Application of ARIMA-GARCH Model in Venture Capital Market Prediction

Jicheng Wang[*], Xiaolong Hu

School of Computer Science, Honhyi Honor College, Wuhan University, Wuhan, Hubei, 430000

[*]Corresponding author: wangjicheng220322@163.com

Keywords: Autoregressive integrated moving average (ARIMA) model; generalized autoregressive conditional heteroscedasticity (GARCH) model; modern portfolio theory (MPT).

Abstract: This paper discusses the autoregressive integrated moving average (ARIMA) model, generalized autoregressive conditional heteroscedasticity (GARCH) model, and their application in risk asset market forecasting. Specifically, we investigated variance error using the GARCH model. In addition, we use ARIMA model to forecast future price trends. Finally, we investigate optimizing risk-return trade-offs in diversified portfolios using modern portfolio theory (MPT).

1. Introduction

Traders in the financial market buy and sell volatile assets frequently. Their only goal is to maximize their total return while bearing the minimum risk. In this paper, we have developed a time series model that uses only the data of past daily prices to determine each day what the trader should do to adjust their assets in their portfolio. In this paper, we also consider the transaction cost of each transaction with commission. We consider one risk-free asset, cash, and two risky assets, gold, and bitcoin.

An autoregressive integrated moving average (ARIMA) model is a statistical analysis model, which uses time-series data to better understand the data set or predict future trends. A statistical model is autoregressive if it predicts future values based on past values [1]. The generalized autoregressive conditional heteroscedasticity (GARCH) is a statistical model, which is used to analyze the time-series data that the variance error is considered as the autocorrelation of the series. The GARCH models assume that the variance of the error term follows an autoregressive moving average process. While ARIMA works on price level or returns, GARCH tries to model the clustering in volatility or squared returns [2].

Traditional econometric models, such as regression analysis and time series analysis, usually assume constant variance when explaining the volatility of returns of risky assets [3]. But a large number of empirical studies show that this assumption is not plausible and cannot fit financial data, because financial time series data is unstable, with large fluctuations in a certain period while small fluctuations in other time periods, resulting in clustering of fluctuations. Therefore, the first order or multi-order difference is used to make it stable for the unstable raw data.

Investors care about certain investments' expected return and risk when building their portfolio.

In this paper, we describe the investment returns and risks in terms of the numerical characteristics: the expectation and variance of portfolio returns. ARIMA-GARCH model is adopted to estimate the expected return rate and variance. On the one hand, we use the ARIMA model to estimate the expected return rate. On the other hand, we use the GARCH model to evaluate the variance of the return rate.

2. Time series analysis

2.1 ARIMA (q, e, r) Model

An ARIMA model is a statistical analysis model that uses time-series data to better understand the data set or predict future trends. A statistical model is autoregressive if it predicts future values based on past values. An ARIMA model can be understood by outlining each of its components as follows:

• Auto regression (AR): refers to a model that regresses a variable on its own lagged term.

• Integrated (I): substitute the adjacent difference of the data values for its raw observation to make the time series stationary.

• Moving average (MA): refers to the dependency between an observation and a residual error from a moving average model applied to lag observations.

ARIMA modeling is essentially an exploratory data-oriented approach with the flexibility to adapt to the structural appropriateness of the data itself and the model [4]. With the help of autocorrelation function and partial autocorrelation function, the stochastic properties of time series can be approximately simulated. From it, we can find information such as trend, random change, periodic component, periodic pattern, sequence correlation and so on. A prediction of the future value of the sequence can be easily obtained with some precision, which can be expressed as:

$$r_n = \sum_{i=1}^{p} \gamma_i \cdot r_{n-i} + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{n-i} + \varepsilon_n \tag{1}$$

Where ε_i is the stationary white noise with a mean value of 0 and variance ofσ^2, p and q are the order of autoregression model and moving average model, respectively. This model is named as an autoregressive moving average sequence with order p of model autoregressive and order q of the model moving average, or ARMA (P, Q) series for short. γ_i Is autoregressive coefficient, θ_i is the moving average coefficient, and these are all parameters to be estimated. ARMA model can only deal with time series of stationary processes. If the nonstationary time series is to be analyzed, it must be made stationary. The most common and simplest method is to perform a difference operation on the original non-stationary time series, that is, to obtain an ARIMA (p, d, *and* q) model, where d is the order of difference. In the ARIMA model, the trend term is extracted through difference operation on time series and transformed into the stationary term. Then, the ARMA model is estimated and transformed after estimation to adapt to the original sequence model before difference operation. $p(q)$ Is the lag order of the AR model (MA model), respectively. And d is the number of times that the raw observations are differenced, also known as the degree of difference.

2.2 GARCH Model

Essentially, observations do not conform to a linear pattern wherever there is heteroscedasticity. Instead, they tend to cluster. Therefore, if statistical models that assume constant variance are used on this data, then the conclusions and predictive value one can draw from the model will not be reliable.

In 1986, BOLLERSLEV, based on ARCH model, introduced the lag terms into conditional variance and obtained the generalized ARCH model, GARCH model. GARCH model is a statistical model used in analyzing time-series data where the variance error is believed to be serially auto correlated. GARCH models assume that the variance of the error term follows an autoregressive moving average process.

The mean equation can be expressed as:

$$r_n = m + \varepsilon_n \quad \varepsilon_n \sim N(0, \sigma_n^2) \tag{2}$$

The variance equation can be expressed as:

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{w} \alpha_i \cdot \varepsilon_{n-i}^2 + \sum_{i=1}^{s} \beta_i \cdot \sigma_{n-i}^2 \tag{3}$$

Where $\sum_{i=1}^{w} \alpha_i + \sum_{i=1}^{s} \beta_i < 1, \alpha_0 > 0, \alpha_i \geq 0, i = 1,2, \ldots w, \beta_i \geq 0, i = 1,2,3, \ldots s.$

2.3 ARIMA-GARCH model

Combining the ARIMA *(p, d,* an*d q)* model with GARCH *(w, s)* model, we can get an ARIMA-GARCH model to estimate the function of risky assets. While ARIMA works on price level or returns, GARCH model tries to model the clustering in volatility or squared returns. As the discrete version of Stochastic Volatility model, GARCH also captures the fat-tail effect in risky asset markets. Therefore combining ARIMA with GARCH is expected to better fit in modeling volatile prices data than one model alone.

3. Model Analysis

Establishing an ARIMA-GARCH model to predict the rate return mainly includes the following steps:

1. Conduct a correlation test on the time series model to check whether autocorrelation exists between data.
2. Use the ADF unit root test to test the stationarity of time series data.
3. The non-stationary sequence is differentially processed, and the original time series is converted into the stationary sequence.
4. Use AIC (Akaike Information Criterion) to order the model and estimate the parameters.
5. ARCH test is performed on the selected ARIMA model to determine whether the model has heteroscedasticity. If heteroscedasticity exists, the GARCH model should be established.
6. Residual test is performed on the established ARIMA-GARCH model. Check whether the residual term conforms to the white noise process. If not, it indicates that there is still relevant information in the residual term that has not been extracted, further improving the model.
7. Forecast the return rate by using the established model.

Suppose tomorrow is the nth day, and we define tomorrow's return rate as a random variabler_n. For risky assets gold and bitcoin, we separately use the ARIMA-GARCH model to analyze their expected return and its variance. We can obtain the expected return rate $E(r_n) = \sum_{i=1}^{p} \gamma_i \cdot r_{n-i}$ and its variance$Var(r_n) = \sigma_n^2$. After we get these two numerical characteristic of random variable r_n, we can build our optimal portfolio based on these values. From our ARIMA-GARCH model, we could estimate tomorrow's conditional standard deviation given all the history data. We will show the results in Fig.1 and Fig. 2.

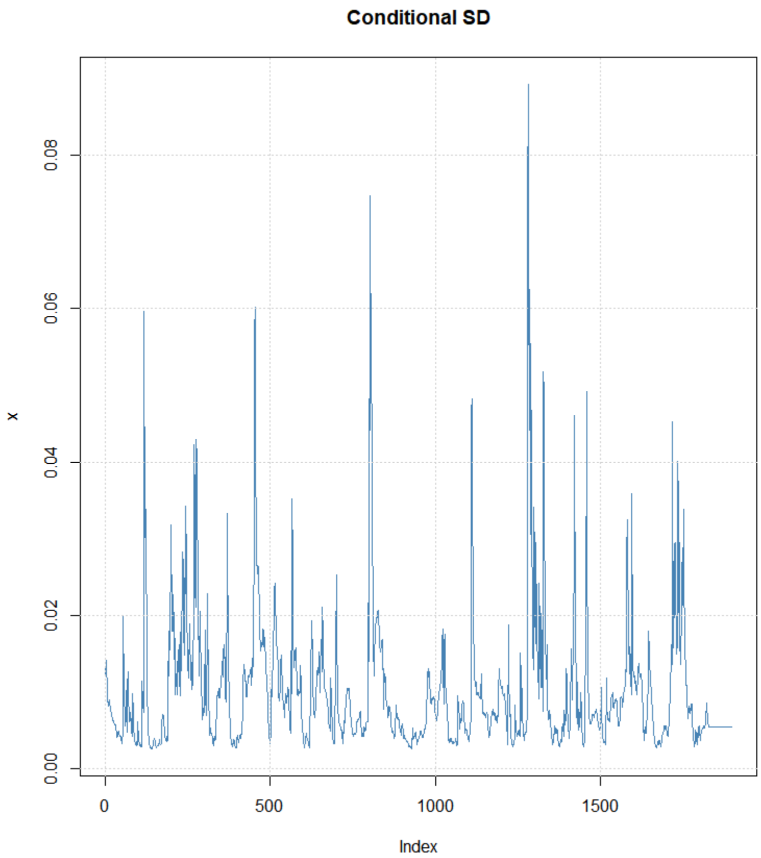

Figure 1. Bitcoin's conditional standard deviation.

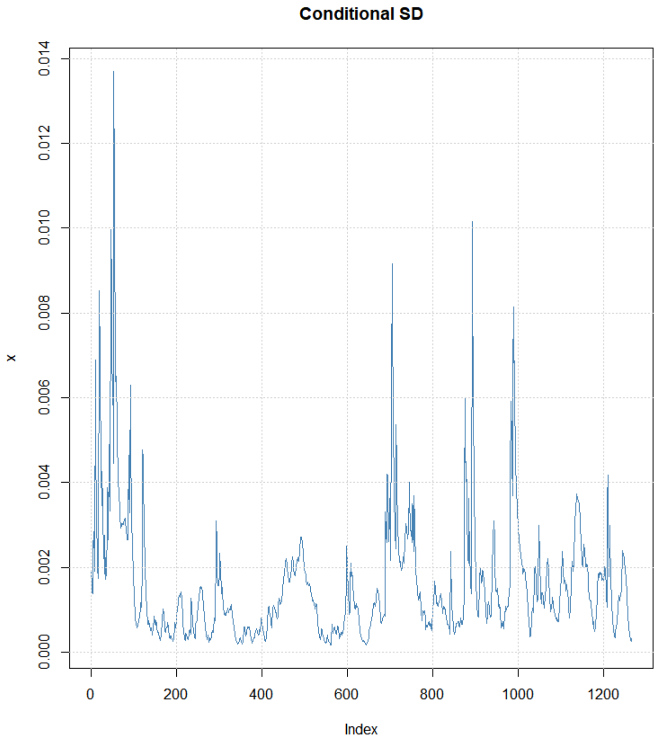

Figure 2. Gold’s conditional standard deviation.

4. Model Improvements

Most investments are either high risk and high return or low risk and low return. Markowitz (1952) argued that investors could achieve their best results by choosing an optimal mix of the two based on an assessment of their tolerance to risk. We adopt the modern portfolio theory (MPT) to select the optimal portfolio from cash, gold and bitcoin. The MPT is a useful tool for investors trying to build diversified portfolios. It is also a practical method for selecting investments to maximize their overall returns within an acceptable level of risk.

Selecting a portfolio may be divided into two stages [5]. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities, which is the main work of this part. The second stage starts with the relevant beliefs about future performances and ends with the portfolio choice. This section is concerned with the second stage.

We consider the rule that the investor should consider expected return a desirable thing and variance of return an undesirable thing since we have assumed that investors are risk-averse. Portfolio construction is usually viewed as comprising two broad tasks:

• The allocation of the overall portfolio to safe assets, such as cash in our model, versus risky assets such as gold and bitcoin in our model.

• The determination of the composition of the risky portion of the complete portfolio.

We will build a portfolio only consisting of 2 risky assets (gold and bitcoin). Let $w_G(w_B)$ be the weight of the portfolio’s value invested in asset gold (bitcoin) such that $w_G + w_B = 1$. and from the results of ARIMA-GARCH model, we can know the expected return of gold (bitcoin), denoted by $\mu_G(\mu_B)$.for simplicity, we denote$w = (w_B, w_G)^T, \mu = (\mu_G, \mu_B)^T$.

The expected return rate of the risky asset portfolio is calculated as a weighted sum of the returns of the individual assets, which can be expressed as:

$$E(r_p) = w^T \mu \tag{4}$$

The portfolio’s risk is a function of the variances of each asset and the correlations of each pair of assets, which can be expressed as:

$$\sigma_p^2 = w^T \sum w \quad (5)$$

Where Σ is the covariance matrix of the asset returns.

Usually, the portfolio with maximum expected return is not the one with minimum variance. There is a rate at which the investor can gain expected return by taking on variance, or reduce variance by giving up expected return. Therefore, we are going to find the optimal weight vector w that maximizes expected return rate per standard deviation, which can be expressed as:

$$\max_{w} \frac{w^T \mu}{\sqrt{w^T \sum w}} \quad (6)$$

Where $w_G + w_B = 1$.

We denote the solution of this optimization problem as w^*, which is the optimal weight invested in risky assets.

In the upper part, we have derived the optimal weight vectorw^*, which determines composition of the risky portion of the complete portfolio. Next, we are going to determine the proportion of risky assets to the overall portfolio, which is called the capital allocation decision. The optimal capital allocation depends partly on the risk–return trade-off the risky port-folio gives. But it also depends on the investor's attitude toward risk, so we need to measure and describe risk aversion

Therefore, a utility function is established to capture the risk aversion, which can be used to rank portfolios with different expected returns and levels of risk. By selecting the overall portfolio with the highest utility, investors optimize the trade-off between risk and return, i.e., they achieve the optimal allocation of capital to risky versus risk-free assets.

We will assume that each investor can assign a utility score to compete portfolios on the basis of the expected return and risk of those portfolios. Higher utility scores are assigned to portfolios with more attractive risk-return profiles. Portfolios receive higher utility scores for higher expected returns and lower scores for higher volatility. The utility function that has been employed in our model assigns a portfolio with expected return $E(\mathrm{r})$, the variance of returns σ^2 and cost c the following utility:

$$U = E(\mathrm{r}) - \frac{1}{2} A\sigma^2 - \mathrm{c} \quad (7)$$

Where U is the utility value, and A is a parameter of the investor's risk aversion.

With a proportion z in the risky portfolio, and 1 – y in cash, the expected rate of return on the complete portfolio can be determined by:

$$E(\mathrm{r}) = \mathrm{y}w^{*T}\mu \quad (8)$$

Variance of return rate of the complete portfolio can be expressed as:

$$\sigma^2 = y^2 w^{*T} \sum w^* \quad (9)$$

The commission cost can be expressed as:

$$\mathrm{c} = \alpha^T |yw^* - y_0 w_0| \quad (10)$$

Then, Investor's utility maximization problem can be expressed as:

$$\max_{y} U = yw^{*T}\mu - \frac{1}{2} Ay^2 w^{*T} \sum w^* - \alpha^T |yw^* - y_0 w_0| \quad (11)$$

Usually, A ranges from 2.0 to 4.0, here, we adopt the value A = 3.

We will apply our portfolio theory to the problem, and the change of value of our portfolio is shown in Fig. 3. Besides, our choice of the proportion of value invested in risky assets (namely, y) is shown in Fig. 4.

We have also compared our model with the popular LSTM model (Long Short-Term Memory model). We find that our model has better accuracy in predicting future prices. We will put some predictions of bitcoin and gold by the LSTM model in Fig.5 and Fig. 6, respectively. From Fig.5 and Fig. 6, we can obviously see that the LSTM model predicts the future price with large errors. Also, because we have adopted the modern portfolio theory, which will beat all other portfolioss almost surely, our strategy is sure to be the best strategy, given only the historical price data.

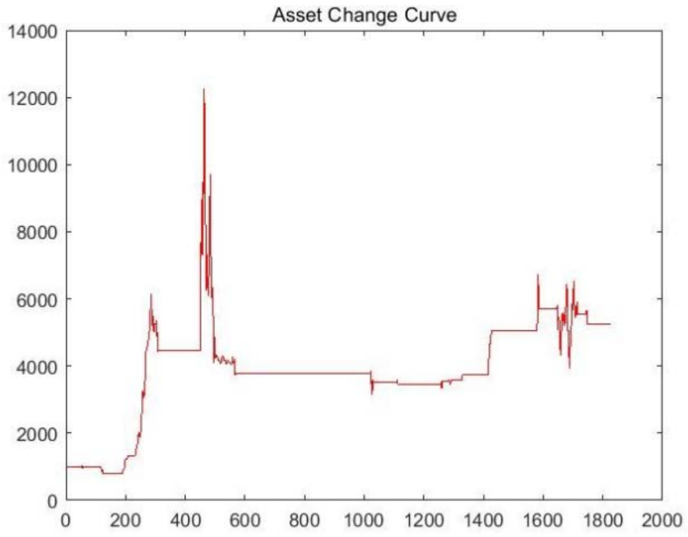

Figure 3. Change of asset value.

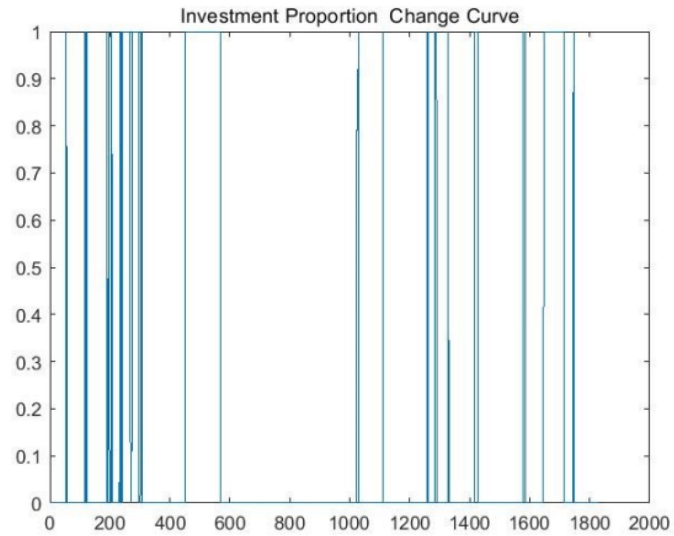

Figure 4. Proportion of value invested in risky assets.

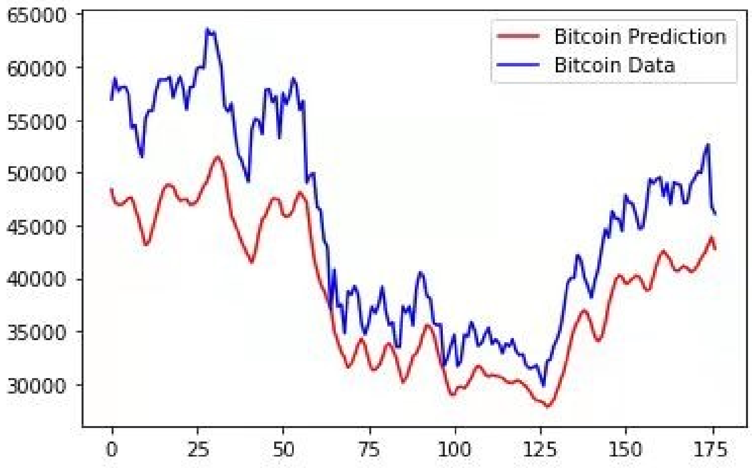

Figure 5. Prediction of price of bitcoin by LSTM model.

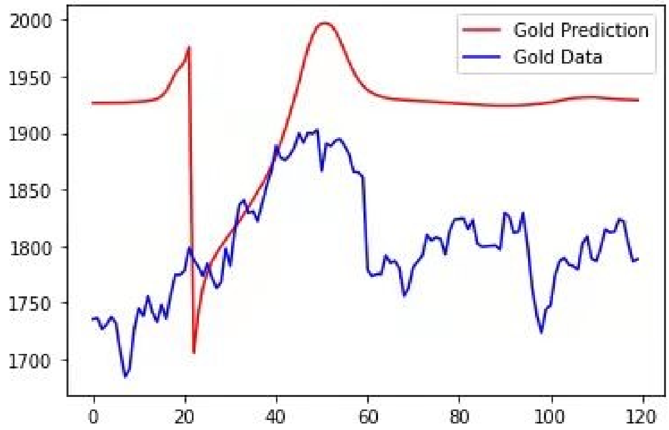

Figure 6. Prediction of price of gold by LSTM model.

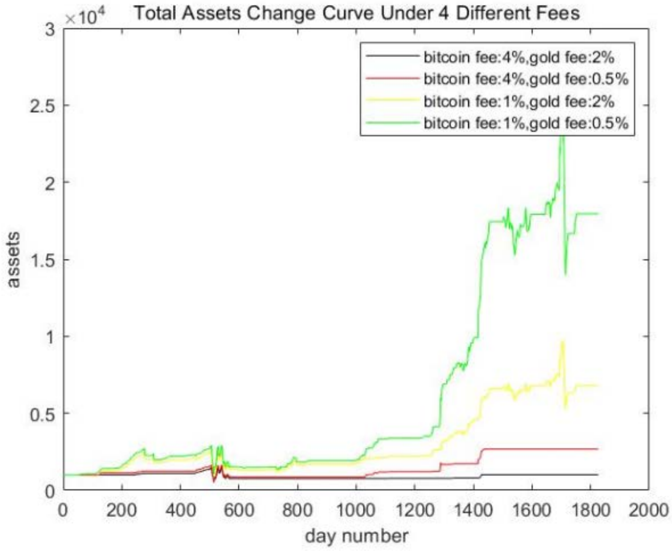

Figure 7. Total values of our optimal portfolio.

5. Sensitivity Analysis

Sensitivity Analysis is a tool used in financial modeling to analyze how the different values of a set of independent variables affect a specific dependent variable under certain specific conditions. It is especially useful in studying and analyzing a 'Black Box Process' where the output is an opaque function of several inputs. An opaque function or process is one that can't be studied and analyzed for some reason. For example, climate models in geography are usually very complex. As a result, the exact relationship between the inputs and outputs are not well understood.

This paper uses sensitivity analysis to study the effect of a small change in commission cost on asset value. We have tested our model by changing the numerical value of commission cost α_{gold} and $\alpha_{bitcoin}$, from $(\alpha_{gold}, \alpha_{bitcoin}) = (1\%, 2\%)$ to $(0.5\%, 1\%)$, $(2\%, 4\%)(0.5\%, 4\%)$ and$(2\%, 1\%)$, respectively. We will plot the curves of total values of our optimal portfolio in these four different conditions, shown in Fig.7.

6. Conclusion

In this paper, we establish ARIMA-GARCH model to predict the expectation and variance of the rate return. While ARIMA works on price level or returns, GARCH tries to model the clustering in volatility or squared returns. As the discrete version of Stochastic Volatility model, GARCH also captures the fat-tail effect in risky asset markets. Therefore, combining ARIMA with GARCH is expected to better fit in modeling volatile prices data than one model alone. Combining the ARIMA (p, d, q) model with GARCH (w, s) model, and we can get an ARIMA-GARCH model to estimate the function of risky assets.

References

[1] M. F. Hayati, I. M. Tahir, Modelling and forecasting S&P 500 stock prices using hybrid arima-garch model, Journal of Physics: Conference Series, 2019, 1366

[2] A. M. Shahidi, A study based on the arima-garch model of bitcoin daily earnings forecasting, Journal of Innovation and Social Science Research, 2019, 6(10).

[3] S. Sitorus, Investment decision making based on value at risk (VAR) analysis for stocks of state own bank in Indonesia, Journal of Economics and Business,2018,2(2).

[4] H. M. Markowitz, Portfolio selection, The Journal of Finance, 1952, 7(1)

[5] A. Georgantas, M. Doumpos, C. Zopounidis, Robust optimization approaches for portfolio selection: a comparative analysis, Annals of Operations Research, 2021(prepublish).