Forecast on gold futures linked with investor sentiment and S&P500 index

Yuan Lu[1,a,†], Zhurong Geng[2,b,†]

[1]School of Mathematics, South China University of Technology, Guangzhou, China

[2]Inner Mongolia University of Finance and Economics, Hohhot, Inner Mongolia Province, China

[a]201830430467@mail.scut.edu.cn, [b]barbara.mize@students.clatsopcc.edu

[†]These authors contributed equally.

Keywords: Investor sentiment, HAR-type models, Gold futures market, Volatility forecasting

Abstract: This paper discusses the relationship between investor sentiment and the realized volatility of gold futures, which is investigated using high-frequency data. For investor sentiment factors, we select four indicators including the volatility index (VIX), the volume, inventory and turnover rate of gold future. To improve our forecasting accuracy, stock market factor is introduced and we select S&P500 index to represent it. Based on the heterogeneous autoregressive (HAR) theory, six new heterogeneous autoregressive (HAR) models are established by combining investor sentiment and S&P500 index. The empirical results show that the accuracy of the new model is better than that of the original HAR model. We found that S&P500 index contain a lot of gold prediction information. In addition, the investor sentiment has a positive impact on the volatility of gold futures. Our work is the first to combine investor sentiment with the S&P500 index to identify more market information. This paper provides a better forecasting method for the Volatility Prediction of gold futures.

1. Introduction

Gold is the simple substance form of the chemical element Au, which is a kind of noble metal with soft texture, yellow luster and corrosion resistance. Due to its good physical properties, stable chemical properties, high ductility and scarcity, it plays an important role as a material in the jewelry industry, electronics industry, modern communications, aerospace industry and other sectors. At the same time, it is also regarded as a special currency, used for savings and investment. The code of gold defined by people in finance is XAU or GOLD. In addition, the source of Au is Aurora, the goddess of dawn in Roman mythology, which means shining dawn. Gold has undergone five changes across the world, from the restoration of the standard to the virtual gold standard, from the Bretton Woods system to the Jamaica Agreement. It has always been the first choice for international savings in various countries, for which is an important part of the world financial system. The main reason why the gold market so indispensable is that it is a special commodity and currency with currency, commodity and financial attributes at the same time. As a general commodity, it can meet the needs of industry and commerce; as a stable currency metal, it is the best storage method and value object; as a world currency, it is the most effective international reserve recognized by governments and central banks; as a means of payment, it can be used not only as an international means of purchase, but also as an international means of payment and a means of international wealth transfer. In addition, the continuous development and soundness of the gold market is conducive to the balance and stability of the financial market. The gold market has played a balanced role in the currency market, foreign exchange market, capital market, and insurance market. After the opening and development of various financial markets, a situation of diversification of financial assets and diversification of investment channels has formed. Various financial markets can restrain and restrict each other to form a balanced situation, which can stabilize market conditions and avoid major fluctuations. In addition, due to the stability of the gold market, investors' participation in gold market transactions can serve as a booster and lubricant for the normal operation of the financial market, which is important for the coordinated and stable development of the money market, securities market, insurance market and foreign exchange.

As a symbol of world wealth, it has an important position in the financial system; people are keen on studying gold and having extremely rich research results. In recent academic journals, more and more people have begun to consider the impact of investor sentiment in financial activities, such as the role of investor sentiment in the two-way impact of S&P500 and gold [1]. The influence of sentiment indicators on the gold futures market is obvious. Therefore, consideration of sentiment indicators is beneficial to improving the accuracy of forecasts.

In recent relevant research, people usually use the Vector autoregression (VAR) model analysis, when several traditional methods were unable to accurately define and measure financial risks. The G30 Group published a report entitled "Practices and Rules for Derivative Products" in 1993 based on the study of derivatives, proposed the VAR to measure market risks, which has become the mainstream method of measuring market risk in the financial world [2]. Then the Risk Metrics risk control model introduced by John Pierpoint. Morgan to calculate VAR is widely adopted by many financial institutions. Since the traditional Asset & Liability Management relies (ALM) too much on report analysis and lacks timeliness; measures risk is too abstract with variance and beta coefficient and reflects only the volatility of the market (or assets); Capital Asset Pricing Model (CAPM), at the same time, cannot be combined financial derivatives.

1993 was the first time to study the nonlinear relationship between gold price and stock market index by applying Markov-switched Bayes VAR model, which contributed to the literature [3]. VAR model can not only calculate a single financial instrument risks, but also calculate the risk of a portfolio consisted of multiple financial instruments, which cannot be achieved by traditional financial risk management. It measures risk concisely and clearly, unifies risk measurement standards, and is easier for managers and investors to understand and master. However, VAR is flawed. VAR uses implicit assumptions inherent in the model. At the same time, there are certain flaws in the use of data and its principles and statistical estimation methods. In fact, it can only be used to study conventional market conditions.

Therefore, in order to solve the problem of the VAR model and upgrade and optimize it, subsequent researchers evaluated the impact of gold and oil price fluctuations on the volatility of the South African stock market and its constituent indexes or sectors (i.e., financial, industrial, and resource sectors). The vector autoregressive asymmetric dynamic conditionally correlated generalized autoregressive conditional heteroscedasticity (VAR-ADCC-GARCH) model is used. The results show that there is a significant volatility spillover between gold and the stock market or between the oil and stock market. This shows the importance of the link between the commodity market and the stock market, the significant of combining gold and stocks as the best strategy for hedging stock risks, especially during the financial crisis [4]. With advancements of technology, people have new ideas about forecasting methods. The concept of Artificial neural networks is proposed, and the results obtained by the ANN method are compared with the results obtained by the econometric forecasting method of VAR. The results show that the artificial neural network method has better predictive ability than the VAR method [5].

Both the regression-based generalized autoregressive conditional heteroscedasticity (GARCH) model and the heterogeneous autoregressive (HAR) model based on high-frequency data can make up for the shortcomings of the previous model; the HAR model performs better [6, 7]. The HAR model, which has obvious advantages in calculating accurate data, is also widely used. In this article, the precise calculation of high-frequency data is also involved with intention, so we constructed a research model related to the HAR model.

When conducting predictive analysis, people usually choose to use the Auto Regressive Moving Average model [8]. However, because this model is a study on stationary data modeling, more and more people have put forward new ideas for ARMA model prediction. Proposed a high-order fuzzy ARMA(p,q) time series solving algorithm based on fuzzy logic group relations, which includes fuzzy MA variables and fuzzy AR variables [9]. At the same time, in the process of volatility research, people will use the improved autoregressive conditional jump intensity (ARJI) model [10] and the extended autoregressive moving average generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) method [11]. GARCH's error Variance was further modeled. It is especially suitable for the analysis and prediction of volatility. Such analysis can play a very important guiding role in the decision-making of investors; and its significance often exceeds the analysis and prediction of the value itself.

Although the world's increasingly connected world today has continuously improved its ability to respond to the financial crisis, under the influence of the COVID-19 epidemic, the gold market in 2020 will be greatly impacted. Following the 2015 bull market cycle of gold, it is inevitable that the price of gold, which should have been in an upward development stage, will fall under the influence of the epidemic. Moreover, in the current complex world, the fluctuations of gold futures are obvious but the trend is unknown. The fluctuations in the gold market will inevitably have a profound impact. Therefore, the research of this article is very important.

This article considers that the New York gold market, as currently the world's largest ones with the largest trading volume, can greatly affect gold prices. The US gold market consists of five exchanges including the New York Mercantile Exchange (NYMEX), Chicago International Mercantile Exchange (IMM), Detroit, San Francisco, and Buffalo, with gold futures trading mainly. Due to the New York's economic status in the United States and even the world, gold trading on the New York Mercantile Exchange has the greatest influence in the US. COMEX gold trading can often dominate the trend of global gold prices and is one of the most important pricing centers for global gold futures. Therefore, we choose 5-minute high-frequency data of US COMEX gold trading for further research and analysis.

2. Data

In consideration of the accuracy of the results, in this study we selected the 5-minute high-frequency data of COMEX gold futures in New York and S&P500, under the assumption that the data interval would directly affect the accuracy of the research results. If the data frequency is too high, there will be a smaller amount of interference. On the contrary, if the data frequency is too low, it cannot express all the market information, for which we would not get the desired result. For the current research we obtained pricing data for March 28, 2018 to September 26, 2020. By calculating the Realised Varience, median, quartile, etc., and eliminating invalid or empty data, we finally got 606 available valid trading days data points.

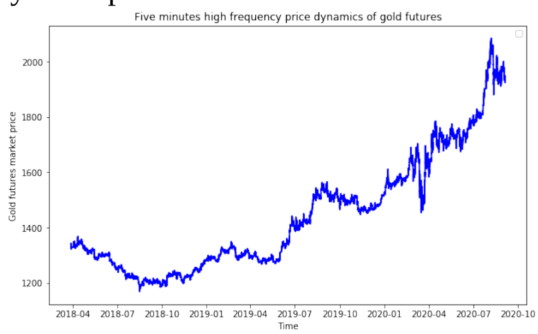

Fig.1. Five minutes high frequency price dynamics of gold futures.

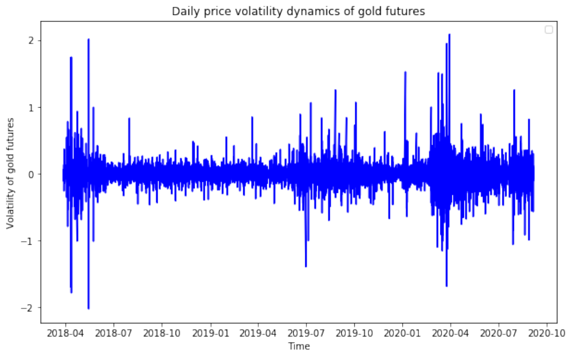

Fig.2. Daily price volatility dynamics of gold futures.

As shown in the figure1 and figure 2, we can see that the price fluctuations of gold futures from the end of March 2018 to the end of December 2019 are basically stable, but from February 2020, the data of COMEX gold fluctuates sharply. Since 2020, COVID-19, a new type of coronavirus, has spread all around the world, for which the impact of this outburst on the economy is huge. The COVID-19 epidemic has gradually developed worldwide from February 2020. The U.S ushered in a large-scale outbreak of the epidemic in March. Since then, the epidemic has been affecting the U.S financial market and escalated again in August 2020. The long-term epidemic has hit the U.S. financial market hardly. Even oil industry, which is as important as gold, suffered severely resulting in a negative trading price. At same time, the per-share market of oil has broken down many times. The U.S. economy has declined. As an important part of the U.S. financial economy, gold futures also have a similar difficult fate in this epidemic. The price fluctuation range of the five-minute high-frequency data in the figure basically confirms this fact. Therefore, the 2020 gold futures data is important for our research, due to highly uncertain and unstable, is irregular and difficult to predict the results.

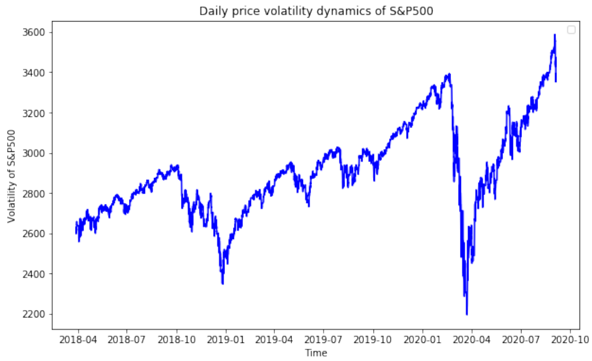

Fig.3. Daily price volatility dynamics of S&P500

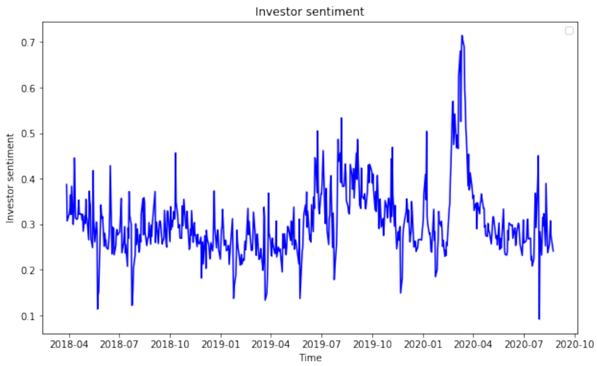

Fig. 4. Investor sentiment

According to Figure 3 and Figure 4 above, we can see that both the S&P500 index and the investor sentiment factor that we built fluctuate greatly during this period. The S&P500 fell greatly during the early period of COVID-19 breakthrough but gradually grew up to about 3000 points. As for the investor sentiment, the higher it goes, the more panic investors feel and the more unstable the market is. It reached its peak of this period at about the time when the COVID-19 broke out in the U.S. and it is reasonable that investors felt panic at that time.

3. Entropy weight method

Entropy was first introduced into information theory by Shannon, and has been widely used in engineering technology, social economy and other fields [12]. According to the explanation of the basic principles of information theory, information is a measure of the degree of order of the system, and entropy is a measure of the degree of disorder of the system; according to the definition of information entropy, for an index, entropy can be used to judge the degree of dispersion of an index , The smaller the information entropy value, the greater the degree of dispersion of the index, the greater the impact of the index on the comprehensive evaluation (i.e., the weight), if the values of an index are all equal, the index will not play a role in the comprehensive evaluation. Therefore, the tool of information entropy can be used to calculate the weight of each indicator to provide a basis for comprehensive evaluation of multiple indicators. In this article, this method is used when confirming the sentiment index ratio. The calculation process is provided as follows.

3.1 Data standardization

In the evaluation system of multiple indicators, due to the differences in the nature of the indicators, if the data with different dimensions and orders of magnitude are directly calculated, then results would become biased when the level difference is very large. Therefore, when we use the entropy method to calculate the weight of sentiment indicators, we must first standardize the obtained market data to ensure the reliability of the results.

In this article, the min-max standardization method is used. The min-max standardization method is to linearly transform the original data. Suppose minA and maxA are the minimum and maximum values of attribute A respectively, and an original value x of A is standardized by min-max and mapped to the value x'in the interval [0,1]. The formula is:

$$Y_{ij} = \frac{X_{ij} - min(X_i)}{max(X_i) - min(X_i)} \quad (1)$$

where Y_{ij} is normalized value and X_{ij} is VIX, volume, inventory and turnover rate.

3.2 Proportion of various indicators

The second step is to calculate the proportion of each data in the overall data. The formula used is:

$$Y_{ip} = \frac{x'_{ip}}{\sum_{i=1}^{m} x'_{ip}} \quad (2)$$

where Y_{ip} is the proportion of P index.

3.3 Determine the information entropy of the P index

Entropy in information theory, also called information entropy, is used to measure the degree of uncertainty of a random variable. The greater the entropy, the greater the uncertainty.

$$E_p = -ln(m)^{-1} \sum_{i=1}^{m} (Y_{ip} * ln\, Y_{ip}) \; (0 \leq E_p \leq 1) \quad (3)$$

where E_p is the information entropy of the P index.

3.4 Information entropy redundancy of the P index

The next step is to calculate the information entropy redundancy of the P index. The formula is shown below:

$$D_p = 1 - E_p \quad (4)$$

where D_p is Information entropy redundancy of the P index and E_p is the information entropy of the P index.

3.5 Determine the weight of the P index

After the above steps, we can finally obtain more accurate and reliable indicator weight values for subsequent research.

$$W_p = \frac{D_p}{\sum_{p=1}^{n} D_p} \quad (5)$$

where W_p is the weight of the P index.

4. Economic models

4.1 Har-RV-type models

In this part, in order to predict the volatility of gold futures market, we introduce eight HAR-type models, including two original models, HAR-RV and HAR-CJ, as benchmark models and six improved HAR-type models, namely HAR-RV-I, HAR-RV-S, HAR-RV-S-I, HAR-CJ-I, HAR-CJ-S, HAR-CJ-I, HAR-CJ-S-I, with considering factors that affect the volatility of gold and S&P500 as well as investor sentiment.

We first introduce the classical HAR-RV model, including linear form and logarithmic form to determine whether RV contains a lot of information. Then, by considering the VIX investor sentiment, turnover rate, inventory as well as volume of COMEX gold futures, the HAR-RV-I model can be developed. In addition, the influence of volatility of S&P500 should be considered and we develop the HAR-RV-S model in order to improve the prediction accuracy. Finally, we take all of the above factors into account simultaneously so that an improved HAR-RV-S-I model is calculated and established.

Also, to analyze the effect of continuous sample path variation and discontinuous jump mutation, the original HAR-CJ model is proposed to predict the volatility. Similarly, by considering investor sentiment, the HAR-CJ-I model is built. In addition, we develop the HAR-CJ-S model to predict the volatility of gold futures and volatility of S&P500. Finally, a HAR-CJ-S-I model is proposed and all of the above factors are included.

Next, we describe these models in detail.

1) HAR-RV model

According to the RV calculation method proposed by Andersen and Bollerslev (1998) [13], we divide the trading day into M segments, and the i^{th} closing price of trading day t is expressed as $P_{t,i}$. The RV of the trading day t is expressed as RV_t^d, which can be expressed as

$$RV_t^d = \sum_{i=1}^{M} r_{t,i}^2 \tag{6}$$

where $r_{t,i}$ is the logarithmic rate of return for the i^{th} period of the trading day t.

$r_{t,i}$ is magnified 100 times for easy observation; this can be expressed as

$$r_{t,i} = \left(lnP_{t,i} - lnP_{t,i-1}\right) * 100 \tag{7}$$

Thus the weekly RV and the monthly RV of the trading day t, denoted as RV_t^w and RV_t^m, respectively, are defined as follow.

$$RV_t^w = \frac{RV_t^d + RV_{t-1}^d + RV_{t-2}^d + RV_{t-3}^d + RV_{t-4}^d + RV_{t-5}^d}{6} \tag{8}$$

$$RV_t^m = \frac{RV_t^d + RV_{t-1}^d + RV_{t-2}^d + \cdots + RV_{t-24}^d}{25} \tag{9}$$

The average RV from the day t to $(t+H)$ is defined as

$$\overline{RV_{t+H}} = \frac{1}{H}\sum_{i=1}^{H} RV_{t+i}^d \tag{10}$$

In addition, (Andersen et al., 2011)[14] discovered that the logarithmic form of the HAR model performs better than the linear HAR model. Therefore, we use the former one to forecast the price volatility of gold futures. The logarithmic form of the HAR-RV model is expressed as

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_d lnRV_t^d + \beta_w lnRV_t^w + \beta_m lnRV_t^m \tag{11}$$

2) HAR-RV-I model

As mentioned above, we use the entropy weight method to assign values to the sentiment indicators.

This is mainly because the entropy weight method greatly optimizes the reliability of the data indicators. In addition, we select four daily variables, VIX, inventory, turnover rate and volume of COMEX gold futures as components of our investor sentiment and calculate the proportion for each parameter. After standardizing them we get the investor sentiment I_t for each day t.

$$I_t = \alpha_{1t}VIX_t + \alpha_{2t}volume_t + \alpha_{3t}inventory_t + \alpha_{4t}turnover - rate_t \quad (12)$$

And the model is

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_d lnRV_t^d + \beta_w lnRV_t^w + \beta_m lnRV_t^m + \beta_i I_t \quad (13)$$

3) HAR-RV-S model

Besides investor sentiment, we also consider the stocks factor and choose the S&P500 as parameter. Using the same method in HAR-RV model, we calculate its RV for each day and use it as S_t in the model

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_d lnRV_t^d + \beta_w lnRV_t^w + \beta_m lnRV_t^m + \beta_s S_t \quad (14)$$

4) HAR-RV-S-I model

In order to test the effect of all the factors that affect the volatility prediction of gold futures market, we develop a class of HAR-RV-S-I model by incorporating structural breakthrough, investor sentiment and volatility of S&P500 in corresponding original HAR-type model. Therefore, the logarithmic form of HAR-RV-S-I model, can be expressed as follow:

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_d lnRV_t^d + \beta_w lnRV_t^w + \beta_m lnRV_t^m + \beta_i I_t + \beta_s S_t \quad (15)$$

4.2 HAR-CJ-type models

1) HAR-CJ model

In the real financial market, due to factors such as information shocks and investor irrationality, the volatility of return on assets is no longer continuous, and jumping fluctuations will exist. Using the quadratic variation theory proposed by Andersen and Bollerslev [13], we decompose the RV into two parts: the continuous sample path variance and the discrete jump variance, which are defined as continuous volatility (CV) and jump volatility (JV), respectively.

In discrete prices, the logarithmic yield volatility is no longer an unbiased estimator of the integral variation (IV), which usually also includes JV components. The logarithmic yield can be expressed as the sum of the integral volatility and the JV after the quadratic variation. The quadratic variation (QV) denoting the logarithmic rate of return from the trading day t -1 to t is defined as

$$QV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<x\leq t} \kappa_s^2 \quad (16)$$

Since the quadratic variation cannot be observed directly, Andersen and Bollerslev [13] pointed out that, when using the discrete return data to estimate the quadratic variation and the sample size tends to infinity, the RV is the consistent estimator of the quadratic variation, which can be represented as

$$RV_t^d \xrightarrow{N\to\infty} QV_t \quad (17)$$

Additionally, Barndorff-Nielsen and Shephard [14, 15] pointed out that the integral volatility can be measured using the realized bi-power variation (RBV). When the sample size tends to infinity, the RBV becomes the consistent estimate of the CV. The RBV on day t, denoted as RBV_t, can be defined as

$$RBV_t = z_1^{-2} \frac{N}{N-2} \sum_{j=3}^{N} |r_{t,j-2}| \, |r_{t,j}| \quad (18)$$

where z_1 is defined as

$$z_1 = E(Z_t) = \sqrt{\pi/2} \quad (19)$$

where Z_t is a random variable subject to the standard normal distribution.

We use the Z_t statistic proposed by Huang and Tauchen [16] to test whether there is a JV component in the RV. According to Andersen et al. [6], in the case where the significance level is an estimate of the daily JV of the trading day t can be obtained.

$$JV_t^d = I(Z_t > \alpha)(RV_t^d - RBV_t) \tag{20}$$

Accordingly, the daily CV of the trading day t is defined as

$$CV_t^d = I(Z_t \leq \alpha)RV_t^d + I(Z_t > \alpha)RBV_t \tag{21}$$

The weekly continuous JV and monthly continuous JV of trading day t are expressed as follows:

$$CV_t^w = \frac{CV_t^d + CV_{t-1}^d + CV_{t-2}^d + CV_{t-3}^d + CV_{t-4}^d + CV_{t-5}^d}{6} \tag{22}$$

$$JV_t^w = \frac{JV_t^d + JV_{t-1}^d + JV_{t-2}^d + JV_{t-3}^d + JV_{t-4}^d + JV_{t-5}^d}{6} \tag{23}$$

$$CV_t^m = \frac{CV_t^d + CV_{t-1}^d + CV_{t-2}^d + \cdots + CV_{t-24}^d}{25} \tag{24}$$

$$JV_t^m = \frac{JV_t^d + JV_{t-1}^d + JV_{t-2}^d + \cdots + JV_{t-24}^d}{25} \tag{25}$$

The logarithmic form of the HAR-CJ model is expressed as

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_{cd} lnCV_t^d + \beta_{cw} lnCV_t^w + \beta_{cm} lnCV_t^m + \beta_{jd} ln(JV_t^d + 1) + \beta_{jw} ln(JV_t^w + 1) + \beta_{jm} ln(JV_t^m + 1) \tag{26}$$

2) HAR-CJ-I model

Based on the HAR-CJ model, we consider the investor sentiment of the market and the logarithmic form of the HAR-CJ-I model can be established as follow

$$ln\overline{RV_{t+H}} = \beta_0 + \beta_{cd} lnCV_t^d + \beta_{cw} lnCV_t^w + \beta_{cm} lnCV_t^m + \beta_{jd} ln(JV_t^d + 1) + \beta_{jw} ln(JV_t^w + 1) + \beta_{cm} ln(JV_t^m + 1) + \beta_i I_t \tag{27}$$

3) HAR-CJ-S model

Based on the HAR-CJ model, the historical trading volume and price fluctuation of stock market are further introduced into the model. Therefore, like formula 14, the logarithmic form of the HAR-CJ-S model is

$$\begin{aligned} ln\overline{RV_{t+H}} = \beta_0 + \beta_{cd} lnCV_t^d + \beta_{cw} lnCV_t^w + \beta_{cm} lnCV_t^m \\ + \beta_{jd} ln(JV_t^d + 1) + \beta_{jw} ln(JV_t^w + 1) \\ + \beta_{jm} ln(JV_t^m + 1) + \beta_s S_t \end{aligned} \tag{28}$$

4) HAR-CJ-S-I model

Based on the HAR-CJ-S model, the investor sentiment factor is further introduced into the model. In this part, we integrate the above factors to evaluate the volatility of gold futures market. The logarithmic form of HARCJ-S-I model can be expressed as follow

$$\begin{aligned} ln\overline{RV_{t+H}} = \beta_0 + \beta_{cd} lnCV_t^d + \beta_{cw} lnCV_t^w + \beta_{cm} lnCV_t^m \\ + \beta_{jd} ln(JV_t^d + 1) + \beta_{jw} ln(JV_t^w + 1) + \beta_{cm} ln(JV_t^m + 1) + \beta_i I_t + \beta_s S_t \end{aligned} \tag{29}$$

5. In-Sample analysis

5.1 Summary statistics

According to the descriptive statistical analysis of the following main variables, as shown in Table 1, the daily fluctuation range of gold futures price ranges from 0.09 to 19.06, with a wide range and strong volatility. The weekly and monthly RV range from 0.11 to 11.47 and 0.20 to 27.80. S&P500 index fluctuates by more than 90, while investor sentiment fluctuates relatively little, about 0.6. The results support the view that the volatility of stock market is more severe than that of gold prices. The mean and variance of S&P500 index fluctuation are far greater than gold price fluctuation, which indicates that the average daily RV of S&P500 is much higher than that of gold. In addition, the daily volatility difference of stock prices is much larger than that of gold.

Table 1. Descriptive statistics of variables.

Variable	mean	Std.dev	min	max
RV	0.9760	3.0021	0.0995	19.0631
S	2.0885	49.8201	0.02116	98.1114
I	0.3065	0.0065	0.0915	0.7148
$\mathbf{RV_t^w}$	0.8463	1.6432	0.1066	11.4689
$\mathbf{RV_t^m}$	0.9187	3.6029	0.2024	27.8047

5.2 Parameter estimations

In this section, we use the OLS method to estimate the parameters of eight HAR-type models, following [17]. We divided eight HAR-type models into HAR-RV-type models (HAR-RV, HAR-RV-I, HAR-RV-S, HAR-RV-S-I) and HAR-CJ-type models (HAR-CJ, HAR-CJ-S, HAR-CJ-I, HAR-CJ-S-I) to discuss the impact of investor sentiment as well as stock prices effect. We find that the stock prices' effect includes the forecasting information of gold futures. The investor sentiment factor may improve the predictability of gold futures.

1)Parameter estimations of HAR-RV-type models

Table 2 shows the parameter estimates for HAR-RV, HAR-RV-I, HAR-RV-S and HAR-RV-S-I model when forecasting gold futures' price volatility at three different horizons (daily, weekly, monthly).

The results of the HAR-RV models show that the weekly and monthly volatility are significantly positive in the 1-day, 1-week and 1-month volatility forecasts, but only short-term (daily) volatility and long-term(monthly) are highly positive in predicting the one cycle gold futures market. This result means that the gold futures market is heterogeneous.

The results of HAR-RV-S model show that the S&P500 index is capable of improving the forecasting results significantly. All the daily RV and the index are highly significant in the forecasting, and the results are a lot greater than the HAR-RV models.

The estimation results of HAR-RV-I model show that the investor sentiment factor calculated by entropy weight method does not improve the prediction as significant as the S&P500 index. The daily and monthly RV play a significant role in forecasting

The estimation results of HAR-RV-S-I model show that when adding the investor sentiment factor, the regression results all improve in three different horizons. Comprehensive consideration of these factors can improve the accuracy of gold futures forecast.

2) Parameter estimations of HAR-CJ-type models

In the HAR-CJ model, all the coefficients of daily, weekly and monthly continuous sample path variation are significantly positive.

However, for gold futures forecast of 1-day, 1-week and 1-month, the discontinuous jump coefficient is not significant. In other words, the research results show that the continuous sample path change of gold futures market contains more gold prediction information, while discontinuous jump change contains little gold volatility prediction information.

For Table 3, we also test the estimated results of the HAR-CJ-S model, which takes into account the S&P500 index factor. The results show that in 1-day, 1-week and 1-month prediction, most of the continuous sample path coefficient of variation is significant and positive. However, all the discontinuous jump coefficients are very small, which is basically consistent with the results of HAR-RV-S model. Through the analysis of the above results, we come to the conclusion that the importance of S&P500 change cannot be ignored in the prediction of gold futures prices.

The estimation results of HAR-CJ-I model show that continuous sample path change and discontinuous jump change have different effects on gold futures price. Continuous sample path variation contains more goldprediction information than discontinuous jump. The results also show that most of the coefficients of trading volume and volatility of gold futures are significant, which means that they play an important role in the prediction of gold market, which is very similar to the HAR-RV-I model. However, the forecasting effect of crude oil futures on gold futures may not be enough.

For the HAR-CJ-S-I model, the results show that in our analysis using the HAR-CJ-S and HAR-CJ-I models, the S&P500 index and investor sentiment have similar effects, which supports the conclusion that these factors should not be ignored when forecasting the price fluctuation of gold futures. Comprehensive consideration of these factors can improve the accuracy of gold futures forecast.

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

	HAR-RV			**HAR-RV-S**			**HAR-RV-I**			**HAR-RV-S-I**		
	1-day	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***
$\boldsymbol{\beta_0}$	-5.66*** (-134.1)	-5.64*** (-131.0)	-5.59*** (-112.8)	0.681*** (10.2)	0.857*** (12.4)	1.83*** (20.9)	-5.75*** (-30.4)	-5.74*** (-29.5)	-5.65*** (-25.3)	2.13*** (73.6)	2.36*** (79.9)	3.60*** (61.2)
$\boldsymbol{lnRV_t^d}$	0.907*** (12.2)	-0.0940 (-1.24)	-0.0871 (-0.996)	0.889*** (49.4)	-0.112*** (-6.03)	-0.108*** (-4.54)	0.897*** (11.7)	-0.103 (-1.32)	-0.0940 (-1.04)	0.999*** (170.6)	0.00137 (0.229)	0.0285* (2.17)
$\boldsymbol{lnRV_t^w}$	0.0969 (0.846)	1.10*** (9.346)	0.122 (0.901)	-0.0859** (-3.077)	-0.0920* (-2.50)	0.129039 (0.970)	0.0869 (0.747)	1.09*** (9.13)	0.115 (0.838)	0.0129 (1.456)	1.01*** (112.2)	0.0282 (1.57)
$\boldsymbol{lnRV_t^m}$	-0.347*** (-3.652)	-0.347*** (-3.56)	0.631*** (5.63)	0.0207 (0.884)	1.06*** (34.3)	0.618098 (5.610)	-0.344*** (-3.60)	-0.344*** (-3.51)	0.634*** (5.64)	-0.00235 (0.319)	0.0115 (1.53)	1.04*** (69.6)
$\boldsymbol{S}$				0.936*** (96.1)	1.10*** (85.2)	0.001538 (0.243)				0.994*** (314.3)	1.02*** (316.7)	1.17*** (181.5)
$\boldsymbol{I}$							0.281 (0.521)	0.265 (0.180)	0.192 (0.302)	-3.08*** (-72.7)	-3.19*** (-73.7)	-3.75*** (-43.5)
$\boldsymbol{Adj-R}$^2	0.4313	0.3565	0.2532	0.9664	0.9616	0.9448	0.4305	0.3556	0.252	0.9967	0.9963	0.9871

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

Table 3. Parameter estimation results of HAR-CJ-type models.

	HAR-CJ			**HAR-CJ-S**			**HAR-CJ-I**			**HAR-CJ-S-I**		
	1-day	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***	***1-day***	***1-week***	***1-month***
$\boldsymbol{\beta_0}$	-2.61*** (-10.3)	-2.43*** (-9.92)	-1.68*** (-6.14)	0.196 (1.61)	0.343** (3.20)	1.40*** (11.3)	-2.87*** (-9.839)	-2.55*** (-9.01)	-1.72*** (-5.43)	1.047*** (7.95)	1.36*** (13.1)	2.66*** (23.2)
$\boldsymbol{lnJ_t^d}$	1.67 (12.4)	-0.0410 (-0.315)	-0.0112 (-0.077)	1.63*** (28.3)	-0.0860. (-1.697)	-0.0612 (-1.05)	1.62 (11.9)	-0.0638 (-0.480)	-0.0181 (-0.122)	1.75*** (33.1)	0.0617 (1.48)	0.122** (2.66)
$\boldsymbol{lnJ_t^w}$	-0.213 (-0.989)	1.70*** (8.14)	0.0808 (0.346)	-0.300** (-3.25)	1.62*** (19.9)	0.0135*** (-0.144)	-0.284 (-1.295)	1.69*** (7.84)	0.0706 (0.297)	-0.120 (33.1)	1.83*** (27.4)	0.254*** (3.45)
$\boldsymbol{lnJ_t^m}$	-0.715*** (-4.55)	-0.867*** (-5.70)	0.583*** (3.426)	0.103 (1.49)	-0.0612 (-1.01)	1.48*** (21.0)	-0.692*** (-4.40)	-0.856*** (-5.61)	0.586*** (3.43)	0.0916 (1.47)	-0.0743 (-1.51)	1.46*** (27.0)
$\boldsymbol{lnC_t^d}$	0.0383*** (3.89)	-0.00747 (-0.785)	-0.00565 (-0.531)	0.0369*** (8.79)	-0.00881* (-2.38)	-0.00714. (-1.670)	0.0350*** (3.51)	-0.00902 (-0.931)	-0.00612 (-0.565)	0.0454*** (11.8)	0.00131 (0.431)	0.00545 (-0.558)
$\boldsymbol{lnC_t^w}$	0.139 (1.634)	0.142. (1.72)	-0.0316 (-0.342)	0.144*** (3.96)	0.146*** (4.56)	-0.0262 (-0.706)	0.135 (1.584)	0.139. (1.69)	-0.0322 (-0.349)	0.156*** (4.77)	0.161*** (6.20	-0.00831 (-0.290)
$\boldsymbol{lnC_t^m}$	1.43*** (10.9)	1.61*** (12.8)	2.04*** (14.5)	0.181** (2.98)	0.376*** (7.00)	0.669*** (10.8	1.43*** (10.9)	1.61*** (12.7)	2.04*** (14.4)	0.125* (2.27)	0.308*** (7.09)	0.585*** (12.2)
$\boldsymbol{S}$				0.897*** (50.8)	0.881*** (56.8)	0.982*** (54.7)				0.950*** (57.4)	0.948*** (72.5)	1.06*** (73.6)
$\boldsymbol{I}$							0.845 (1.77)	0.403 (0.869)	0.123 (0.236)	-2.23*** (-11.6)	-2.66*** (-17.6)	-3.31*** (-19.8)
$\boldsymbol{Adj-R}$^2	0.5477	0.5448	0.5	0.9175	0.9311	0.9193	0.5493	0.5446	0.4992	0.9331	0.9551	0.952

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

5.3 Comparison of the in-sample fitting capacity

We discuss the effect of investor sentiment and S&P500 index on the gold futures market. Mid-term and long-term price volatility information is contained within these factors. However, whether the established models can improve the accuracy of in-sample analysis still needs to be examined. In this part, we test the results by using Adjusted R-squares' methods and compare the different models that we have developed. Table 4 shows the different Adjusted R-squares of the models proposed above divided into three groups: 1-day, 1-week, and 1-month. As the table shows, when performing in-sample analysis, if added S&P500 index factor and investor sentiment factor, both HAR-CJ models and HAR-RV models have a better performance, but the results vary. Decomposing the RV into CV and JV can improve the fit of the HAR model to some extent, as the HAR-CJ and HAR-CJ-I model perform better than the HAR-RV and HAR-RV-I model. Moreover, when considering S&P500 factor, the established models behave better from the perspective of Adjusted R-squares than do the original HAR-RV and HAR-CJ models (Adjusted R-squares of the HAR-RV-I, HAR-RV-S, HAR-RV-S-I, HAR-CJ-I, HAR-CJ-S, HAR-CJ-S-U models are better than are those of the HAR-RV and HAR-CJ models, in that they contain more predictive information). We will discuss the influences of these individual factors in detail.

Table 4 also describes the different adjusted R-square of HAR-RV-type models and HAR-CJ-type models in 1-day, 1-week, and 1-month of gold futures' forecasting. If the adj. R^2 is high, the model performs better. A more detailed discussion is provided in the following. The HAR-RV-S and HAR-CJ-S models, respectively, improve the fit of the HAR-RV and HAR-CJ models in forecasting the price volatility of gold futures, and the fit improvement is greater than that brought from the HAR-RV-I and HAR-CJ-I models to the HAR-RV and HAR-CJ models. Compared with the HAR-RV-I and HAR-CJ-I models, the fit of the HAR-RV-S-I and HAR-CJ-S-I models is improved significantly. Whereas, compared with the HAR-RV-S and HAR-CJ-S models, we find only a little improvement. This shows that the introduction of S&P500 index can improve the ability of the HAR models to explain the price volatility of gold futures. Specifically, historical trading volume, the volatility and turnover rate of gold and VIX contain limited forecasting information.

Compared with the fit of HAR-RV, HAR-RV-S, HAR-CJ and HAR-CJ-S models, the fit of HAR-RV-I, HAR-RV-S-I, HAR-CJ-I and HAR-CJ-S-I models improved significantly in 1-day volatility forecasting. Therefore, although historical trading volume and the volatility of gold futures can partly explain the volatility changes in gold futures, the ability to interpret short-term volatility changes increases significantly after the introduction of S&P500 index and investor sentiment factor. Combined with previous analysis, it can be speculated that the short-term volatility changes in gold futures prices are affected greatly by the S&P500 index. No matter how long the forecasting period is, the HAR-R-S-I model always has the best forecasting performance.

Table 4. Adjusted R-squares of HAR-RV-type and HAR-CJ-type models.

	1-day	**1-week**	**1-month**
HAR-RV	0.4313	0.3565	0.2532
HAR-RV-I	0.4305	0.3556	0.2520
HAR-RV-S	0.9664	0.9616	0.9448
HAR-RV-S-I	0.9967	0.9963	0.9871
HAR-CJ	0.5477	0.5448	0.5000
HAR-CJ-I	0.5493	0.5446	0.4992
HAR-CJ-S	0.9175	0.9311	0.9193
HAR-CJ-S-I	0.9331	0.9551	0.9520

6. Loss function to out-of-sample forecasting

To evaluate the impact of investor sentiment and S&P500 index on gold markets, we also perform out-of-sample forecasting. Data from April 11, 2018 to August 24, 2020, in a total of 583 observations are collected and we obtain corresponding forecasting values for the same period.

We estimate the parameters of the eight models and calculate the respective forecast values with the rolling window estimation. The loss function evaluation method is the basic way to know which model's forecasting effect is better. If the loss function number of the model is smaller, the model has better forecasting behavior. We select four loss functions, including the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), the Root Mean Square Error (RMSE), and the Theil Inequality Coefficient (TIC), to calculate the loss function values and compare the forecasting loss of each model. The four loss functions are defined as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |RV_t - \widehat{RV}_t| \tag{30}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|RV_t - \widehat{RV}_t|}{RV_t} \tag{31}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(RV_t - \widehat{RV}_t)^2} \tag{32}$$

$$Theil\ inequality\ coefficient = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(RV_t-\widehat{RV}_t)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(RV_t)^2}+\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\widehat{RV}_t)^2}} \quad (33)$$

where RV_t is the real value of the RV, and $\widehat{RV}_t$ is the predicted value of the RV.

The loss function values (As seen from Table 5) show that, except for the MAPE computed from 1-day volatility forecasting, the remaining loss function values reduce gradually as the forecasting period increases and the historical trading volume, the volatility of both gold and crude oil, structural breaks, and the day-of-the-week effect are introduced in models. Generally, the predicted loss of the HAR-CJ-S-I model is the smallest in 1-day volatility forecasting; the predicted loss of the HAR-RV-S-I and HAR-CJ-S-I models are relatively small in 1-week volatility forecasting; and when the forecasting period increases to 1 month, the forecasting loss of the HAR-CJ-S-I model and the HAR-RV-S-I model, are relatively small.

Table 5. Loss function values for out-of-sample forecasting (window is 583).

		MAE	MAPE	RMSE	TIC
1-day	***HAR-RV***	0.978491	0.995607	2.009039	0.995364
	HAR-RV-I	0.978487	0.995606	2.009000	0.995362
	HAR-RV-S	0.975354	0.993358	2.001660	0.978772
	HAR-RV-S-I	0.974567	0.992952	2.000089	0.973668
	HAR-CJ	0.976547	0.994859	2.007209	0.991631
	HAR-CJ-I	0.976569	0.994878	2.007257	0.991783
	HAR-CJ-S	0.973002	0.993020	1.996562	0.969463
	HAR-CJ-S-I	**0.972158**	**0.992703**	**1.994266**	**0.963021**
1-week	***HAR-RV***	0.841017	0.995561	1.543548	0.995263
	HAR-RV-I	0.841013	0.995560	1.543510	0.995220
	HAR-RV-S	0.838745	0.993109	1.540980	0.985464
	HAR-RV-S-I	0.838201	0.992351	1.540815	0.982289
	HAR-CJ	0.838151	0.994948	1.541454	0.993928
	HAR-CJ-I	0.838155	0.994952	1.541440	0.993935
	HAR-CJ-S	0.836188	0.993050	1.537998	0.984098
	HAR-CJ-S-I	0.835739	0.992722	1.537621	0.981064
1-month	***HAR-RV***	0.920541	0.995302	2.123891	0.996320
	HAR-RV-I	0.920540	0.995302	2.123903	0.9996326
	HAR-RV-S	0.917111	0.991411	2.121920	0.988294
	HAR-RV-S-I	0.916257	**0.990537**	2.121114	0.985491
	HAR-CJ	0.873816	0.994425	1.787889	0.994259
	HAR-CJ-I	0.873819	0.994427	1.787901	0.994277
	HAR-CJ-S	0.871180	0.991532	1.784756	0.986181
	HAR-CJ-S-I	**0.870543**	0.990903	**1.783738**	**0.983355**

Notes: in this table, the bold number means the best evaluation in the loss function among 1-day,1-week and 1-month. The smaller of the number, the better of the model performs.

Table 6. Adjusted R-squares of sample 1 and sample 2.

	1-day		**1-week**		**1-month**	
	Sample1	***Sample2***	***Sample1***	***Sample2***	***Sample1***	***Sample2***
HAR-RV	0.8242	0.9620	0.8163	0.9536	0.8136	0.9413
HAR-RV-I	0.8242	0.9713	0.8162	0.9649	0.8134	0.9555
HAR-RV-S	0.9799	0.9889	0.9787	0.9864	0.9668	0.9827
HAR-RV-S-I	0.9991	0.9969	0.9989	0.9962	0.9887	0.9951
HAR-CJ	0.7528	0.9132	0.8050	0.9052	0.8272	0.9208
HAR-CJ-I	0.7688	0.9204	0.8106	0.9124	0.8265	0.9273
HAR-CJ-S	0.9462	0.9241	0.9740	0.9195	0.9758	0.9361
HAR-CJ-S-I	0.9524	0.9244	0.9889	0.9225	0.9949	0.9418

7. Robustness test for in-sample regression

In this part, we change the sample range and test the robustness with in-sample regression. Since the prediction results are similar to those in section 5 and 6, we only discuss the robustness results in this section.

In order to test whether the prediction results are still credible in different samples, we divide 583 samples into two sub samples. Subsample 1 contains samples from 1 to 292, and subsample 2 contains samples from 293 to 583. Then, we perform intra sample regression on subsample 1 and subsample 2, and calculate the adjusted R square to test the prediction accuracy of the model. The results are shown in Table 6.

Table 6 shows that when the HAR-RV and HAR-CJ models are used for estimation, the adjusted R-squared values of subsample 1 and subsample 2 are quite different and lack of robustness. When using HAR-RV-S and HAR-CJ-S models to estimate short-term volatility, the adjusted R-squared difference between subsample 1 and subsample 2 is significant, and the robustness is poor. In the medium- and long-term volatility estimation, the adjusted R square is almost the same between the two sub samples, which shows that each model is relatively stable.

Using HAR-RV-S, HAR-RV-S-I, HAR-CJ-S and HAR-CJ-S-I models, the adjusted R-squares of subsample 1 and subsample 2 are estimated and the results are basically the same. Therefore, we can conclude that these four models are robust in the short-term, medium- and long-term volatility estimation of gold futures.

8. Conclusion

This paper mainly uses 5-minute high frequency data to forecast the volatility of gold futures market. On the basis of HAR-type models, the influence of investor sentiment and S&P500 index on gold futures market is considered.

Based on the HAR-RV and HAR-CJ models, we first study the S&P500 index factor, then investor sentiment factor is introduced into the model. Therefore, we have established the HAR-RV-I, HAR-RV-S, HAR-RV-S-I, HAR-CJ-S, HAR-CJ-I, HAR-CJ-S-I models, which are used for the in-sample analysis of gold futures price RV. Some conclusions are put forward.

In the in-sample analysis prediction, the model with S&P500 index performs better than the model without S&P500 index. However, the model considering investor sentiment has little improvement on the fitting degree and prediction accuracy. It has a positive impact on the price fluctuation of gold futures. In terms of prediction accuracy, HAR-RV-S-I and HAR-CJ-S-I provide the most powerful explanation for gold futures price. We provide a new perspective of analysis and further explore the specific impact of gold futures price on various factors by using the HAR-type models. This is in consistency with common sense that when there is great shock in stock market, people tend to turn to precious metal to decrease the instability of the market.

In this paper, the improved HAR-type models consider the investor sentiment and S&P500 index, which significantly improves the robustness and effectiveness of the prediction. This is conducive to the improvement of gold futures market function and comprehensive risk management and is an important supplement to the existing literature. The shortcomings of this paper mainly include two aspects. First, we have to sacrifice the simplicity of some of the new models (for example, the HAR-RV-S-I and HAR-CJ-S-I models) to take all factors into account and make them a little more complicated. Further decomposition of some volatility characteristics of gold futures will lead to further decomposition of some components of volatility. At present, many researchers are considering the impact of leverage effect, spillover effect and geopolitical risk on gold price fluctuation. In the future, we will combine leverage effect and geopolitical risk into the model to predict the future volatility of gold price.

References

[1] Piñeiro-Chousa, J., López-Cabarcos, M.Á., Pérez-Pico, A.M., Ribeiro-Navarrete, B., 2018. Does

social network sentiment influence the relationship between the S&P 500 and gold returns? Int. Rev. Financ. Anal. 57, 57–64. https://doi.org/10.1016/j.irfa.2018.02.005

[2] G30 Group 1993 "Practice and Rules of Derivative Products"

[3] Akgül, I., Bildirici, M., Özdemir, S., 2015. Evaluating the Nonlinear Linkage between Gold Prices and Stock Market Index Using Markov-Switching Bayesian VAR Models. Proc. 4th Int. Conf. Leadersh. Technol. Innov. Bus. Manag. ICLTIBM-2014 210, 408–415. https://doi.org/10.1016/j.sbspro.2015.11.388

[4] Morema, K., Bonga-Bonga, L., 2020. The impact of oil and gold price fluctuations on the South African equity market: Volatility spillovers and financial policy implications. Resour. Policy 68, 101740. https://doi.org/10.1016/j.resourpol.2020.101740

[5] Aydin, A.D., Cavdar, S.C., 2015. Comparison of Prediction Performances of Artificial Neural Network (ANN) and Vector Autoregressive (VAR) Models by Using the Macroeconomic Variables of Gold Prices, Borsa Istanbul (BIST) 100 Index and US Dollar-Turkish Lira (USD/TRY) Exchange Rates. IISES 3rd 4th Econ. Finance Conf. 30, 3–14. https://doi.org/10.1016/S2212-5671(15)01249-6

[6] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), Modeling and Forecasting Realized Volatility. Econometrica, 71: 579-625. https://doi.org/10.1111/1468-0262.00418

[7] Gaoxiu Qiao, Yuxin Teng, Weiping Li, Wenwen Liu,Improving volatility forecasting based on Chinese volatility index information: Evidence from CSI 300 index and futures markets,The North American Journal of Economics and Finance,Volume 49,2019,Pages 133-151,ISSN 1062-9408,https://doi.org/10.1016/j.najef.2019.04.003.

[8] Ray, B.K., 1993. Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. Int. J. Forecast. 9, 255–269. https://doi.org/10.1016/0169-2070(93)90009-C

[9] Kocak, C., 2017. ARMA(p,q) type high order fuzzy time series forecast method based on fuzzy logic relations. Appl. Soft Comput. 58, 92–103. https://doi.org/10.1016/j.asoc.2017.04.021

[10] Zhang, C., Shi, X., Yu, D., 2018. The effect of global oil price shocks on China's precious metals market: A comparative analysis of gold and platinum. J. Clean. Prod. 186, 652–661. https://doi.org/10.1016/j.jclepro.2018.03.154

[11] C. E. Shannon, "A mathematical theory of communication," in The Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, July 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[12] Andersen, Torben G., and Tim Bollerslev. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." International Economic Review, vol. 39, no. 4, 1998, pp. 885–905. JSTOR, www.jstor.org/stable/2527343. Accessed 10 Nov. 2020.

[13] Andersen, T.G., Bollerslev, T., Huang, X., 2011. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. Realiz. Volatility 160, 176–189. https://doi.org/10.1016/j.jeconom.2010.03.029

[14] Barndorff-Nielsen, O. E., & Shephard, N. (2004). Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in

[15] Financial Economics. Econometrica, 72(3), 885–925. https://doi.org/10.1016/j.jeconom.2005.01.009

[16] Barndorff-Nielsen, O. E., & Shephard, N. (2006). Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes. Journal of

[17] Econometrics, 131(1–2), 217–252. https://doi.org/10.1016/j.jeconom.2005.01.009

[18] Huang, X., & Tauchen, G. (2005). The Relative Contribution of Jumps to Total Price Variance 3 DOI:10.1093/jjfinec/nbi025.

[19] Corsi, F., 2009. A Simple Approximate Long-Memory Model of Realized Volatility. J. Financ. Econom. 7, 174–196. https://doi.org/10.1093/jjfinec/nbp001.