	<b>0.0038</b>

The DM test results in Table 3 are for the GARCH-MIDAS-RV model as the benchmark model. The bolded numbers in the table indicate that the DM statistic is significantly negative, suggesting better forecasting than the benchmark model. As can be seen from the table, RV+US EPU has a significantly positive DM statistic for both the MAE and MSE loss functions, indicating that the US Economic Policy Uncertainty Index hardly improves the predictive accuracy of Chinese stock market volatility. In contrast, GEPU and VIX outperform the benchmark model under both loss functions. Similarly, the newly constructed first principal component and simple average variables in this paper also significantly improve the model forecasts. Our economic policy uncertainty index and geopolitical risk index have only one loss function that outperforms the benchmark model in terms of prediction.

**Table 4.** Out-of-sample prediction MCS test

	1 DAY		5 DAY		22 DAY	
	MAE	MSE	MAE	MSE	MAE	MSE
RV	0.007	0.073	0.002	0.052	0.000	0.003
RV+GEPU	0.281	<b>1.000</b>	0.052	<b>1.000</b>	0.000	0.058
RV+US_EPU	0.006	0.073	0.002	0.052	0.000	0.012
RV+China EPU	<b>1.000</b>	<b>0.837</b>	<b>1.000</b>	<b>0.730</b>	<b>1.000</b>	<b>1.000</b>
RV+VIX	0.057	<b>0.837</b>	0.010	<b>0.711</b>	0.000	0.013
RV+PC1	0.027	<b>0.837</b>	0.006	<b>0.606</b>	0.000	0.012
RV+mean	0.028	<b>0.837</b>	0.006	<b>0.624</b>	0.000	0.012
RV+GPR	0.027	0.273	0.006	0.091	0.000	0.012

To further investigate the forecasting ability between the various comparative models, the MCS test is next applied to the models. Here, predictions of future fluctuations of one day, one week and one month are considered, corresponding to fluctuations of 1 day, 5 days and 11 days respectively. As shown in Table 4, p-values greater than 0.5 are bolded in the font. It can be seen that for forecasting 1-day volatility, the three models RV and RV+US EPU and RV+GPR do not have p-values above 0.5 under both loss functions, while the models incorporating GEPU, VIX, PC1 and EPU mean all only test for p-values above 0.5 under the MSE loss function, and it is worth noting that only the model incorporating the China EPU index performs better under both loss function. This suggests that the China EPU Index contains more information on China's stock market volatility than the other models. A similar result is shown in testing the predicted weekly volatility. In terms of predicting monthly volatility, it remains the case that our EPU index is the best predictor in terms of predicting our volatility.

## 5. Conclusion

In this paper, macroeconomic variables are introduced into the volatility model to study the volatility of the Chinese equity market through the lens of macro factors. Because of the different frequency of the data, the GARCH-MIDAS model is used for the economic policy uncertainty index and the geopolitical risk index for which only monthly data are available. This model allows for the decomposition of realised volatility into short-term and long-term components, and then speaks of the long-term component modelled with macro variables. The following conclusions are obtained from the empirical analysis.

First, there is volatility aggregation in the Chinese stock market, and the in-sample estimation results show that except for the US economic policy uncertainty index, all other uncertainty indices and the first principal component constructed from the EPU indices of 17 countries and regions have a significant impact on the volatility of the Chinese stock market.

Second, the results of the DM test for out-of-sample forecasts show that the addition of uncertainty indicators to the GARCH-MIDAS-RV model as the baseline model can all improve the forecasting ability of the model significantly to some extent.

Thirdly, and finally we did the MCS test, which showed that our EPU index showed significant strength in predicting daily, weekly and monthly volatility, while the GEPU, VIX, PC1 and the mean of the EPU index were only good predictors of daily and weekly volatility and underperformed in predicting monthly volatility. In summary, investors, policy makers and other market participants should pay attention to changes in uncertainty indices such as EPU, GPR and VIX in addition to traditional supply and demand fundamentals, speculation and interest rates when making investment, market regulation and purchase decisions.

## References

- [1] Bollerslev, T. 1986. Generalized Autoregressive Conditional Heterskedastici. *Journal of Econometrics*. 31: 307-27.
- [2] Taylor S J. 1986. *Modelling Financial Time Series* [M]. New York: John Wiley & Sons Ltd. 183-195
- [3] Corsi, F. . 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), p. 174-196.
- [4] Asgharian, H., Hou, A. J., & Javed, F.. 2013. The importance of the macroeconomic variables in forecasting stock return variance: a garch -midas approach. *journal of forecasting*, 32(7).
- [5] Likun Lei et al. 2020. Forecasting the volatility of Chinese stock market: An international volatility index. *International Journal of Finance & Economics*, 26(1), pp. 1336-1350. ; *Economics*, 26(1), pp. 1336-1350.
- [6] Baker, Scott R. and Bloom, Nicholas and Davis, Steven J.. 2016. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), pp. 1593-1636.
- [7] Honghai Yu and Libing Fang and Wencong Sun. 2018. Forecasting performance of global economic policy uncertainty for volatility of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 505pp. 931-940.
- [8] Mei Dexiang et al. 2020. Geopolitical risk uncertainty and oil future volatility: Evidence from MIDAS models. *Energy Economics*, 86(C),
- [9] Libing Fang and Honghai Yu and Wen Xiao. 2018. Forecasting gold futures market volatility using macroeconomic variables in the United States. *Economic Modelling*, 72pp. 249-259.
- [10] Diebold, F. X. , & Mariano, R. S. . 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 134-144.
- [11] Hansen, P. R. , Lunde, A. , & Nason, J. M. . 2011. the model confidence set. *econometrica*.
- [12] Yaojie Zhang and Feng Ma and Yin Liao. 2020. forecasting global equity market volatilities. *international Journal of Forecasting*, 36(4), pp. 1454- 1475.