International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 170-183

https://doi.org/10.70517/ijhsa46312

Wavelet transform-based characterization of anomalous data fluctuations during audit period capture study

Shibao Yang[1,*]
[1] Department of Economics, The Engineering & Technical College, Chengdu University of Technology, Leshan, Sichuan, 614000, China
Corresponding authors: (e-mail: 15228192887@163.com).

Abstract This paper denoises the collected stock market return data during the audit period based on wavelet transform and constructs a GARCH-MIDAS model to capture the volatility characteristics of the return data. The autoregressive conditional heteroskedasticity (ARCH) test, Q-test and ADF test are used to demonstrate the reasonableness of the GARCH-MIDAS model construction in this paper. The parameter estimation of the GARCH-MIDAS model is carried out using the great likelihood estimation method to further illustrate the validity of the model in this paper. The results show that the denoised yield data using wavelet transform is smoother than the original data, and can retain the main volatility characteristics in the original data, providing good data support for the subsequent capture of volatility characteristics. The macroeconomic and economic policy uncertainty variables are basically significant in the parameter estimation of one-factor and two-factor GARCH-MIDAS models, which can effectively reflect the overall and long-term volatility characteristics of yield data.

Index Terms wavelet transform, GARCH-MIDAS, ADF test, great likelihood estimation, volatility characteristics

I. Introduction

With the continuous expansion of enterprise scale and the increasing complexity of business, the amount of data and complexity of information faced by the audit has increased exponentially. Traditional auditing methods are inefficient and prone to human errors and omissions, and most of them are based on given financial data, and it is difficult to grasp the other volatile data of the enterprise, and it is also not easy to detect the anomalies that exist in them, and at the same time, there is a significant difference in the results of the sampling audits [1]-[3]. At the same time, the timeliness and accuracy of the enterprise audit requirements are also increasingly high, the need for audit can quickly find problems and provide effective solutions. And the development of digital technology has brought new opportunities for auditing. The emergence of big data, cloud computing, artificial intelligence and other technologies enables auditors to access and analyze data more quickly and comprehensively, and to discover abnormalities in audit data so as to perceive potential risks [4]-[6]. In the digital environment, audit data are no longer limited to paper documents and spreadsheets, but include a variety of structured and unstructured data, such as transaction records in databases, information on social media, and data generated by IoT devices [7]. However, with the continuous increase of audit data, the climbing of enterprise cash flow transaction data, the cyclical overlap of transaction data, and the complexity of transaction networks make the audit data show fluctuating characteristics, which improves the difficulty of uncovering data anomalies [8], [9]. And wavelet transform is a signal processing technology, which can decompose the signal into frequency components of different scales, so as to be able to better understand the characteristics of the signal [10]. Through wavelet transform analysis, users can obtain important information such as the frequency characteristics and time-frequency characteristics of the signal, which enables a deeper understanding of the intrinsic laws and characteristics of the signal, and thus can be used for image processing, feature extraction, and anomaly detection, etc. [11]-[13]. This provides a new paradigm for bill verification, data feature extraction and anomaly detection in auditing.

In this paper, we analyze the principle of wavelet transform and wavelet threshold denoising to explore the processing method of data with abnormal fluctuations during the audit period. In order to fully capture the characteristics of data fluctuations, this paper introduces the idea of frequency mixing on the basis of GARCH model, establishes GRACH-MIDAS model, and uses the great likelihood method to estimate the parameters of the model. The return data of the SSE 50 index from 2013 to 2024 are collected, and descriptive statistics are performed on the data, as well as testing the applicability of the data about the GARCH model. The wavelet transform method is utilized to reduce the noise of the collected data to ensure that the data are smoother under the premise of ensuring the original volatility characteristics, which facilitates the subsequent use of the processed data for further analysis. Parameter estimation using one-factor GARCH-MIDAS model and two-factor GRACH-MIDAS model, respectively, was conducted to explore the significance of the macroeconomic and economic policy uncertainty variables, and

the volatility of the macroeconomic factor and economic uncertainty factor, respectively, was estimated to better capture the volatility characteristics of the yield data. It provides a basis for decision-making to identify and deal with unusual volatility data during the audit period.

II. Wavelet denoising method

II. A.Wavelet transform

The wavelet transform is a transform analysis method commonly used in mathematics and engineering [14], which solves the problem of the Fourier transform in which the window size does not fit the frequency transform.

Let $\Phi(t)$ square be productible, i.e., $\Phi(t) \in L^2(R)$, if its Fourier transform $\hat{\Phi}(t)$ is satisfied:

$$C_\Phi = \int_R \frac{\left|\hat{\Phi}(t)\right|^2}{|t|} dt < \infty \tag{1}$$

Call $\Phi(t)$ a wavelet mother function or a fundamental wavelet, Eq. (1) is also known as the admissibility condition for the wavelet function. The wavelet mother function is obtained as a function under the action of the telescoping scale factor a and the translation factor b:

$$\Phi_{ab}(t) = a^{-\frac{1}{2}} \Phi\left(\frac{t-b}{a}\right), a, b \in R, a \neq 0 \tag{2}$$

The $\Phi_{ab}(t)$ is known as the continuous wavelet generated by $\Phi(t)$, i.e., the wavelet basis function. Expanding the signal $f(t)$ in $L^2(R)$ space under the wavelet basis yields the continuous wavelet transform of the signal:

$$WT_f(a,b) = \left\langle f(t), \Phi_{a,b}(t) \right\rangle = \frac{1}{\sqrt{a}} \int_R f(t) \Phi^*\left(\frac{t-b}{a}\right) dt \tag{3}$$

In practice, in order to prevent signal loss, it is common to discretize a, b so that the parameters $a = a_0^{-j}$ and $b = kb_0$, and the discrete wavelet basis function can be obtained as follows:

$$\Phi_{a_0^{-j}, kb_0(t)} = a_0^{-\frac{j}{2}} \Phi\left[a_0^j \left(t - kb_0\right)\right], j, k \in Z \tag{4}$$

The discrete wavelet transform is defined as:

$$WT_f\left(a_0^{-j}, kb_0\right) = \int_{-\infty}^{+\infty} f(t) \Phi^*_{a_0^{-j}, kb_0}(t) dt \tag{5}$$

II. B.Multi-discrimination analysis

Multi-discriminatory analysis decomposes the signal at different resolutions, and the resulting sum of the component signals at different scales is equal to the original signal; the decomposition algorithm is the Mallat algorithm. Multi-discriminative analysis only targets the low frequency part of the signal, not the high frequency part. Taking the three-layer decomposition map as an example (cA represents low frequency and cD represents high frequency), the decomposition relation of multidiscriminative analysis is: $S = cD1 + cD2 + cD3 + cA3$.

II. C.Wavelet thresholding denoising

The principle of wavelet threshold denoising is that by selecting a suitable wavelet basis function and the number of decomposition layers in advance, then the signal containing noise is decomposed by wavelet transform to obtain a set of wavelet coefficients containing important information about the original signal [15]. By setting a threshold, wavelet coefficients smaller than the threshold are considered to be generated by noise and are set to 0. Wavelet coefficients larger than the threshold are retained.

The implementation steps of wavelet threshold denoising are as follows:

(1) Choose an appropriate wavelet function and number of decomposition layers, and transform the original signal onto the wavelet domain by wavelet to obtain a set of wavelet decomposition coefficients.

(2) Determine a threshold value λ, the wavelet coefficients obtained from the decomposition are compared with the threshold value, when the wavelet coefficients are less than λ, it is considered that the coefficients consist of noise, so they are set to 0 and discarded. When the wavelet coefficient is greater than λ, the coefficient is retained.

(3) Perform signal reconstruction on the processed wavelet coefficients to obtain the denoised signal.

The thresholding method is divided into two kinds of hard thresholding and soft thresholding, in which the wavelet soft thresholding function is:

$$\hat{d}_{j,k}=\begin{cases}\operatorname{sgn}\left(d_{j,k}\right)\left(\left|d_{j,k}\right|-\lambda\right) & \left|d_{j,k}\right|>\lambda \\ 0 & \left|d_{j,k}\right|<\lambda\end{cases} \tag{6}$$

The wavelet hard threshold function is:

$$\hat{d}_{j,k}=\begin{cases}d_{j,k} & \left|d_{j,k}\right|>\lambda \\ 0 & \left|d_{j,k}\right|<\lambda\end{cases} \tag{7}$$

Where $d_{j,k}$ is the original wavelet coefficients, λ is the threshold, and $\hat{d}_{j,k}$ is the wavelet coefficients after thresholding.

In addition to the selection of the threshold function, it is also very important to determine the appropriate threshold value, and the following are four commonly used threshold selection methods.

(a) Fixed threshold (sqtwoog criterion)

i.e:

$$\lambda=\sigma\times\sqrt{\ln(N)} \tag{8}$$

where σ is the variance of the noise, N is the sampling length of the signal, $\sigma=median\left(\left|d_{j,k}\right|\right)/0.6745$, and the $median$ function is the median taken over the absolute value of all wavelet coefficients.

(b) Stein unbiased likelihood estimation threshold (rigrsure criterion)

Set up a vector D, the elements in the vector are the squared wavelet coefficients in ascending order, i.e., $D=\left[d_1^2,d_2^2,d_3^2,\cdots,d_n^2\right]$, $d_1^2<d_2^2<\cdots<d_n^2$, where $d_k,k=1,2,\cdots,n$ denotes the wavelet coefficients and n denotes the signal length.

Set up another risk vector $R=\left[r_1,r_2,r_3,\cdots,r_n\right]$. The k th element in R is expressed as: $r_k=\left[n-2k+(n-k)\times d_k^2+\sum_{j=1}^{k}d_j^2\right]/n$ to find the smallest element in R, $r_{\min}$ and the corresponding k -value, then find the corresponding d_k, and finally get the threshold value: $\lambda=\sqrt{d_k^2}$.

(c) Heuristic thresholding (heursure criterion)

It combines fixed thresholding and stein unbiased likelihood estimation thresholding by first defining two variables X and Y, $X=\left(\sum_{i=1}^{n}d_i^2-n\right)/n$, $Y=\sqrt{\left(\log_2 n\right)^3/n}$. The threshold selection rule is: when $X<Y$, take a fixed threshold. When $X>Y$, take the smaller of the fixed threshold and the threshold of stein unbiased likelihood estimation.

(d) Minimum-maximum criterion threshold (minimaxi criterion)

i.e:

$$\lambda=\begin{cases}0 & n\geq 32 \\ 0.3936+0.1829\log_2 n & n<32\end{cases} \tag{9}$$

III. GARCH-MIDAS models

III. A. The form of the GARCH family of models

The GARCH model is further extended from the ARCH model and can be regarded as an infinite-order ARCH model [16].The GARCH model assumes that volatility is time-varying and that changes in volatility can be explained by its own historical volatility and by external influences. Compared to the ARCH model, the GARCH model is able to better portray the long-run nature of volatility. $GARCH\left(p,q\right)$ The standard model theory form is as follows:

$$\begin{cases}r_t-E_{t-1}(r_t)=\sigma_t\varepsilon_t \\ \sigma_t^2=m+\sum_{i=1}^{p}\alpha_i\sigma_{t-i}^2\varepsilon_{t-i}^2+\sum_{j=1}^{q}\beta_j\sigma_{t-j}^2\end{cases} \tag{10}$$

where r_t represents the rate of return, which is typically converted to logarithmic form, $E_{t-1}(\cdot)$ represents the conditional expectation of the rate of return, σ_t represents the variance, ε_t is the new interest term, and α_i and β_j are coefficients, which measure the $t-1$-period information on the magnitude of the effect of the t-period information on the t-moment, and it is clear that $\alpha_i > 0$ and $\beta_j > 0$. In empirical applications, the $GARCH(1,1)$ model usually has a better fitting effect, so it is widely used in finance, economy and other fields.

In order to better study the correlation nature of financial time series, the GARCH model can be extended, and the class of models based on the extension of the GARCH model is collectively known as the GARCH family, and the common ones are GARCH-M, IGARCH, EGARCH, and so on. The GARCH-MIDAS model discussed in this paper improves the effectiveness of the estimation of each parameter of the model by introducing the idea of mixing into the GARCH model.

III. B. Setup of the standard GARCH-MIDAS model

The GARCH-MIDAS model in this paper extracts two components of volatility, a short-run volatility component, which is generally regarded as obeying a $GARCH(1,1)$ model, and a long-run volatility component, which uses a Beta-weighting scheme to estimate low-frequency explanatory variables. The impact of macroeconomic or financial information on market volatility is generally analyzed using a log-linear model as the main analytical framework. The daily logarithmic stock market return $r_{i,t}$ in the t th period of a time period (e.g., month, quarter) is expressed as follows:

$$r_{i,t} - E[r_{i,t} \mid \Phi_{i-1,t}] = \sqrt{\tau_t \cdot g_{i,t}}\,\varepsilon_{i,t} \tag{11}$$

where $E[r_{i,t} \mid \Phi_{i-1,t}]$ denotes the conditional expectation of the set of all information as of the $i-1$ th trading day in a given t period. And assuming that the expected return is a constant μ and $\varepsilon_{i,t} \mid \Phi_{i-1,t}$ obeys a standard normal distribution $N(0,1)$, under these two assumptions, taking the conditional variance on both sides reveals that there is $\sigma_{i,t}^2 = \tau_t g_{i,t}$. Equation (10) shows that the volatility of the stock market has two components: a short-term volatility component $g_{i,t}$ and a long-term volatility component τ_t, with the short-term volatility component being related to short-term influences such as stock liquidity and the long-term component being related to macroeconomic variables.

Disregarding the asymmetry of short-term volatility, i.e., short-term volatility reacts differently in magnitude to good and bad news, the general form of the short-term volatility component is given as follows:

$$g_{i,t} = (1-\alpha-\beta) + \alpha(r_{i-1,t} - \mu)^2 / \tau_i + \beta g_{i-1,t} \tag{12}$$

Considering the asymmetric mean-reversion model, i.e., the impact of macroeconomic variables on short-term volatility is generally asymmetric, in order to improve the accuracy of model estimation, more in line with the subsequent study of leverage, this paper introduces the asymmetric effect into the short-term volatility component, adopting the assumptions of the GARCH-MIDAS-A model, assuming that the short-term volatility component $g_{i,t}$ obeys the $GARCH(1,1)$ model The modification of the above equation is given:

$$\begin{aligned} g_{i,t} &= \left(1 - a - \beta - \gamma/2\right) \\ &\quad + \left(a + \gamma \cdot I_{\{r_{i-1,t}-\mu<0\}}\right)\frac{\left(r_{i-1,t}-\mu\right)^2}{\tau_i} + \beta g_{i-1,t} \end{aligned} \tag{13}$$

The model requires $\alpha > 0$, $\beta > 0$ and $\alpha + \beta + \gamma/2 < 1$, and I is an indicative function, i.e., it takes 1 when the condition is satisfied and 0 when it is not, which is set up in such a way as to allow asymmetry to be reflected by the shift in returns from positive to negative. Obviously, the model guarantees that the expectation of the short-term volatility component is 1, i.e., $E\left[g_{i,t}\right] = 1$, and the parameter γ contains the asymmetric information.

Current research on GARCH-MIDAS models focuses on the portrayal of the long-run volatility component, which is usually characterized by a Beta weight function for lagged macroeconomic variables. In this paper, a general form of long-run volatility component equation is used for the weight construction:

$$\log(\tau_t) = m + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) X_{t-k} \tag{14}$$

The purpose of adopting $\log(\tau_t)$ instead of τ_t is to ensure that the long-term volatility is constantly positive, m is a constant term, K is the lagged order of the macroeconomic variable, and $\varphi_k(\omega_1,\omega_2)$ is the weight function corresponding to the macroeconomic variable, which is determined by three parameters ω_1,ω_2,k determined by the specific expression:

$$\varphi_k(\omega_1,\omega_2)=\frac{\left(k/(K+1)\right)^{\omega_1-1}\cdot\left(1-k/(K+1)\right)^{\omega_2-1}}{\sum_{l=1}^{K}\left(l/(K+1)\right)^{\omega_1-1}\left(1-l/(K+1)\right)^{\omega_2-1}} \tag{15}$$

Clearly, for k in the range, there is (ω_1,ω_2).

Beta weight functions can produce attenuated, hump-shaped or U-shaped weight distributions. The unrestricted weight function can produce both decaying and hump-shaped weight distributions, while the restricted weight function can only produce decaying weight distributions, and the rate of decay is determined by the parameter ω_2, and the larger ω_2 is, the faster the rate of decay is. The larger ω_2 is, the faster the decay rate is. The smaller ω_2 is, the slower the decay rate is.

Lagged macroeconomic variables are weighted by using a weighting function. Different choices of weighting functions affect the role of macroeconomic information in the long-run volatility component, where the parameter (ω_1,ω_2) is used to measure the extent of that effect. Since different weighting functions assign different weights to the lagged term, this can lead to differences in the estimated value of (ω_1,ω_2) and the predictive effect of macroeconomic variables on long-term market volatility. This provides a basis for this paper to study the volatility of mixed-frequency models and macroeconomic variables.

III. C. Model estimation methods

In this paper, we use the great likelihood estimation method to estimate the parameters of the GARCH-MIDAS model, and denote the set of parameters by Ω, which can be obtained by combining Eq. (13) as well as Eq. (14) in the previous section:

$$\begin{aligned} g_t(\Phi)&=\left(1-\alpha-\beta-\gamma/2\right)\\ &\quad+\left(a+\gamma\cdot I_{\{r_{t-1,t}-\mu<0\}}\right)\frac{\left(r_{t-1,t}-\mu\right)^2}{\tau_t(\Phi)}+\beta g_{t-1}(\Phi)\end{aligned} \tag{16}$$

$$\log\left[\tau_t(\Phi)\right]=m+\theta\sum_{k=1}^{K}\varphi_k(\omega_1,\omega_2)X_{t-k} \tag{17}$$

where $\Phi=(\alpha,\beta,\gamma,m,\theta,\omega_1,\omega_2)\subset\Omega$, the logarithmic form of the parameter estimates of the standard GARCH-MIDAS model is derived by quasi-great-likelihood estimation as follows, and there are as many as 8 parameters to be estimated for the overall model, which is based on the distributional function and model setting, the sequences of long- and short-term volatility components $\hat{\tau}_t$ and $\hat{g}_{i,t}$ estimated by the mixed-frequency model are obtained by quasi-great likelihood estimation:

$$LLF=-\frac{1}{2}\sum_{i=1}^{T}\left[\log g_t(\Phi)\tau_t(\Phi)-\frac{\left(r_t-\mu\right)^2}{g_t(\Phi)\tau_t(\Phi)}\right] \tag{18}$$

IV. Empirical analysis

IV. A. Selection of Indicators and Descriptive Statistical Analysis

In order to verify the effectiveness of the method proposed in this paper, after examining many stock market data such as GEM and NSE, and considering factors such as data quality and sample size, it was finally decided to use the data of SSE 50 index as the research sample, which is one of the most famous stock indices in China, and it is a sample of 50 leading listed companies in Shanghai, covering a wide range of industries and sectors, and it is the Chinese stock market's representative index. Secondly, the data quality and reliability of SSE 50 Index are unquestionable, with the advantages of strong representativeness and high data reliability. Therefore, studying the volatility (RV) of SSE 50 index can better understand the characteristics and behaviors of the financial market and provide more accurate information and suggestions.

Macroeconomic indicators considered for inclusion in the GARCH-MIDAS model are China's quarterly GDP, inflation rate CPI (YoY). The economic policy uncertainty aspect contains China's Geopolitical Risk Index (GPR), Economic Policy Uncertainty Index (EPU), Infectious Disease Pandemic Index (EMV-ID) and Trade Policy

Uncertainty Index (TPU). Table 1 shows the descriptive statistics of the returns and the indicators included in the model for the SSE 50 index data. In the table, DB and DP denote skewness and kurtosis, respectively, J-B is the Jarque-Bera's normal distribution test, and ***, **, and * denote the rejection of the original hypothesis at 1%, 5%, and 10% significance levels, respectively, as follows.

In this paper, in order to study the stock market volatility using mixed-frequency data model, the data of the SSE 50 index is selected, and for the convenience of the study, it is chosen from January 4, 2013 to September 31, 2024, the data is from the wind database, and the data of the geopolitical risk index is from the website of economic policy uncertainty. The yield data are daily frequency data with a total of 2827 observations, the GDP data are quarterly data with a total of 47 observations, and the rest of the indicators are monthly data with a total of 141 observations. As can be seen from the table, the SSE 50 return is for the evidence and close to 0, indicating that the average investment return of investors during the sample period is still positive although close to 0. The standard deviation of return is larger than volatility, the change of return is more frequent and the kurtosis value is larger than 3. The J-B statistic also rejects the original hypothesis at 1% significant level, which indicates that the return does not obey the normal distribution and has the characteristics of spiky and thick-tailed.

Table 1: Descriptive statistics of stock market indicators

	Mean	Std	DB	DP	Obs	J-B
$r_{i,t}$	0.0193	1.4423	-0.2374	7.9423	2917	2973.548***
RV	2.1267	2.3189	3.0583	15.1187	141	1125.679***
GPR	0.5783	0.2194	1.1062	4.4238	141	42.5348***
CPI	0.1962	0.5168	0.1379	3.0791	141	58.6927***
GDP	197832.64	60387.41	0.3548	2.2693	47	132.0847***

In Table 2, autoregressive conditional heteroskedasticity (ARCH) test and Q-test are performed on the relevant influences such as SSE 50 return and realized volatility by considering three different lags, and the ADF test (using no convergent term and no constant term) shows that the return series of SSE 50 index obtained after logarithmic differencing is smooth and can be modeled directly without further transformation.

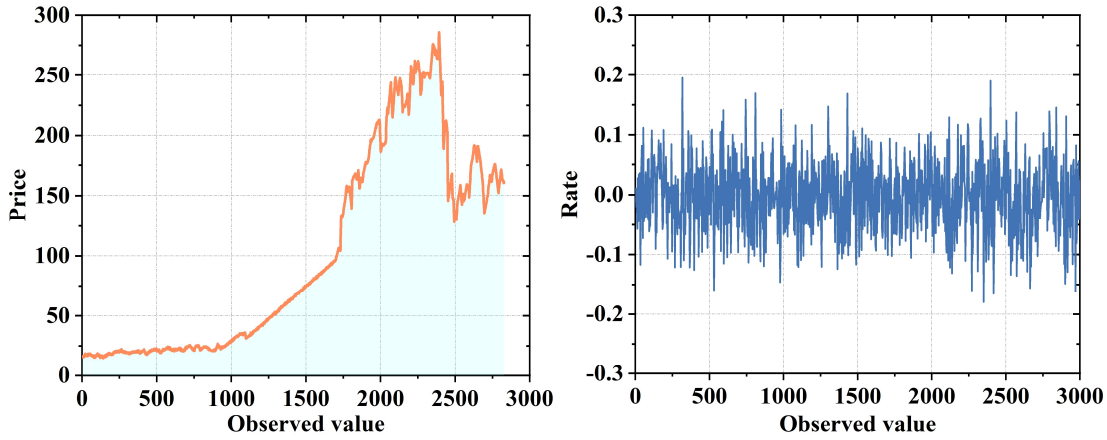
Table 2: Basic analysis of stock market index data

	$r_{i,t}$	RV	GPR	CPI	GDP
ARCH(5)	274.58***	2874.621***	2789.931***	2359.128***	2903.457***
ARCH(10)	306.73***	2813.952**	2764.143***	2674.931***	2899.635***
ARCH(20)	374.86***	2809.483***	2758.192***	2649.676***	2887.842***
Q(5)	24.583***	96.471***	96.847***	17.834***	140.579***
Q(10)	32.689***	139.642***	130.956***	26.956***	205.672***
Q(20)	67.842***	174.868***	181.325***	86.238***	224.531***
ADF	-14.395***	-8.372***	-6.538**	4.175	-2.637

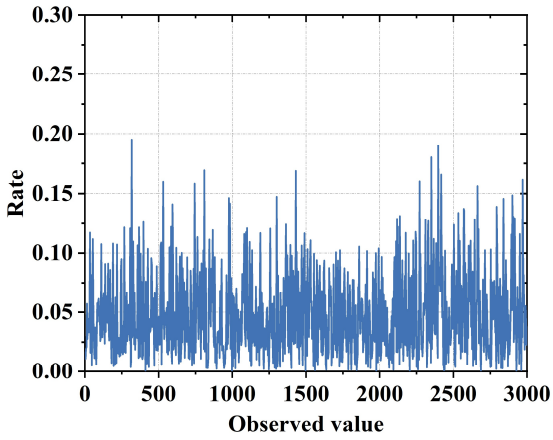
Figures 1(a)-(c) show the time series of the observed data y_t, the return $r_{i,t}$ and the absolute value of the return $|r_{i,t}|$, respectively. As can be seen in Figure (a), between 2013 and 2024, the closing price of the SSE 50 Index has experienced multiple peaks and troughs. From 2013 to the first half of 2018, the share price of the SSE 50 is generally on an upward trend due to the favorable economic situation and the boost of AI, and then produces a decline with a large magnitude and keeps fluctuating from the second half of 2018 to 2024. In short, the share price of SSE 50 is volatile. And it is obvious from Figure (b) that there is a significant fluctuation during the period from October 2023 to November 2024, there is volatility time-variation, and there is heteroskedasticity in the series. The trend graph of absolute value of return, on the other hand, reflects that the volatility of SSE 50 return is characterized by volatility agglomeration.

As the volatility of financial time series audit period generally has a long memory characteristics, so the reveal to calculate the autocorrelation coefficient of each delay order of return, absolute value of return and return squared respectively, the results are shown in Table 3. The 95% confidence range of the autocorrelation coefficient is calculated as $\pm 1.96/\sqrt{n}$, and the amount of data for the yield is 2827, and the confidence range is calculated to be ±0.037, by which the autocorrelation nature of the data can be detected. The autocorrelation coefficients of each delay order of $r_{i,t}$ in the table basically fall into the confidence range, and the delayed first-order autocorrelation coefficient of $r_{i,t}$ is negatively correlated at -0.021, while the delayed first-order autocorrelation coefficients of $|r_{i,t}|$

and $r_{i,t}^2$ are 0.189 and 0.114, respectively. Have high positive correlation, which implies that financial time series in the market are not well regulated and the collected data are not realizations of independent identically distributed processes.



(a) Daily closing price trend (b) Yield trend



(c) Trend of the absolute value of yield

Figure 1: Descriptive analysis of yield changes

Table 3: The self-correlation coefficient of each delay order

	$r_{i,t}$	$\lvert r_{i,t} \rvert$	$r_{i,t}^2$
1	-0.021	0.189	0.114
2	0.003	0.137	0.026
3	0.032	0.126	0.047
7	-0.027	0.048	0.008
10	-0.023	0.096	0.019
18	0.025	0.091	0.048
29	0.007	0.046	0.013
36	0.018	0.069	0.006

Figure 2 shows the autocorrelation coefficient curves for each order of delay, and the blue dashed line in the figure shows the range of 95% confidence intervals. From the figure, it can be seen that $\lvert r_{i,t} \rvert$ and $r_{i,t}^2$ have roughly the same trend, indicating that there is a high degree of positive correlation between them, which implies that there may be long memory in the financial time series.

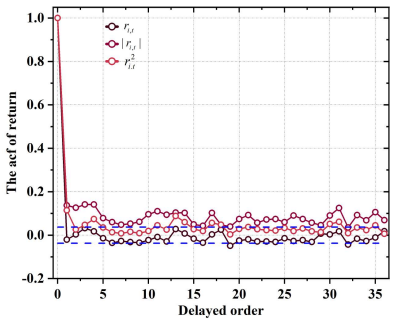


Figure 2: Delay the order self-correlation coefficient

Next, the ARCH-LM test for returns is continued and the results obtained are shown in Table 4. From the table, it can be seen that the P-value obtained after the test of each delayed order is much less than 0.05, which rejects the original hypothesis, indicating that there is a higher-order ARCH effect in the original series, i.e., there is a GARCH effect, and therefore a GARCH model is considered.

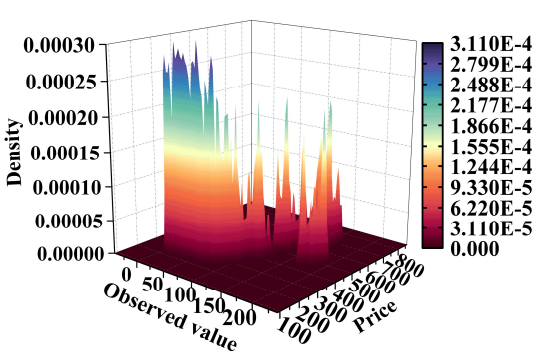
Table 4: ARCH-LM test

Lag	Chi-squared	P-value
1	124.83	2.4E-17
2	157.64	2.4E-17
3	178.31	2.4E-17
4	217.94	2.4E-17
6	219.73	2.4E-17
8	228.59	2.4E-17
12	272.47	2.4E-17
21	283.54	2.4E-17
30	298.67	2.4E-17
36	305.28	2.4E-17

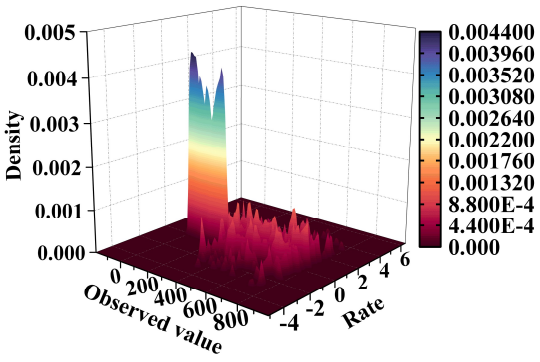
IV. B. Wavelet denoising process

Before establishing the GARCH model, wavelet denoising is performed on the data considering the possible noise in the data. The db3 wavelet function is selected and the data is decomposed into three layers, which ensures in-depth analysis of the data, and at the same time avoids the important information in the signal of the first floor due to too many layers of decomposition.

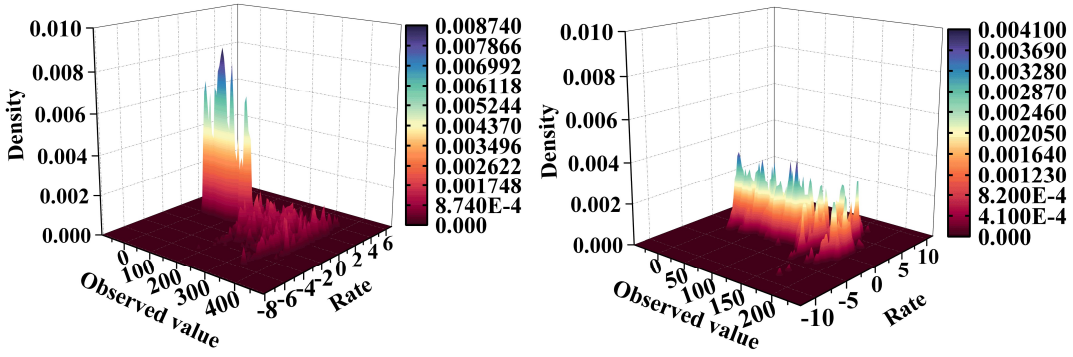
Using Python to wavelet decomposition of the original data, decomposition of the three layers to obtain the low-frequency coefficients A03 and high-frequency coefficients D01,D02,D03, as shown in Fig. 3, (a)-(d) represent the decomposition results under the four coefficients of A03, D01, D02, D03, respectively.



(a) A03



(b) D01



(c) D02 (d) D03

Figure 3: Wavelet decomposition coefficients

After that the obtained high frequency coefficients are denoised by choosing a combination of fixed thresholding and hard thresholding denoising methods, viz:

$$w_{\lambda} = \begin{cases} w, & |w| \geq \lambda \\ 0, & |w| < \lambda \end{cases} \tag{19}$$

where $\lambda = \sqrt{2\log(N)}$ and N is the signal length.

For the low-frequency coefficients and the high-frequency coefficients after the threshold denoising process, wavelet reconstruction is utilized to produce new denoised data, and the results are shown in Fig. 4. Comparing Fig. 4 with Fig. 1(a), it can be found that the wavelet denoising processed data is smoother and also retains the main fluctuation characteristics of the data, indicating that wavelet denoising can achieve the purpose of removing the noise as well as retaining the useful information in the original data. Therefore, in the next study, denoising the data can effectively improve the accuracy of the study.

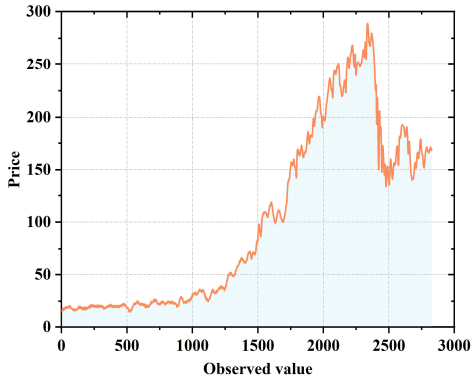


Figure 4: Trend after wavelet threshold denoising

IV. C. Estimation of one-factor GARCH-MIDAS models

According to the results of the above research, there is a high-order ARCH effect after wavelet denoising of the SSE 50 index return data in this paper, and the establishment of a GARCH model should be considered. This paper introduces the idea of frequency mixing in the GARCH model and constructs the GARCH-MIDAS model. This section and the subsequent sections will mainly study the estimation of the GARCH-MIDAS model, so as to verify the validity of the model and explore the volatility characteristics of the SSE 50 index return after wavelet transform.

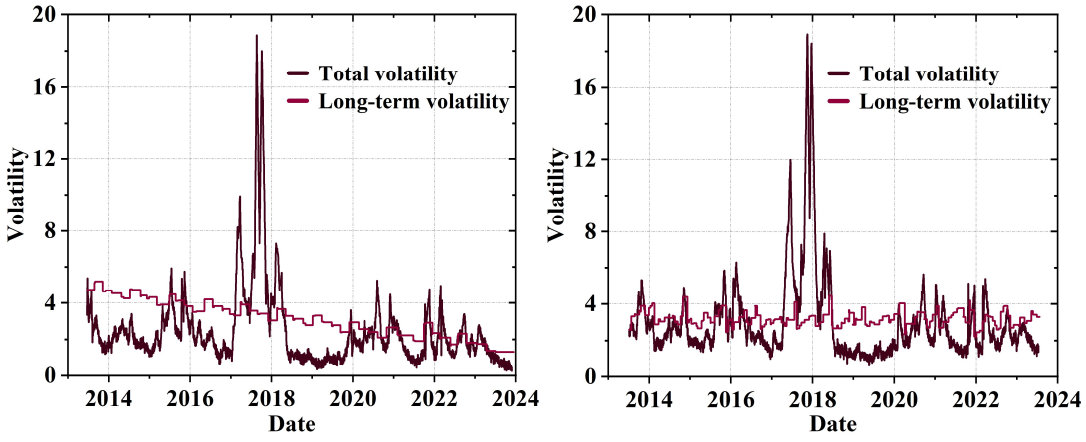
Table 5 shows the parameter estimation results of the one-factor mixed-frequency volatility model constructed based on each indicator, with robust standard errors of the coefficients in parentheses, LLF is the value of the great likelihood function, and BIC is the information criterion, and ***, **, and ** denote that they are significant at the 1%, 5%, and 10% levels, respectively, and are the same as the same below.The parameters of the GARCH-MIDAS model, (α, β), the α and β parameters are significant, and the estimates of the parameter θ represent the extent to which each influencing factor affects the long-run volatility of returns under a one-factor mixed-frequency

model. All of them are significantly positive except for the EPU and TPU models, which are negative, suggesting that rising GDP, inflation rate, infectious disease pandemic level and geopolitical risk all increase the level of long-run volatility of the SSE 50 index, while increasing economic policy uncertainty and trade policy uncertainty decrease the level of long-run volatility of the SSE 50.

Table 5: Single factor model parameter estimation

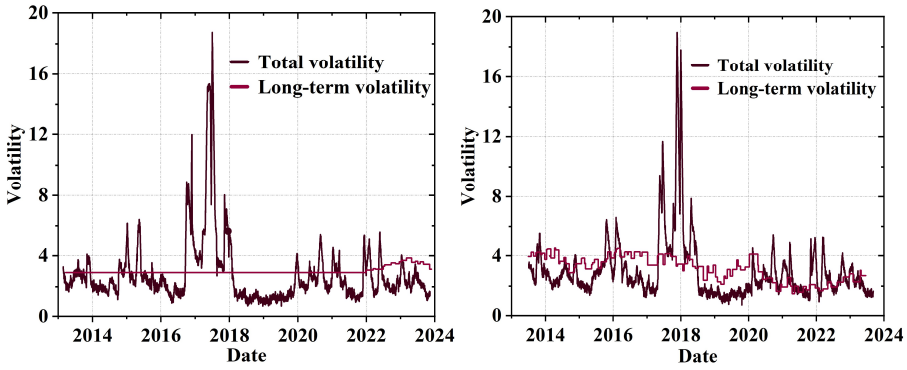
	GDP	CPI	EPU	EMV	GRP	TPU
μ	0.0473**	0.0489**	0.0498**	0.0417*	0.0562**	0.0436**
	(0.0208)	(0.0209)	(0.0213)	(0.0201)	(0.0237)	(0.0215)
α	0.0712***	0.0674***	0.0657***	0.0691***	0.0819***	0.0682***
	(0.0167)	(0.0147)	(0.0163)	(0.0163)	(0.0198)	(0.0161)
β	0.9378***	0.9459***	0.9258***	0.9209***	0.9138***	0.9283***
	(0.0164)	(0.0179)	(0.0181)	(0.0215)	(0.0237)	(0.0189)
γ	-0.0063	-0.0048	-0.0013	0.0043	-0.0029	-0.053
	(0.0171)	(0.0159)	(0.0172)	(0.0204)	(0.0231)	(0.0169)
m	-2.1857***	0.8439***	1.3364***	0.8375**	-0.2478	1.0284***
	(0.4413)	(0.3248)	(0.3698)	(0.3357)	(0.8192)	(0.3267)
θ	0.0006***	0.3096***	-0.1583***	0.0167***	2.1897**	-0.0008*
	(0.0004)	(0.1134)	(0.0572)	(0.0224)	(1.0543)	(0.0005)
ω_2	4.8174***	32.0894***	32.7984***	1.0084	1.0006	33.8746**
	(1.0328)	(7.2857)	(4.5679)	(1.5976)	(1.1273)	(6.0185)
K	15	10	10	20	36	8
LLF	-4673.832	-4779.963	-4803.224	-4396.443	-3521.742	-4639.237
BIC	9074.629	9387.215	9478.752	8917.309	7058.693	9208.794

Figure 5 shows the macroeconomic one-factor model volatility estimation, (a) and (b) are the results of two macroeconomic variables, GDP and CPI, respectively. Figure 6 shows the volatility estimation of the one-factor model of economic uncertainty, (a)-(d) are the results of four economic uncertainty variables, EMV, EPU, GPR and TPU, respectively. As can be seen from the figure, the volatility charts fitted by the one-factor models with GDP and CPI as the as-variables are roughly the same, but the long-run volatility of the two models fluctuates up and down considerably. The shapes of the volatility of the SSE 50 fitted by the single-factor models of economic uncertainty are roughly the same, whereas the long-term volatility of the EMV and GPR models is smoother, the long-term volatility of the EPU model has more ups and downs, and the long-term volatility of the TPU model has a higher frequency of change after 2018.

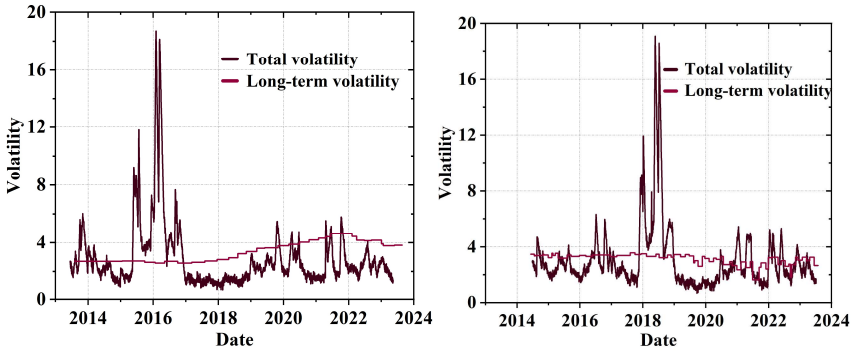


(a) GARCH-MIDAS(GDP) (b) GARCHMIDAS(CPI)

Figure 5: Estimation of volatility in macroeconomic signal factor models



(a) GARCH-MIDAS(EMV) (b) GARCHMIDAS(EPU)



(c) GARCH-MIDAS(GPR) (d) GARCHMIDAS(TPU)

Figure 6: Estimation of economic uncertainty single factor model volatility

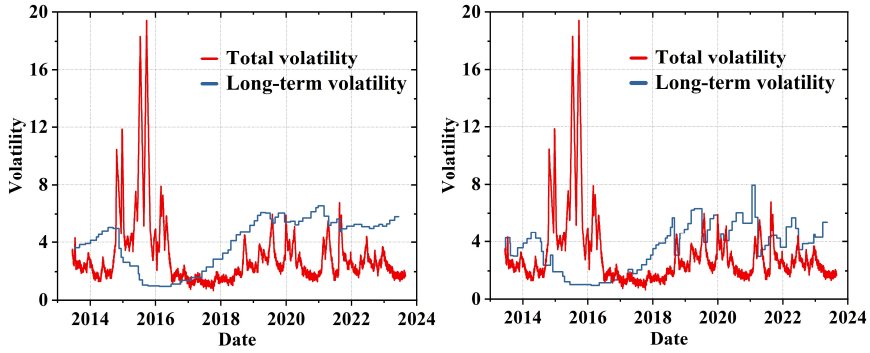
Table 6: Estimation results of the dual factor model in stock market

	RV+GDP	RV+CPI	RV+EPU	RV+EMV	RV+TPU	RV+GPR
μ	6.0248**	0.0672***	0.0589***	0.0573**	0.0589**	0.0487
	(0.0239)	(0.0258)	(0.0257)	(0.0263)	(0.0238)	(0.0821)
α	0.0823***	0.0784***	0.0714***	0.0732***	0.0713***	0.0683
	(0.0187)	(0.0168)	(0.0183)	(0.0173)	(0.0184)	(0.0804)
β	0.9236***	0.9245***	0.9167***	0.9218***	0.9267***	0.9176***
	(1.7564)	(0.0154)	(0.0254)	(0.0194)	(0.0231)	(0.0413)
γ	-0.0013	-0.0176	-0.0005	-0.0029	-0.009	0.0003
	(0.0217)	(0.0168)	(0.0178)	(0.0203)	(0.0021)	(0.2136)
m	3.1126*	2.9386**	2.6894***	2.9465**	2.8647***	2.9158***
	(1.6193)	(1.1467)	(0.6832)	(1.4006)	(1.0069)	(1.0078)
θ_1	-1.0736	-1.0926**	-0.7086**	-0.9614	-0.8304*	-0.5547
	(7.2293)	(0.5578)	(0.2534)	(0.6687)	(0.4268)	(0.3412)
θ_2	0.0004***	0.3547***	-0.1173	-0.0347	-0.0009**	-1.7288*
	(0.0001)	(0.1322)	(0.0904)	(0.4789)	(0.0005)	(1.3426)
ω_{12}	2.1716***	2.4274***	1.0005***	1.7635***	1.0228	1.0003*
	(0.7783)	(0.7531)	(0.3547)	(0.5763)	(0.6937)	(0.5847)
ω_{22}	2.5698	282.7483***	1.0008*	1.0001	154.082***	7.1043**
	(6.2471)	(44.2861)	(0.5916)	(18.9472)	(37.5143)	(3.5628)
K_1	45	47	35	40	35	25
K_2	45	47	35	40	35	25
LLF	-3259.643	-3179.834	-3716.483	-3488.231	-3631.022	-3903.074
BIC	6582.179	6423.561	7503.282	7147.586	7332.083	7876.573

IV. D. Two-factor GARCH-MIDAS model estimation

Table 6 shows the parameter estimation results of the two-factor model of the stock market, θ_1, θ_2 reflect the realized volatility RV and the degree of influence of each macroeconomic and economic uncertainty variable on the long-run volatility of the SSE 50 index, respectively, and it can be seen that the parameters of the GDP and CPI models θ_2 are significantly positive, indicating that the long-term volatility of the SSE 50 Index also rises when GDP and inflation rise, while the parameters θ_2 of the TPU and GPR models are significantly negative, indicating that the long-term volatility of the SSE 50 Index declines when trade policy uncertainty and China's geopolitical risk rise.

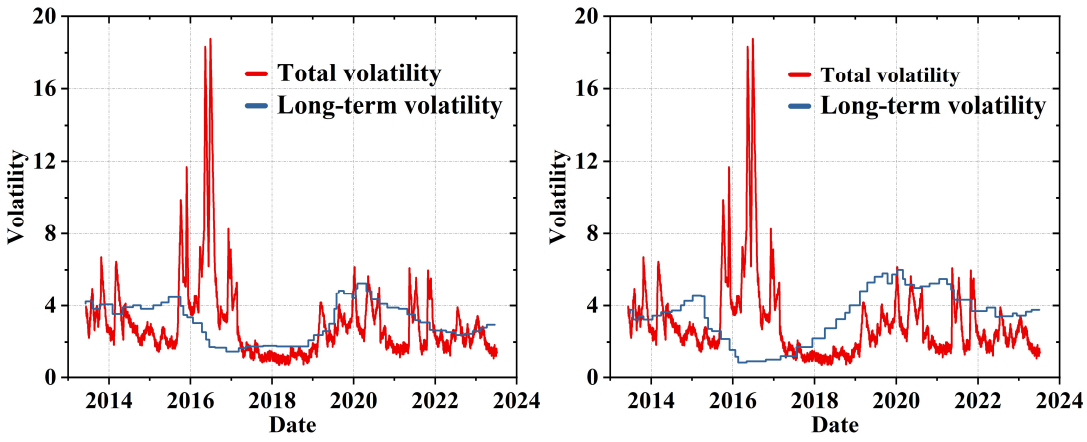
Figure 7 shows the estimation of the volatility of the two-factor model of macroeconomics, (a) and (b) for RV+GDP and RV+CPI, respectively. Figure 8 shows the estimation of the volatility of the two-factor model of economic uncertainty, (a)-(d) are the estimation results of RV+EPU, RV+EMV, RV+TPU and RV+GPR, respectively. From the results in the figure combined with the parameter estimation table, it can be seen that the volatilities fitted by the six two-factor models have similar shapes, with the GDP model having a large upward fluctuation in long-term volatility in 2015-2017 and the CPI model having a large ups and downs in long-term volatility after 2018. The four economic uncertainty models also have large changes in long-term volatility after 2018, with the most pronounced change in the TPU model.



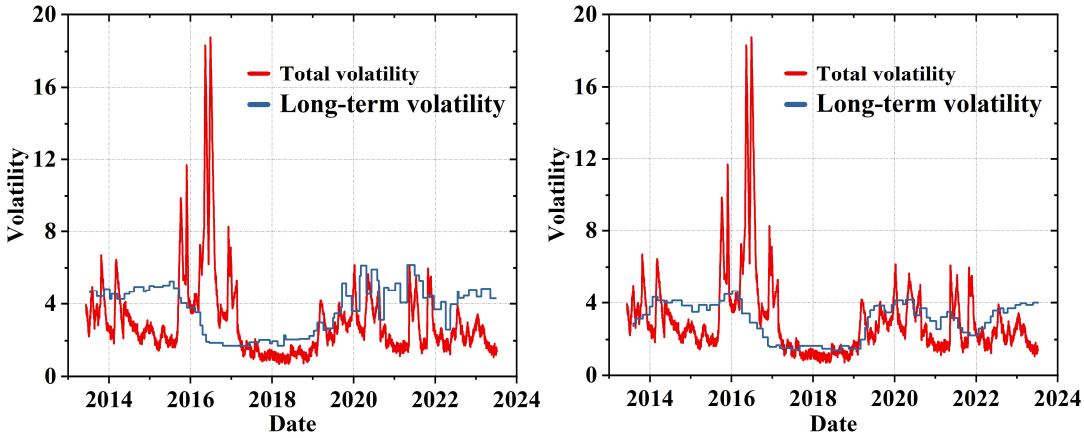
(a) GARCH-MIDAS(RV+GDP) (b) GARCHMIDAS(RV+CPI)

Figure 7: Estimation of volatility in macroeconomic dual factor models

Figure 9 shows the Beta weight plots of the θ_2 significant four two-factor models, from which it can be seen that the effects of GDP and GPR on the volatility of the SSE 50 index take 42 and 16 months to be fully absorbed respectively, while the effects of CPI and TPU on the SSE 50 index take only 2 months to be fully absorbed. It shows that the period of influence of changes in the macroeconomic side of the stock market volatility is long, and changes in trade policy uncertainty and fluctuations in geopolitical risk indices do not affect the stock market for a long period of time.

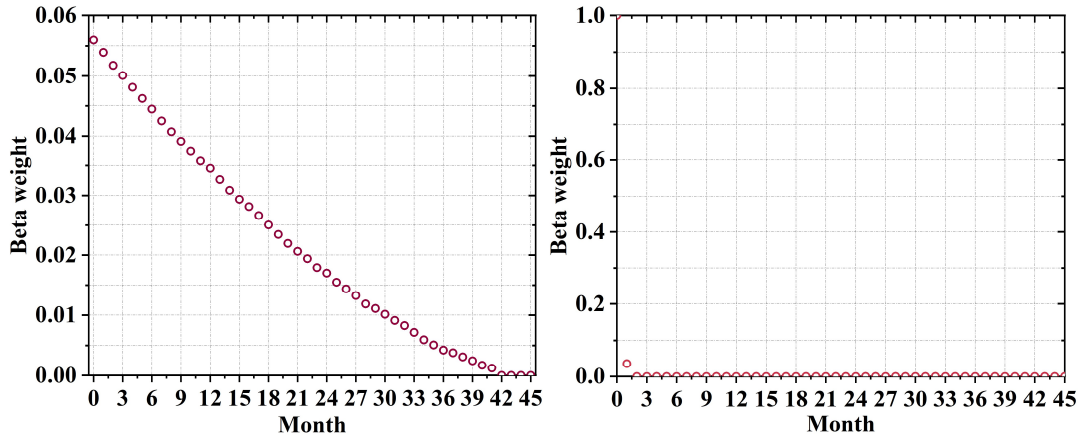


(a) GARCH-MIDAS(RV+EPU) (b) GARCHMIDAS(RV+EMV)

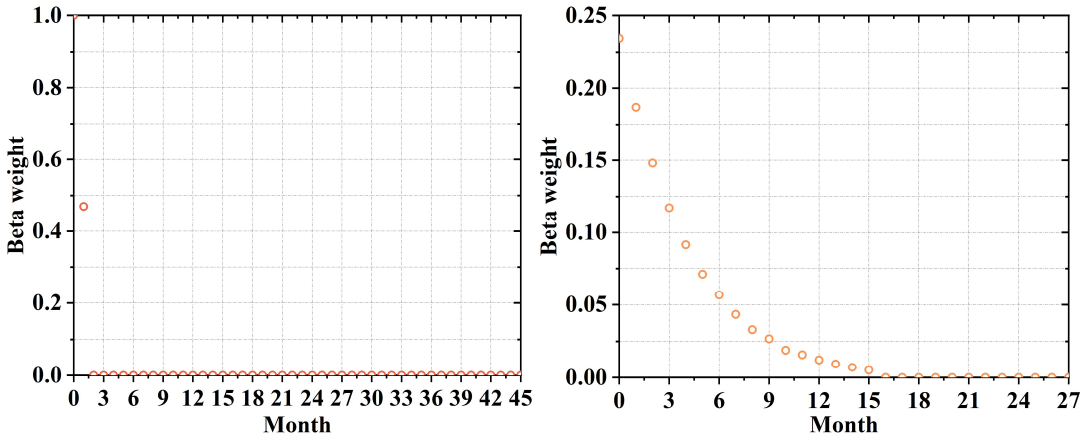


(c) GARCH-MIDAS(RV+TPU) (d) GARCHMIDAS(RV+GPR)

Figure 8: Estimation of economic uncertainty dual factor model volatility



(a) RV+GDP (b) RV+CPI



(c) RV+TPU (d) RV+GPR

Figure 9: Dual factor model beta weight

In summary, through the estimation results of the one-factor model and two-factor model, it can be obtained that the GARCH-MIDAS model can effectively capture the fluctuation characteristics of the data during the auditing

period based on the wavelet threshold denoising processed data, and it has an important application value in the in-depth analysis of the data.

V. Conclusion

In this paper, the wavelet transform is used to realize the denoising of data containing fluctuations, and the idea of frequency mixing is introduced into the GARCH model to establish the GARCH-MIDAS model for capturing the fluctuation characteristics in the data. The experimental results verify the effectiveness and reasonableness of the method proposed in this paper, which provides a new technical means for auditing and helps auditors to more quickly and accurately detect irregularities and risky operations in the field of economic activities and other areas. Future research can further optimize the data processing method and the fluctuation feature extraction algorithm to adapt to a more complex audit data environment. At the same time, the number and type of samples that can be collected for the study should be expanded to make the study more general and representative.

About the Author

2023 General Project of Sichuan Key Research Base of Humanities and Social Sciences in Colleges and Universities - Water Transportation Economy Research Center: “Research on the Impact of Resource and Environment Audit on the High-Quality Development of the Upper and Middle Reaches of the Yangtze River Economy” (SYJJ2023B01).

References

[1] Usul, H., & Alpay, M. F. (2024). From Traditional Auditing to Information Technology Auditing: A Paradigm Shift in Practices. European Journal of Digital Economy Research, 5(1), 3-9.

[2] Changqing, L. (2023). Analysis of traditional financial audit industry based on improved algorithm of block chain technology. Journal of Commercial Biotechnology, 28(2).

[3] Freiman, J. W., Kim, Y., & Vasarhelyi, M. A. (2022). Full population testing: Applying multidimensional audit data sampling (MADS) to general ledger data auditing. International Journal of Accounting Information Systems, 46, 100573.

[4] Abdelwahed, A. S., Abu-Musa, A. A. E. S., Badawy, H. A. E. S., & Moubarak, H. (2025). Investigating the impact of adopting big data and data analytics on enhancing audit quality. Journal of Financial Reporting and Accounting, 23(2), 472-495.

[5] Mistri, P. (2022). Cloud Security Audit: A necessity in the cloud computing environment. The Management Accountant Journal, 57(7), 64-67.

[6] Ganapathy, V. (2023). AI in auditing: A comprehensive review of applications, benefits and challenges. Shodh Sari-An International Multidisciplinary Journal, 2(4), 328-343.

[7] Eilifsen, A., Kinserdal, F., Messier Jr, W. F., & McKee, T. E. (2020). An exploratory study into the use of audit data analytics on audit engagements. Accounting Horizons, 34(4), 75-103.

[8] Duan, H., Du, Y., Zheng, L., Wang, C., Au, M. H., & Wang, Q. (2022). Towards practical auditing of dynamic data in decentralized storage. IEEE Transactions on Dependable and Secure Computing, 20(1), 708-723.

[9] Shang, T., Zhang, F., Chen, X., Liu, J., & Lu, X. (2019). Identity-based dynamic data auditing for big data storage. IEEE Transactions on Big Data, 7(6), 913-921.

[10] Zhuang, C., & Liao, P. (2020). An improved empirical wavelet transform for noisy and non-stationary signal processing. IEEE Access, 8, 24484-24494.

[11] Othman, G., & Zeebaree, D. Q. (2020). The applications of discrete wavelet transform in image processing: A review. Journal of Soft Computing and Data Mining, 1(2), 31-43.

[12] Al-Qerem, A., Kharbat, F., Nashwan, S., Ashraf, S., & Blaou, K. (2020). General model for best feature extraction of EEG using discrete wavelet transform wavelet family and differential evolution. International Journal of Distributed Sensor Networks, 16(3), 1550147720911009.

[13] Yao, Y., Ma, J., & Ye, Y. (2023). Regularizing autoencoders with wavelet transform for sequence anomaly detection. Pattern Recognition, 134, 109084.

[14] Bohan Ma,Yushan Xue,Yuan Lu & Jing Chen. (2025). Stockformer: A price–volume factor stock selection model based on wavelet transform and multi-task self-attention networks. Expert Systems With Applications,273,126803-126803.

[15] Shuyang Jiang,Hongjuan Zhang,Yi Lu,Ruobing Han,Yan Gao,Yu Wang & Baoquan Jin. (2025). SNR enhancement for Raman distributed temperature sensors using intrinsic modal functions with improved adaptive wavelet threshold denoising. Optics and Lasers in Engineering,189,108949-108949.

[16] Matúš Horváth & Tomáš Výrost. (2025). No shortfall of ES estimators: Insights from cryptocurrency portfolios. Finance Research Letters,73,106685-106685.