

Volatility Forecasting of Copper Futures Based on HAR-RV Model

Ziyi Fang^{1, *, †}, Chenyang Zhao^{2, a, †}, Zichun Zhong^{3, b, †}

¹School of Finance and Statistics Hunan University, Changsha, China

²School of Economics and Finance Queen Mary University of London, UK

³School of Computer Science and Engineering Central South University, Changsha, China

*Corresponding author: guanghua.ren@gecademy.cn, ^azcy98612@163.com,
^b2069838869@qq.com

[†]These authors contributed equally.

Abstract. As an important part of the international futures market, copper price prediction is important for international financial market research. This paper selects the high-frequency data every 5 minutes from the database and uses the HAR-RV model based on realized volatility. By introducing investor sentiment and the day of week effects, we have established three new types of non-uniform autoregressive models. Empirical analysis shows that the weekly and monthly fluctuations of copper futures prices are relatively small, while the daily fluctuations are relatively large. The prediction model is more accurate when predicting the long-term volatility, and the stability test shows that the HAR-RV model is relatively stable when predicting the long-term volatility. Investor sentiment has a negative impact on the price volatility of copper futures in the medium and long-term forecasts. Weekend effects have a negative impact on the medium and long-term forecasts of copper futures. This paper complements the existing literature and improves the prediction ability of copper price fluctuations, which is very important to promote effective hedging, risk transfer, and price discovery in the futures market.

Keywords: HAR-RV model; Copper Futures Market; Weekend Effects Investor Sentiment.

1. Introduction

As a metal element, copper is widely present in nature and widely used in many industries such as electric power, energy, chemical industry, national defense and military industry, and transportation. For a long time, due to copper standardization and easy storage characteristics, copper spot trading is highly market-oriented and has extremely high liquidity [1]. Therefore, copper import trade has become a good carrier for financing-oriented trade under the Chinese unique financing environment. A large amount of financing trade has made domestic copper prices lower than foreign copper prices for a long time [2]. On November 19, 2020, international copper futures were officially listed for trading at Shanghai International Energy Exchange, using "international platforms, net price trading, bonded delivery and RMB pricing" as the listing model, fully introducing foreign investors to participate, filling the international trade of the copper industry chain. The blank of RMB-denominated hedging instruments is also the first time that the domestic futures market in China has achieved internationalization in a "dual contract" model. On March 22, 2021, the first international copper futures contract BC2103 of Shanghai Futures Energy, a subsidiary of Shanghai Futures Exchange, was successfully delivered, with a delivery volume of 6,225 tons and an amount of 370 million yuan. Since its listing, the international copper futures market has generally been operating steadily, with a steady expansion of scale, a reasonable investor structure, and the gradual development of market functions [3].

Improving the ability to predict the volatility of copper prices is of great significance to promoting the effective use of functions such as hedging, risk transfer, and price discovery in the futures market. This paper uses the HAR-RV model to conduct relevant research on the volatility prediction of copper futures.

Copper prices, as a key signal, can be sensitive to changes in economic and political events. It can show investor sentiment. Since the 21st century, the average copper price in China has fluctuated sharply. Before 2006, the product price gradually increased, rising rapidly from about 18,000 yuan per ton to nearly 80,000 yuan per ton, the lowest price fell to 25,000 yuan per ton, fell more than double; with the economic recovery, 2009-2011, then continued until 2016, prices continued to fluctuate steadily in 2017, 2017-2019, the middle of the COVID-19 epidemic in early 2020 [2]. In the existing studies, Chen Lin et al. [4] can use the ARIMA model to predict and analyse the price of copper futures. Chen Xiaodong [5] Through the GARCH model analysis, the Shanghai-copper futures price fluctuations are characterized by sharp peaks, thick tail, agglomeration, and sustainability. Han Liyan et al. [6] analyses the influence of various domestic and foreign factors on the price of commodity futures based on the enhanced VAR model system. This paper will try to study the price volatility of copper futures using the HAR-RV model based on high-frequency data, which can supplement the existing literature.

In the existing research results, the HAR-RV model has had more research and applications. Andersen et al. [7] first take high-frequency metric data computing as a brand-new noise volatility high-frequency measurement calculation, studying problems associated with price prediction using realized volatility (realized volatility, RV) based on high frequency metric data. Moreover, Andersen et al. [8] found that simple models based on RV can have stronger analysis and prediction on the probability of financial asset price fluctuations than the models of GARCH, etc. Then, using the heterogeneity market hypothesis as to its theoretical basis, Corsi [9] proposed the heterogeneity autoregressive realized volatility model (HAR-RV model). It regarded the market fluctuations as the result of the joint action of high frequency, medium frequency, and low-frequency traders. It better characterized the process superposition of fluctuation autoregressive through three different time sizes of day, week, and January, indicating that the HAR-RV model is significantly more accurate than predicting future volatility models of GARCH and ARFIMA-RV. The HAR-RV model proposed by Corsi largely has a better predictive ability for the financial market's volatility ratio and has a good financial significance. This important study will push the study of the importance of price volatility ratio prediction in our financial market to another new level, and a large number of researchers have found that different types of financial markets are significantly stronger than those of GARCH, SV, VAR-RV, or ARFIMA-RV. Andersen et al. [10] propose the HAR-RV-J and HAR-RV-CJ models that distinguish between jump and continuous fluctuations and empirically state that introducing square root and logarithmic transformations can improve the fitting ability of the model. Zhang Xiaoyong et al. [11] considered the influence of overnight information based on the HAR-RV-CJ model, expanded the model, and finally formed the HAR-RV-CJN model. Qu Hui et al. [12] pointed out that HAR family models have better prediction accuracy than previous models based on low-frequency data. Due to the excellent prediction performance of the model, many researchers have recently adopted the HAR family model to study the volatility of the financial market [13]. So far, the HAR family model has been widely used.

The earliest study of the intra-week effect was the article called "Stock Price: A Problem in Verification" written by Fields in 1931 [14], which first mentioned the intra-week effect and analysed the market week effect, attracting much attention. Due to a large number of pursuers of excess returns in the market, the discovery of the intra-week effect means that there may be arbitrage space in the market. Subsequently, more researchers continued to study the market visions based on the analysis of Fields, further enriching the relevant theory of stock market visions. Cross [15] took a sample of the S & P 500 1953-1970 yield and found negative Monday and positive Friday effects in the U. S. market. French [16] selected new data to verify the effectiveness of the US stock market again, with the same conclusion as Cross. At the same time, different markets have been proved to have a weekly effect. Still, the weekly effect may differ between countries due to cultural background and securities market development degree. Jaffe and Westerfield [17] found that Britain and Canada, significantly negative Monday effect, Australia and Japan weekly effect consistent, there is a significantly negative Tuesday effect, different countries have their own unique form of expression. Aggarawal and Rivoli

[18] found emerging markets such as the Philippines with negative abnormal gains Monday and Tuesday. Thus, there are different manifestations in the performance of different periods and different markets.

In this paper, the HAR-RV model uses five-minute high-frequency data. Through an in-sample analysis, we find relatively small weekly and monthly fluctuations in the prices of copper futures, and the daily fluctuations are relatively large. In terms of the prediction accuracy, we find that the HAR-RV model, HAR-RV-V model, HAR-RV-W model, and HAR-RV-WV model have a better fit in predicting 1 week and January volatility. While having a poor fit capability when predicting the 1-day volatility, it shows that only long-term (monthly) volatility is predicted using these models. This phenomenon is also present in the predicted 1-week volatility and 1-day volatility. When predicting the fixed periodic volatility, the volatility short than this period has no sufficient predictive significance.

Otherwise, we also consider investor sentiment. Baker, Wurgler, and Yuan argue that investor sentiment generates irrational trading behaviors (e.g., markets' overreactions and underreactions) and affects the dynamics of asset returns and trading volumes. If investors' trading decisions are contaminated by irrationalities such as overconfidence, selection bias, and optimism, investor sentiment can directly determine investors' trading behaviors. Karam Kim and Doojin Ryu analyze the Korean stock market. Speculative trading prevails in this leading emerging market, and individual investors' participation rates are high, implying that the market is likely to respond sensitively to behavioral and sentiment issues [19]. In this paper, we add VIX Index to form a new model HAR-RV-V. VIX Index is the index obtained after the weighted average of the implied volatility of index options. When the VIX Index is higher, it means that market participants expect the degree of future market volatility to be more intense; On the contrary, if the VIX Index is lower, it reflects the mentality of market participants who expect the degree of future volatility to ease. Therefore, the VIX Index is also known as the investor panic index.

The weekend effect is a vision against the effective market hypothesis prevalent in the stock futures market, where the average yield is higher or lower and statistically significant than the average yield on any other day of the week. The first to study the futures market calendar effect was Chiang and Tapley [20]. They studied 21 commodity futures varieties traded on the Chicago Futures Exchange, which showed a negative Monday effect and the largest trading volume in the week. Cornell [21] discussed the weekly calendar distribution of the SP500 index and SP500 recent month futures contract income, and whose selected time span was 27 months from May 1982 to July 1984. The results show that the income of the SP500 index has a weekly calendar effect, but the income of the SP500 futures price index does not exist. Gay and Kim [22] examined the weekly calendar effect based on the 29-year Futures Price Index (Futures Price Index) released by the Commodity Research Agency (Commodity Research Bureau) and found that Wednesday and Friday's earnings were significantly greater than zero, Friday the highest of the week, and Monday the lowest and negative number of the week. In addition, Agrawal and Tandon [23], Bowers, and Dimson [24] studies show that weekly calendar effects vary between the study assets and the study time span. By considering Sunday effects and investor sentiment factors, we establish a HAR-RV-W model with Sunday effects and HAR-RV-WV factors with investor sentiment and Sunday effects.

The sample analysis shows that only the above four models have long-term (month) fluctuations when using the HAR-RV model, HAR-RV-V model, HAR-RV-W model, and HAR-RV-WV model. This phenomenon also exists in predicting one-week fluctuations and one-day fluctuations. When volatility is less than that period, it is predicted. There is no sufficient predictive significance. In terms of prediction accuracy, we find that these four models have a better fit in predicting 1-week and 1-month fluctuations and poorly in predicting one-day fluctuations.

The rest of the paper is organized as follows: part 2 describes the samples and data, part 3 introduces the HAR-RV model, part 4 examines the sample analysis, and the last part is the study's conclusions.

2. Data

Due to the influence of many factors, the price of copper futures is in a state of large fluctuations in spreads. The sharp price fluctuations have brought huge challenges to copper product demand and supply companies. To protect the steady development of copper-related industries and the interests of related entities, analysing the price volatility influencing factors is very necessary [25]. Since the concept of Realized Volatility (RV) was proposed, Andersen [7] first proposed the use of high-frequency data as a new volatility measurement method. Wei Yu [26] demonstrated high-frequency trading through empirical research. The data helps to improve the prediction accuracy of the volatility model and the prediction accuracy with 5 minutes as the sampling frequency is the highest, which contains as much effective information as possible without too much noise. Therefore, we select trading data with a frequency of 5 minutes to measure the price fluctuations of copper futures. This paper has obtained copper futures trading data for the past 3 years through the wind database. The sample period starts on January 2, 2018, and ends on June 2, 2021. There are 72,944 copper futures transaction data. After calculating the RV and excluding the vacant value, the final result is from January 13, 2018, to May 2021. Data for 957 trading days on the 28th.

Figure 1 shows that the price continued to fluctuate and stabilize from 2018 to 2019, but it remained low. At the beginning of 2020, due to the severe impact of the COVID-19, the price of copper futures fell severely in the first three months. After a clear trough, the epidemic situation was initially brought under control. The copper futures market gradually recovered, and the price continued to rise, rising rapidly, reaching a new peak in May 2020, with a trend of breaking through 80,000.

It can be clearly seen from Figure 2 that there were two serious structural mutations in the volatility of copper futures. One was in February 2020 due to the impact of the COVID-19. The sharp drop in copper futures prices caused major fluctuations, and the other was in 2020. After the initial control of the new crown pneumonia in April 2008, the rapid recovery of copper futures prices led to significant fluctuations. It can be seen that the new crown epidemic has indeed had a great impact on the economy. We cannot ignore this factor when predicting the price trend of copper futures.

From Figure 3, we can see that the copper futures market has a relatively significant yield asymmetry. A stronger downward movement often accompanies the upward movement of copper futures prices. That is, a negative price shock will lead to greater returns. The rate volatility is common in the stock, currency, and bond markets. At the same time, it can be found that most of the sample logarithmic returns are concentrated near 0, and only a few samples are far away, showing an obvious "spike and thick tail" phenomenon.

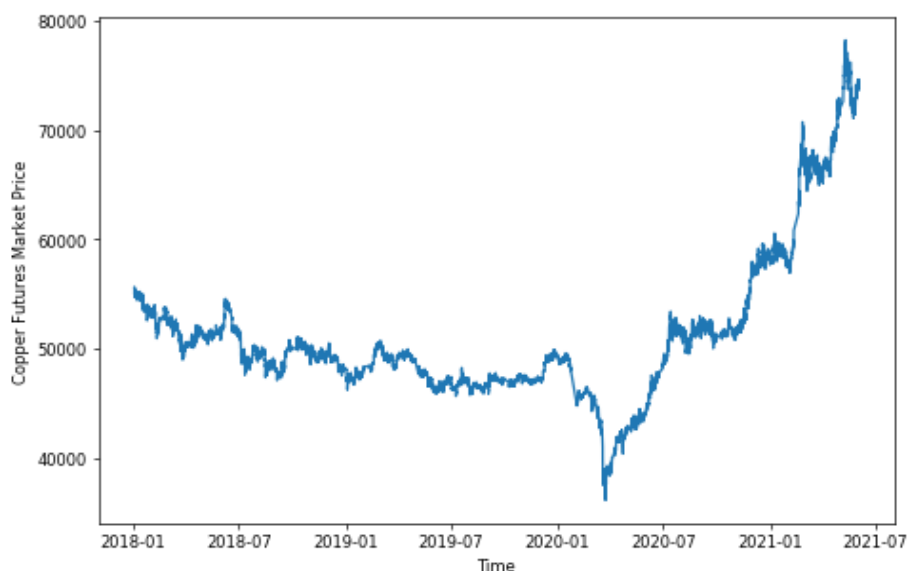


Figure 1. Five-minute high-frequency price dynamics of copper futures.

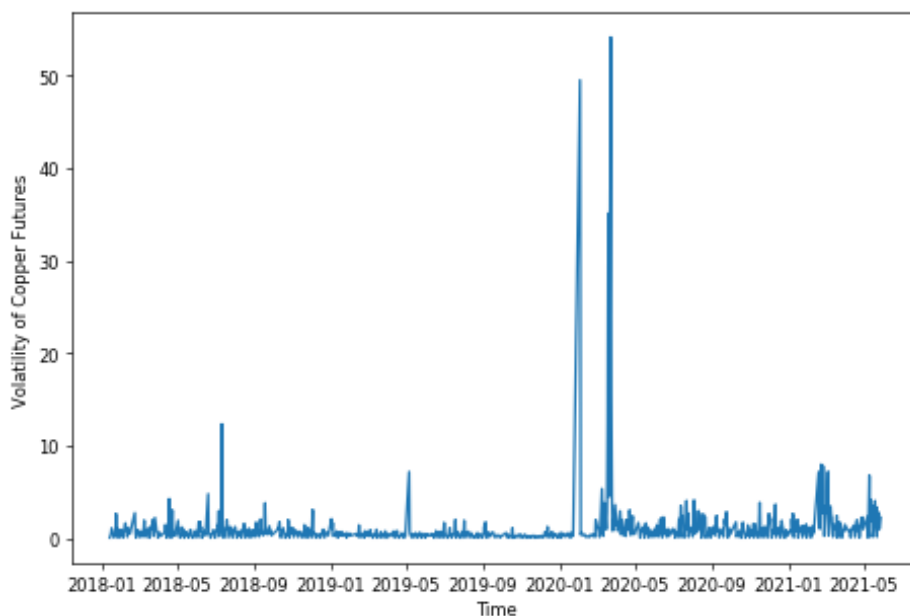


Figure 2. Daily price fluctuation dynamics of copper futures.

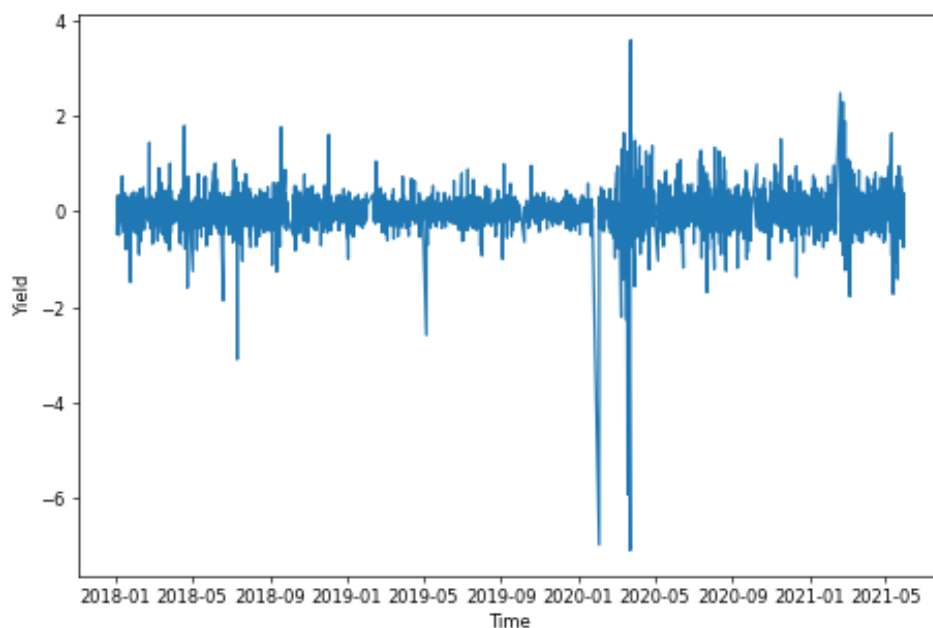


Figure 3. Five-minute high-frequency logarithmic return rate of copper futures.

In the HAR-RV-V model, we mainly added the VIX index as another possible influencing factor. The implied volatility VIX (Volatility Smile) is formed because the volatility of the parity series will be lower than that of the out-of-price series. Market participants are more willing to avoid risks when the index is falling than when the index is rising. Therefore, when the index falls, the demand for hedging to buy put options will increase, and it will also push up the implied volatility of deep out-of-the-money put options. VIX reflects the views of option market participants on the volatility of the market outlook. It is often used to judge the market's long and short contrarian indicators. We obtained 740 sets of VIX index data from the wind database from January 2, 2018, to May 28, 2021, corresponding to copper futures' effective trading day data and excluding the null value obtained from January 2018 700 VIX data from 15th to May 28th, 2021.

It can be seen from Figure 4 that both 2018 and 2019 reached the VIX peak at the beginning of the year, which means that investors believe that the market will fluctuate very sharply, and then fluctuate at a lower level in a small range. In 2020, the VIX index will show a relatively great the big change

was that the VIX index broke through 80 in mid-March and reached more than twice the peak value of the previous two years. This is due to the large-scale outbreak of foreign epidemics and the rapid increase in investor panic, followed by strong epidemic blockade control policies and flows. The VIX index fell to the shoulders in April (around 60). In May, the foreign epidemic was initially controlled, and the VIX index was controlled at around 30. However, the VIX index fluctuated at a relatively high level in the following year.

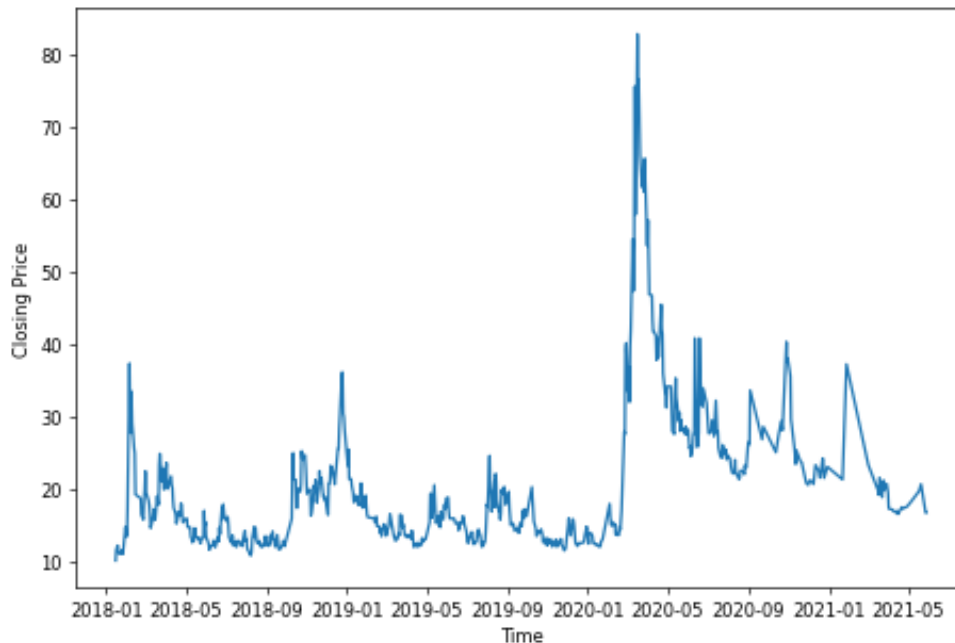


Figure 4. VIX index fluctuation chart.

3. Theoretical basis of the HAR-RV model

The HAR-RV model is a first-order autoregressive model established for the realized volatility at different time intervals [27]. Therefore, it is necessary to introduce the relevant definition of the implemented volatility model. With the average yield is zero, the realized volatility is usually estimated using the following formula:

$$\sigma^2(T, \Delta t) = \sum_{j=1}^n r^2(\Delta t, t_j) \tag{1}$$

Here, $r(\Delta t, t_j) = \ln(P(t_j)) - \ln(P(t_j - \Delta t))$ represents the logarithmic yield of each time interval. T represents the length of time for calculated volatility; Δt represents the time interval for calculated realized volatility; n represents the number of time interval to calculate the realized volatility; $T = n\Delta t$; t_j represents the calculated time point in time; $\sigma^2(T, \Delta t)$ represents the realized volatility $\sigma^2(s)$ with the length of time T and time interval Δt .

According to the research results of Dong Dianhua et al. [27], when there is no sequence correlation between the high-frequency data yields in the same day, it can be proved by the secondary variation theory:

$$p \lim_{n \rightarrow \infty} \{ \sum_{j=1}^n r^2(\Delta t, t_j) \} = \int_{t_0}^{t_0+n\Delta t} \sigma^2(s) ds \tag{2}$$

That is, when the data frequency is high enough, the volatility convergence to the integral volatility (real volatility). The consistency of the realized volatility is based on the assumption that the asset price is continuously observed with no measurement errors. However, the sampling frequency is

bound to be restricted by the trading mechanism in the actual trading. Because the high-frequency prices are affected by the market micro-noise. For example, the transaction bidding jump, nonsynchronous transaction, closed market effect, etc., so the real price is not observable. The presence of microscopic market noise will lead to a bias in the estimates of the realized volatility.

Suppose the observed unbiased high-frequency logarithmic asset price to be:

$$Y_{t_j} = Y_{t_j}^* + \varepsilon_{t_j} \tag{3}$$

Among them, $Y_{t_j}^*$ represents the observed price represents the real potential price. Further, the microscopic market noise is assumed to mean 0 and a variance of $\eta \varepsilon_{t_j} 2 \varepsilon_{t_j} Y_{t_j}^* r_{t_j}^*$. And obey an independent and homogeneous stochastic process. Independent from each other. Note as effective yield, then high frequency continuous composite yield follows a MA (1) process:

$$r_{t_j} = r_{t_j}^* + \varepsilon_{t_j} - \varepsilon_{t_{j-1}} \tag{4}$$

The volatility is defined by the above:

$$RV_n = \sum_{i=1}^M (r_{t_j}^*)^2 + 2 \sum_{j=1}^M r_{t_j}^* (\varepsilon_{t_j} - \varepsilon_{t_{j-1}}) + \sum_{j=1}^M (\varepsilon_{t_j} - \varepsilon_{t_{j-1}})^2 \tag{5}$$

$$E[RV_n] = E[RV_n^*] + 2M\eta 2 \tag{6}$$

We can see that the realized volatility is disturbed by two errors: measurement and microstructural errors. By observing the above definition, we can find that the choice of frequency is very critical: on the one hand, if the frequency selection is too high, although the measurement error will be reduced, the microstructural error will be very large; on the other hand, if the frequency selection is too low, the interference of the microstructural error will be very small, but the measurement error rises to the main error because it will be greater than the microstructural error. Therefore, the optimal frequency selection must be balanced between these two classes of errors. When the realized volatility is currently calculated, most scholars in the current literature believe that the estimated 3- or 5-minute sampling frequency is optimal [28]. Therefore, the realized volatility will be measured using 5-minute-high frequency data.

Based on realized volatility, the HAR-RV model is a first-order autoregressive model established for realized volatility calculated at different time intervals. To facilitate the analysis, we generally divide the traders into three categories: daily traders, weekly traders, and monthly traders, corresponding to the short, medium, and long-term traders, respectively. To this end, we calculate the daily realized, weekly, and monthly realized volatility of copper futures. Generally, there are 5 trading days a week and 22 trading days a month. $(RV_t^d)(RV_t^w)(RV_t^m)$ The calculation formula is shown as follows:

$$RV_t^d = \sum_{j=1}^n r^2(\Delta t, t_j) = \sum_{j=1}^{48} r^2(\Delta t, t_j) \tag{7}$$

$$RV_t^w = \frac{1}{5} \sum_{i=1}^5 RV_{t-i}^d \tag{8}$$

$$RV_t^m = \frac{1}{22} \sum_{i=1}^{22} RV_{t-i}^d \tag{9}$$

Under the framework of the heterogeneous market hypothesis, short-term realized volatility is affected not only by the short-term lag period but also by the mid-term realized volatility at the same time. Similarly, the realized volatility in the medium term will be affected by its own lag in the first

phase and the long-term realized volatility at the same time. Long-term realized volatility is only affected by its own lagging phase. That is:

$$RV_t^d = c^m + c_1^m RV_{t-1}^m + \varepsilon_t^m \tag{10}$$

$$RV_t^w = c^w + c_1^w RV_{t-1}^w + c_2^w E_{t-1}[RV_t^m] + \varepsilon_t^w \tag{11}$$

$$RV_t^m = c^d + c_1^d RV_{t-1}^d + c_2^d E_{t-1}[RV_t^w] + \varepsilon_t^d \tag{12}$$

Adding and simplify the above asanas, we can get the HAR-RV model:

$$RV_t^d = c + c_1 RV_{t-1}^d + c_2 RV_{t-1}^w + c_3 RV_{t-1}^m + \varepsilon_t \tag{13}$$

The HAR-RV model can be estimated on the model through simple linear regression methods, such as the ordinary least squares method, and the estimated partial regression coefficients respectively represent the marginal contribution of the behavior of different types of traders to the whole volatility, so it has some economic significance. Although the form of the model is relatively simple, many scholars have found that the HAR-RV models can well characterize the long memory characteristics of the volatility [27].

4. In-Sample analysis

In this section, we first performed descriptive statistics on variables and understood their characteristics based on their mean, standard deviation, maximum and minimum values. Then the parameters of the HAR-RV model are estimated, and the significance of variables for volatility prediction is discussed. Finally, analyse the fitting ability of the model.

4.1 Summary statistics

According to the descriptive statistical analysis of the following main variables (Table 1), the daily fluctuations of copper futures prices range from 0.0049 to 54.1572, with large fluctuations. The weekly fluctuation range of copper futures prices is 0.0029 to 0.6974, and the monthly fluctuation range is 0.0038 to 0.2143. This shows that the weekly and monthly fluctuations of copper futures prices are relatively small. In addition, the mean and standard deviation of daily volatility is much larger than the mean and standard deviation of weekly volatility and monthly volatility, which is consistent with the fluctuation range of daily volatility, weekly volatility, and monthly volatility, which proves the price of copper futures. Weekly volatility and monthly volatility are relatively small, while daily volatility is relatively large.

Table 1. Descriptive statistics of variables

Variables	Mean	Std. dev	Min	Max
RV_d	0.9698	2.8411	0.0049	54.1572
RV_w	0.0307	0.0663	0.0029	0.6974
RV_m	0.0288	0.0361	0.0038	0.2143
Vix	20.1221	10.1551	10.16	82.69

4.2 Parameter estimations

In this section, we have constructed the HAR-RV model, HAR-RV-V model, HAR-RV-W model, HAR-RV-WV model and used OLS to estimate the parameters of the above four models.

4.2.1 Parameter estimation of HAR-RV model:

The parameter estimation results of the HAR-RV model (Table 2) show that in the 1-day volatility forecast, the daily volatility is significantly negative at the 1% significance level, and the monthly volatility is significant at the 10% significance level. If it is negative, the weekly volatility is significantly positive at the 1% significance level. In the 1-week volatility forecast, the weekly volatility and the monthly volatility are significantly positive at a significant level of 1%. In contrast, the daily volatility is not meaningful enough for the 1-week volatility forecast. Only the monthly volatility was significantly positive at the 1% significance level in the January volatility forecast, and the daily volatility and weekly volatility did not have sufficient predictive significance. The results show that the medium-term (weekly) volatility and long-term (monthly) volatility of the copper futures market contain a large amount of forecast information about RV, but only the long-term (monthly) volatility has predictive significance in the January volatility forecast, and the short-term (daily) volatility can only predict the 1-day copper futures market.

4.2.2 Parameter estimation of HAR-RV-V model:

The estimation results of the HAR-RV-V model show that in the 1-day volatility forecast, the addition of the VIX index has a significant positive impact on the weekly volatility and monthly volatility; in the 1-week volatility forecast, the weekly volatility and the monthly volatility has a significant negative impact; in the January volatility forecast, it has a significant positive impact on the weekly volatility, but has a negative impact on the daily volatility and monthly volatility. The results show that most of the coefficients of the HAR-RV-V model are consistent with the HAR-RV model, which further supports our results and conclusions. However, there is still a certain impact on the value of the coefficients, which shows that we cannot ignore the VIX. The predictive significance of the index on the volatility of copper futures prices.

4.2.3 Parameter estimation of HAR-RV-W model:

Estimates of the HAR-RV-W model show significant positive effects on weekly and monthly fluctuations and significant Friday effects, and in month fluctuations. Still, they have no significant effect on Japanese and weekly fluctuations. The results show that the HAR-RV-W model is roughly consistent with the HAR-RV model. It further supports our results and conclusions, but still has some influence, indicating that we still need to pay attention to the predictive significance of the intra-week effect on the price fluctuation index of copper futures.

Table 2. Parameter estimation results of HAR-RV-type models.

	HAR-RV			HAR-RV-V			HAR-RV-W			HAR-RV-WV		
	1-day	1-week	1-month	1-day	1-week	1-month	1-day	1-week	1-month	1-day	1-week	1-month
β_0	0.357* ** (3.76)	- 0.000 02 (-0.03)	0.000 15 (1.15)	0.736* ** (2.97)	- 0.004* * (-2.31)	0.002* ** (5.09)	0.476* ** (2.47)	-0.001 (0.87)	0.000 26 (0.96)	0.865* ** (2.70)	-0.002 (-1.05)	0.002 (3.94)
$\ln RV$	- 0.169* ** (-5.29)	0.000 05 (0.26)	- 0.000 01 (-0.38)	- 0.176* ** (-4.69)	0.0000 6 (0.26)	- 0.0004 (-0.79)	- 0.171* ** (-5.28)	0.000 02 (0.13)	0.000 02 (-0.47)	- 0.179* ** (-4.74)	0.0000 2 (0.10)	- 0.0000 4 (-0.78)
$\ln RV$	30.903 *** (19.17)	0.952 *** (83.57)	0.001 6 (0.71)	32.307 *** (15.90)	0.939* ** (65.64)	0.008* ** (2.99)	30.934 *** (19.13)	0.953 *** (83.51)	0.001 7 (0.74)	32.387 *** (15.87)	0.941* ** (65.64)	0.008* ** (2.96)

$\ln RV$	-	0.05*	0.994	-4.578	0.022	1.003*	-	0.05*	0.995	-4.524	0.022	1.003*
	5.907*	**	***	(-1.42)	(0.96)	**	5.808*	**	***	(-1.40)	(0.97)	**
	(-2.34)	(2.83)	(275.67)			(228.78)	(-2.30)	(2.80)	(274.51)			(228.20)
VIX				-	0.0003	-				-	0.0003	-
				0.026*	***	0.0001				0.026*	***	0.0001
				(-1.86)	(2.67)	(-5.31)				(-1.87)	(2.64)	(-5.29)
Tue							-0.137	0.000	-	-0.167	0.0003	-
							(-0.54)	9	3	(-0.53)	(0.12)	0.0003
								(0.53)	(-0.88)			(-0.66)
Thu							0.075	-0.002	0.000	0.056	-0.003	0.0003
							(0.30)	(-1.29)	1	(0.18)	(-1.35)	(0.63)
Fri							-0.505**	-	0.000	-0.363	-	0.0000
							(-1.97)	0.003	04	(-1.14)	0.004*	8
								(-2.06)	(0.12)		(-1.87)	(0.19)
Sta							-0.062	-0.001	-	0.222	0.002	-
							(-0.23)	(-0.75)	2	(0.15)	(-0.18)	0.0003
									(-0.54)			(-0.13)
N	956			699			956			699		
Adj. R ²	0.355	0.941	0.992	0.354	0.942	0.993	0.355	0.941	0.992	0.352	0.942	0.992

Note: t statistics in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

4.2.4 Parameter estimation of HAR-RV-WV model:

Estimates of the HAR-RV-WV model show that the addition of the VIX index and weekly and monthly effects; significant Friday effects, and in January fluctuation prediction, significant positive effects on weekly and monthly fluctuations, but negatively on Japanese fluctuations. The results show that the HAR-RV-WV model is roughly consistent with the HAR-RV model. It further supports our results and conclusions, but still has some influence, indicating that we still need to pay attention to the predictive significance of the intra-week effect on the price fluctuation index of copper futures.

To judge the fitting ability of the model, Table 1 shows the coefficient of determination of the HAR-RV model, HAR-RV-V model, HAR-RV-W model, HAR-RV-WV model in 1 day, 1 week, and 1 month. $Adj. R^2$. There is no big difference in the coefficients of determination of these four models. Both the $adj. R^2$ of 1 week, and January tends to 1. Therefore, these four models are predicting the results of 1 week and January. The fitting ability is better when the volatility is used, but the fitting ability is poor when predicting the one-day volatility.

5. Robustness test for in-sample regression

In this section, we test the robustness of within-sample regression prediction. Since the prediction results are similar to Section 5 and Section 6, we only discuss the robustness results in this section.

To test whether the prediction results are still credible under different samples, we divide 956 samples into two sub-samples. Subsample 1 contains samples from 1 to 478, and subsample 2 contains samples from 478 to 956. Then, we perform intra-sample regression on subsample 1 and subsample 2 and calculate the coefficient of determination $Adj. R^2$ to test the prediction accuracy of the model. The results are shown in Table 3.

It can be seen from Table 3 that when the HAR-RV model is used to estimate short-term and mid-term volatility, the coefficient of determination $Adj. R^2$ of sub-sample 1 and sub-sample 2 is significantly different, and the robustness is poor. In the long-term volatility estimation, the coefficient of determination $Adj. R^2$ between the two sub-samples is almost the same, and the HAR-RV model is relatively stable in predicting long-term volatility.

Table 3. Adjusted R-squares of sample 1 and sample 2

	1-day		1-week		1-month	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
HAR-RV	0.2698	0.3483	0.8373	0.9392	0.9811	0.9906
HAR-RV-V	0.2576	0.3518	0.8622	0.9411	0.9797	0.9919
HAR-RV-W	0.3255	0.3453	0.8401	0.9393	0.9906	0.9816
HAR-RV-WV	0.3110	0.3452	0.8650	0.9409	0.9798	0.9918

Note: different sample's adjusted R-squares have been shown in this table. When the value is larger, it means the forecasting results are better

6. Conclusion

This paper uses the HAR-RV model HAR-RV-V model, HAR-RV-W model, and HAR-RV-WV model to use five-minute high-frequency data to predict the volatility of the copper futures market. The data shows that the weekly and monthly volatility of copper futures prices are relatively small. In contrast, the daily volatility is relatively large, so the prediction of this study The model is more accurate in predicting long-term volatility. The HAR-RV model is a first-order autoregressive model established for the realized volatility of different time intervals. Through observation of historical volatility, We have further built three new models by adding investor sentiment VIX index and Sunday effect to improve copper futures price volatility prediction accuracy.

Through the in-sample analysis, this article can find that the weekly and monthly fluctuations of copper futures prices are relatively small, while the daily fluctuations are relatively large. At the same time, the same problem can be found when using the HAR-RV model, HAR-RV-V model, HAR-RV-W model, and HAR-RV-WV model to predict the price of copper futures. That is, the above four models are predicting 1. In the case of monthly volatility, only long-term (monthly) volatility has predictive significance. This phenomenon also exists in predicting one-week volatility and one-day volatility. When predicting volatility in a predetermined period, the volatility is shorter than the period. There is not enough predictive significance. In terms of prediction accuracy, we found that these four models have the better fitting ability when predicting the volatility of one week and one month but poor fitting ability when predicting the volatility of one day.

Although this paper has achieved some results, there are still some shortcomings. The specific manifestation is that the overfitting situation is not discussed when evaluating the model's fitting ability. At the same time, because of the few variables set, the model's ability to predict copper futures prices needs further verification. In the next step of the study, structural mutations will be introduced based on the HAR-RV model to further improve the model's accuracy. High-frequency volatility usually has structural mutations, and the impact of structural mutations will also affect price fluctuations in the futures market.

References

- [1] Wang Jiheng, in the East Preface. China copper futures and spot price correlation study [J]. *Price Monthly*, 2021 (03): 22 - 29.
- [2] Sh. International Copper Futures is about to "break out of the cocoon" [N]. *Futures Daily*, 2020-11-18 (006).
- [3] Tu Wen. The CSRC approved the Shanghai International Energy Trading Center to conduct international copper futures trading [J]. *Nonferrous Metals in China*, 2020 (22): 23.
- [4] Chen Lin, Huang Zhangshu. Futures Price Analysis and Forecast Based on the ARIMA Model [J]. *Journal of Fuzhou University (Philosophy and Social Sciences)*, 2010, 24 (03): 32 - 37.
- [5] Chen Xiaodong. Comparative study of high-frequency volatility prediction model in China Financial Futures Market [J]. *Management Engineering Journal*, 2016 (3): 114 - 119.
- [6] Han Liyan, Yin Libo. Speculative behavior or actual demand? —— General Perspective Analysis of International Commodity Price Impact [J]. *Economic Research*, 2012 (12): 83 - 96.
- [7] ANDERSEN T G, BOLLERSLEV T. Answering the critics: yes, ARCH models do provide good volatility forecasts [R]. National Bureau of Economic Research, 1997.
- [8] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, 71 (2): 579 - 625.
- [9] Corsi F.A simple approximate long-memory model of realized volatility [J]. *Journal of Financial Econometrics*, 2009, 7 (2): 174 - 196.
- [10] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X. Roughing it up: Including jump components in the measurement, Modeling, and forecasting of return volatility [J]. *The Review of Economics and Statistics*, 2007, 89 (4): 701 - 720.
- [11] Zhang Xiaoyong, Ren Deping. Study on the volume and price relationship between Shanghai and Shenzhen 300 Stock Index futures based on high-frequency data [J]. *Journal of Hunan University (Social Sciences Edition)*, 2013, 27 (02): 48 - 54.
- [12] Qu Hui, Li Jie, Cheng Xin. Prediction accuracy comparison of HAR and GARCH models —— is based on an empirical analysis of high frequency prices of the CSI 300 index [J]. *System Engineering*, 2015,33 (03): 32 - 37.
- [13] Zhu Xuehong, Zou Jia grain, Han Feiyan, Chen Jinyu. Modeling and prediction of high-frequency volatility of China copper futures market by introducing external impact [J]. *China Management Science*, 2018, 26 (09): 52 - 61.
- [14] Fields M. J. Stock Prices: A Problem in Verification. *Journal of Business* [J]. 1931 (10): 415 - 418.
- [15] Cross F. The Behavior of Stock Prices on Fridays and Mondays [J]. *Financial Analysts Journal*, 1973, 29 (6): 67 - 69.
- [16] French, k. Stock returns and the weekend effect [J]. *Journal of Finance*, 1980, 579 - 596.
- [17] Jaffe J, R Westerfield. The Weekend Effect in Common Stock Returns: The International Evidence [J]. *Journal of Finance*, 1985, 40 (2): 433 - 454.
- [18] Aggarawal R..Seasonal and Day of the Week Effects in Four Emerging Stock Markets [J]. *The Finance*,1985, (40): 433 - 454.
- [19] Kim Karam, Ryu Doojin. Does sentiment determine investor trading behavior? [J]. *Applied Economics Letters*, 2021, 28 (10):
- [20] Chiang, R.and Tapley, T. C. (1983): "The Day of the Week Effect in the Futures market, *Review of Research in Futures Markets*", 2, 356 ~ 410.
- [21] Cornell, B. (1985): "The Weekly Pattern in Stock Returns: Cash Versus Futures: A Note, *Journal of Finance*," 40, 583 ~ 588.
- [22] Gay, G.D.and Kim T. H. (1987): "An Investigation into Seasonality in the Futures Market, *The Journal of Futures Markets*,"7, 169 ~ 181.
- [23] Agrawal, A.and Tandon, K. (1998): "Anomalies or Illusions? Evidence from Stock Markets in Eighteen Countries," *Journal of International Money and Finance*, 13, 83 ~ 106.

- [24] Bowers, J. and Dimson, E. (1998): "Introduction, in: E. Dimson (ed.) *Stock Market Anomalies*," Cambridge University Press, Cambridge, 3 ~ 15.
- [25] He Zhaohui, Liu Qingling. Empirical Analysis of the Impact Factors of copper futures price fluctuation in China [J]. *Mass Investment Guide*, 2020 (21): 11 - 12.
- [26] Wei Yu. Research on the volatility prediction model of Shanghai-Shenzhen 300 stock index futures [J]. *Journal of Management Science*, 2010,13 (02): 66 - 76.
- [27] Dong Dianhua, Zhang Dai. Study on heterogeneity of China Securities Market Based on HAR-RV Model [J]. *Journal of Xi'an Institute of Finance and Economics*, 2011, 24 (05): 30 - 34.
- [28] Li, W., Cheng, Y., & Fang, Q. (2020). Forecast on silver futures linked with structural breaks and day-of-the-week effect. *The North American Journal of Economics and Finance*, 53, 101192.