Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 7, pp. 1381–1391
DOI : 10.47974/JSMS-1214

Gold market risk evaluations using GARCH incorporate with machine learning

Lee Yong Xin
Chin Wen Cheong *
Department of Mathematics
Xiamen University Malaysia
43900 Sepang
Selangor Darul Ehsan
Malaysia

Gloria Teng Ai Hui
School of Mathematical Sciences
University of Nottingham Malaysia
Jalan Broga
43500 Semenyih
Selangor Darul Ehsan
Malaysia

Lim Min
Department of Mathematics
Xiamen University Malaysia
43900 Sepang
Selangor Darul Ehsan
Malaysia

Abstract

This paper utilizes the Support Vector Regression (SVR) and Artificial Neural Network (ANN) integrated with a GARCH model in analyzing volatility within the gold market. We used the root of mean square error to compare the performance between the econometric model and various ML-GARCH models in forecasting the stock price of COMEX Gold Futures. SVR model with RBF kernel is found to be the most successful model in predicting the future stock prices of COMEX Gold Futures with an extremely low RMSE value among

* *E-mail:* wcchin@xmu.edu.my

the machine learning models. For the market risk evaluations, we found that the gold future market become less volatile during the COVID-19 pandemic as compared to before pandemic.

Subject Classification: *91B25, 91B84, 92B20.*

Keywords: *Machine learning, Neural network, Support vector regression, Volatility analysis, Value at risk.*

1.0 Introduction

Volatility prediction of financial markets is one of the major concerns of data scientists and investors. Econometrics and artificial intelligence often complement each other to enhance the precision of volatility predictions by leveraging the strengths of both approaches. In order to provide high accuracy of price movements and volatility in financial markets, advanced econometrics approaches [1, 2, 3] and artificial intelligence methodologies [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14] have been embedded in the financial data analytics. Studying the fluctuations in financial returns and economic variables is noted to make meaningful contributions to society.

Throughout recorded history, human civilizations have been enamored with gold. Empires and kingdoms rose and fell in pursuit of this precious metal and the wealth it symbolized. With the progression of societies, gold solidified its status as a dependable currency. Ultimately, gold's impact has consistently surpassed that of any other physical commodity, enduring through the ages without waning. However, the worldwide financial markets, commodities markets, and economic activity in all nations have been impacted by the Covid-19 outbreak. Given the erratic nature of gold prices, modelling and forecasting are crucial to investors since these assisted them in comprehending how the gold market functions and anticipating its future path. Additionally, technical analysis of the stock market provides crucial data for risk allocation.

The principal objective of this research is to forecast and model various conditional variance models for COMEX Gold Futures using Python programming. The GARCH model, artificial ANN model, and SVR model [15,16] are the models employed in this study. The COMEX Gold data is partitioned into training and validation sets, with the validation set serving the purpose of evaluating forecasting precision. Root Mean Square Error (RMSE) will be used as accuracy metrics in order to evaluate the forecast accuracy. Furthermore, the volatility analysis results are used to further analyse some financial applications including the measurements using Value-at-Risk (VaR) and Conditional VaR or sometimes notated as Expected Shortfall (ES).

2.0 Methodology

2.1 *Volatility models*

Volatility represents a statistical metric that quantifies the extent of returns' dispersion for a specific security or market index. Derivative contracts can be used to build strategies to profit from volatility. Volatile market certainly bring uncertainty to the achievement of investor's goal. Hence, volatility is a very useful indicator which acts as a warning of both short-term threats and long-term opportunities. Data analysis that corresponding to realized volatility and value-at-risk will be performed and discussed in detail in this section. A generalized model was later proposed by Bollerslev with GARCH(p, q) model which has the specification

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 . \tag{1}$$

This standard specification has a drawback of assumption that the current conditional variance and the conditional variance on prior period is linear. In order to overcome this linearity assumption issue, SVR and ANN can be incorporated in the GARCH model. [17] SVR begins by nonlinearly transforming the input space into a high-dimensional feature space. Then, linear regression is performed in the output space, effectively leading to nonlinear regression in the original, lower-dimensional input space. An essential aspect of SVR design is the identification of support vectors, which are a subset of the training data representing stable characteristics of the entire dataset. Given a training data set, $\{(x_t, y_t)\}_{t=1}^{N}$ with vector inputs $x_t \in \Re^m$, scalar output, $y_t \in \Re$, three main different main kernel functions. [17] are used in the optimization procedures.

Polynomial kernel: $K(x, y) = (x^T y + c)^d$

Radial Basis (Gaussian) kernel: $K(x, y) = e^{-\gamma \|x-y\|} 2, \gamma > 0.$

Exponential Kernel:

$$K(x, y) = e^{-\gamma |x-y|} \tag{2}$$

Next, ANN serve as the foundational elements for deep learning. In a neural network, data undergoes processing through multiple stages to reach a decision. Every neuron receives the result of a dot product as input and utilizes an activation function [17] to determine its output as follow:

$$y = \omega_1 x_1 + \omega_2 x_2 + b \tag{3}$$

where b is bias, w is corresponding weight and x is the input data. Throughout this process, input data undergoes various mathematical manipulations in both hidden and output layers. The model acquires knowledge of coefficients as it progresses from the input layer to the hidden layer. The activation functions will perform nonlinear transformations in the hidden layers. For performance evaluation of each model, the model evaluation is based on the performance of root mean squared error (RMSE). RMSE is an effective measure of forecast accuracy. Consider the observed or recorded data as x_t and the forecast value as $x\hat{}_t$, then the definition of RMSE is:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N} (x_t - \hat{x}_t)^2} \tag{4}$$

2.2 *Market Risk Evaluation*

The risk is measured and forecasted using two ways in this study. Value at Risk (VaR) and Conditional VaR are standard metrics for managing financial risk. Many forms of study employ the VaR and Conditional VaR to quantify the level of risks for financial derivatives.

In the case where the return series is anticipated to follow a normal distribution, the VaR will be formulated as follows:

$$VaR = \mu + \sigma \Phi^{-1}(c) \tag{5}$$

where μ, σ and $\Phi^{-1}(c)$ represent the conditional mean, standard deviation, and inverse one-tail standard normal distribution function at confidence level. For Conditional VaR, this method entails calculating the average of the portfolio's tail distribution returns at a selected confidence level. Assuming the return series is modeled as a normal distribution with a mean μ and standard deviation, then the Conditional VaR is

$$\text{Conditional } VaR = \mu + \sigma \frac{e^{-\frac{\Phi^{-1}(c)^2}{2}}}{\sqrt{2\pi}(1-c)} \tag{6}$$

3.0 Empirical study

This study utilizes daily price indices of COMEX Gold Futures. The data set is obtained from the website of CME Group. Gold COMEX data used in this study covers the period from 10th October, 2012 to 7th October, 2022 for a total of 2547 observations after adjustment. 10 years of data is significant to undergo volatility analysis and draw new conclusions.

3.1 *Volatility estimation using GARCH model*

GARCH model is fitted into the realized volatility of the COMEX Gold return index. The automated process in Python fitting different GARCH Model to the realized volatility and report the ranking based on AIC. The GARCH model with the least AIC value is selected as the best-fitting model. All in all, the Table 1 shows GARCH(1,2) model appears to be the most appropriate GARCH model to describe the volatility of the COMEX Gold returns.

Table 1

Fitted Parameter for GARCH(1,2) Model

GARCH(1,2) Model with Normal Distribution				
Parameter	**Coefficient**	**Std.error**	**T-statistics**	**P-value**
Constant	0.0175	1.468×10^{-2}	1.193	0.233
α_1	0.0511	2.566×10^{-2}	1.993	4.625×10^{-2}
β_1	0.1999	0.422	0.473	0.473
β_2	0.7313	0.418	1.748	8.047×10^{-2}

It can be seen from Table 1 that all parameter estimates, except for the constants and β_1 in the model are highly significant indicating that GARCH(1,2) model obtained a satisfactory fit. Following that, the

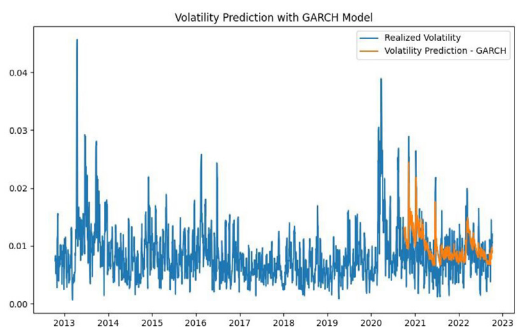

Figure 1

Volatility Prediction by GARCH Model

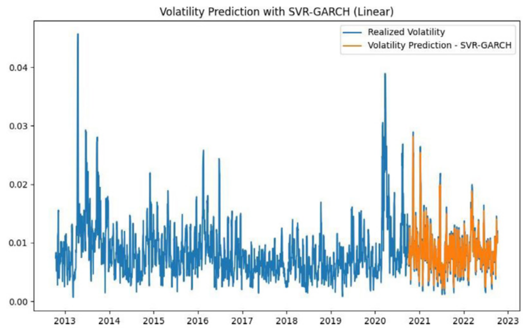

Figure 2

Volatility Prediction by SVR Model with Linear Kernel

estimated fitted GARCH(1,2) model is used to predict the COMEX Gold returns for the next few months and the predicted outcomes are compared to the validation data. Figure 1 shows the actual COMEX Gold returns associated with the predicted returns.

Clearly, the predicted returns value cannot capture all variability well, and the GARCH model failed to accurately predict most extraordinarily high and meager returns. The RMSE value for the GARCH Model is 0.0896.

3.2 *Volatility estimation using SVR-GARCH model*

SVR with three different kernel methods are applied in this paper to investigate the prediction accuracy of machine learning models. Three figures namely the Figure 2, Figure 3 and Figure 4 plot the observed and predicted realized volatility of COMEX Gold Futures. It can be seen clearly

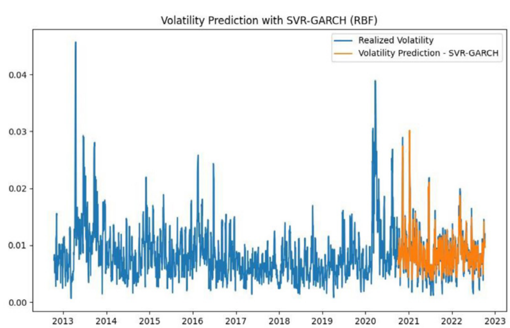

Figure 3

Volatility Prediction by SVR Model with RBF Kernel

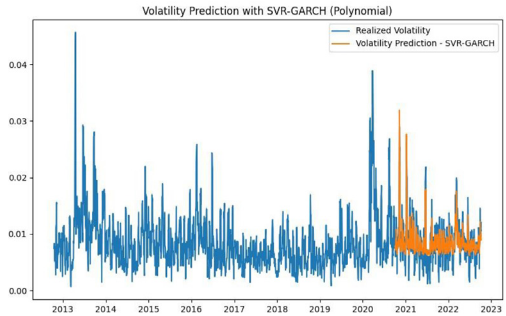

Figure 4

Volatility Prediction by SVR Model with Polynomial Kernel

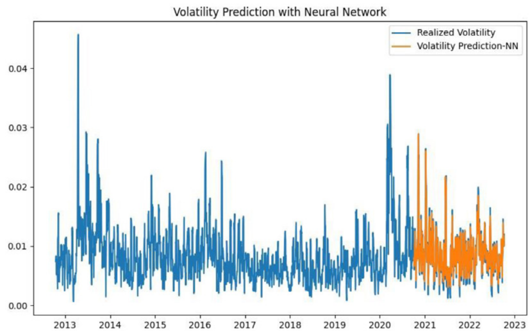

Figure 5

Volatility Prediction by Neural Network Model

that the SVR model with linear and RBF kernels fit well with the observed values. Most of the extreme high and low volatility values are successfully captured by the linear and RBF kernel. The predicted values follow accordingly and closely with the actual ones. On the contrary, the polynomial kernel performs slightly worse than the other two kernels from Figure 4. This conclusion is further supported by the RMSE values shown in Table 2.

Table 2

RMSE Value of Support Vector Regression Model

SVR Models	RMSE Value
SVR with Linear Kernel	0.000641
SVR with RBF Kernel	0.000605
SVR with Polynomial Kernel	0.001651

3.3 *Volatility estimation using Neural Network*

ANN model is used in this study to forecast the volatility of COMEX Gold Futures' returns. The application of ANN for volatility prediction is performed using the MLPRegressor module from scikit-learn in Python [17]. The model is constructed with three hidden layers, each with a different number of neurons. The forecasting plot along with the observed values is shown in Figure 5. The prediction power of the ANN is quite strong as it can be clearly observed that largely of the predicted values follow closely to the observed values, including some notably high ones from Figure 5. The RMSE of ANN model is 0.000731.

3.4 *Comparison and Evaluation of Predicted Results*

Based on the results in the preceding sections, all five models have some predictive power. However, it is common to choose a model that performs the best on a hold-out test data set. In light of that, RMSE has been introduced and chosen to compare the performance between prediction models. Both machine learning models are performing better than the GARCH Model in overall. This could be due to their abilities in capturing the variability in the sample.

From Table 3, SVR model with RBF kernel outperforms the other four models and being reported as the best model that fit COMEX Gold Futures return the most. SVR model with linear kernel ranks second. However, results from ANN model surpasses SVR with polynomial kernel by at least 50% under RMSE criterion. Overall, GARCH model behaves the worst among all five models.

3.5 *Market Risk Valuation*

In this research, VaR and Conditional VaR are computed under the assumption of a total invested capital of \$100,000 for the upcoming one-day ahead trading day, considering a 95% confidence level. Note that the forecast is started from 8th October 2022 which is the next trading day after the validation data set. The VaR and Conditional VaR for the COMEX Gold Futures returns by using SVR model with RBF kernel are considered under the period before and during the COVID-19 pandemic. The results are as follows:

Before COVID-19 period (10th October, 2012 to 28th February, 2022):
95%1-day VaR = \$100, 000 × 2.58174% = \$2581.74
95%1-day Conditional VaR = \$100, 000 × 3.17225% = \$3172.25

Table 3
RMSE Value of All Volatility Models

Volatility Models	RMSE Value
GARCH (1,2)	0.089600
SVR (Linear Kernel)	0.000641
SVR (RBF Kernel)	0.000605
SVR (Polynomial Kernel)	0.001651
ANN	0.000731

During COVID-19 period (10th October, 2012 to 7th October, 2022):
95%1-day VaR = $100, 000 × 1.50404% = $1504.04
95%1-day Conditional VaR = $100, 000 × 1.96998% = $1969.98

It has been observed that COMEX Gold Futures exhibit lower volatility during the COVID-19 period compared to the period preceding the pandemic. As one of the alternative safe investment, gold price and stock market has a negative correlation or inverse relationship. Several studies indicated that the stock market volatilities become more intense [18, 19, 20] during COVID-19. Therefore, when the stock market is volatile, the gold market become relative less volatile. It is worth noting that the relationships between gold and stock markets can also be influenced by other factors such as adjustment of interest rate, currency exchange value and even geopolitical events in a specific region.

4.0 Conclusion

This study analyzed the COMEX Gold Futures daily prices from 10th October 2012 to 7th October 2022. Five models were considered. They were, the GARCH model with GARCH(1,2), SVR with linear kernel, SVR with RBF kernel, SVR with polynomial kernel, and ANN. Judging by RMSE, SVR with RBF kernel was found to be best model with RMSE 0.000605. The results also suggested that machine learning models in general have better prediction power than the GARCH models. The predicted volatility is used in examining the market risk of gold futures. It is found that the gold futures volatility has a tendency to move in the opposite direction of stock market. This can be observed from the market risk is higher before the COVID-19 as compared during the pandemic. As a conclusion, the investment in gold can be considered as a hedge against market uncertainty due to pandemic where investors may flock to gold as a safe haven.

References

[1] T. Liu and Y. Shi, "Forecasting crude oil future volatilities with a threshold zero-drift GARCH model," *Mathematics*, vol. 10, p. 2757 (2022).

[2] R. Maheshwari and V. Kapoor, "Estimating the volatility of stock price index for the Indian market using GARCH model," *J. Stat. Manag. Syst.*, vol. 25, no. 7, pp. 1523-1530 (2022).

[3] P. J. Venter and E. Maré, "Univariate and multivariate GARCH models applied to Bitcoin futures option pricing," *J. Risk Financial Manag.*, vol. 14, p. 261 (2021).

[4] K. D. Ashok and S. Murugan, "Performance analysis of Indian stock market index using neural network time series model," in Proc. 2013 Int. Conf. Pattern Recognition, Informatics and Mobile Engineering, pp. 72–78 (2013).

[5] L. Chuong and D. Nikolai, "Forecasting of realised volatility with the random forests algorithm," *J. Risk Financial Manag.*, vol. 11, no. 4, p. 61 (2018).

[6] J. Huang and W. Pan, "Performing stock closing price prediction through the use of principle component regression in association with general regression neural network," *J. Discrete Math. Sci. Cryptogr.*, vol. 12, no. 6, pp. 717-728 (2009).

[7] M. Seo and G. Kim, "Hybrid forecasting models based on the neural networks for the volatility of Bitcoin," *Applied Sciences*, vol. 10, p. 4768 (2020).

[8] W. K. Liu and M. K. P. So, "A GARCH model with artificial neural networks," *Information*, vol. 11, p. 489 (2020).

[9] A. Petrozziello, L. Troiano, A. Serra, I. Jordanov, G. Storti, R. Tagliaferri, and M. L. Rocca, "Deep learning for volatility forecasting in asset management," *Soft Comput.*, vol. 26, pp. 8553–8574 (2022).

[10] L. Tze and P. S. Samuel, "Stochastic neural networks with applications to nonlinear time series," *J. Amer. Stat. Assoc.*, vol. 96, no. 455, pp. 968–981 (2001).

[11] H. S. Vatsal, "Machine learning techniques for stock prediction," *Foundations of Machine Learning*, vol. 1, no. 1, pp. 6–12 (2007).

[12] W. J. Chen, J. J. Yao, and Y. H. Shao, "Volatility forecasting using deep neural network with time-series feature embedding," *Economic Research Ekonomska Istraživanja*, vol. 36, no. 1, pp. 1377-1401 (2023).

[13] S. H. Yang, "Stock price direction prediction by directly using prices data: an empirical study on the Kospi and HSI," *Int. J. Bus. Intell. Data Mining*, vol. 9, no. 2, pp. 145–160 (2014).

[14] H. Q. Yang, L. W. Chan, and I. King, "Support vector machine regression for volatile stock market prediction," in Proc. Int. Conf. Intell. Data Eng. Autom. Learn., pp. 391–396 (2002).

[15] V. Jyothi and M. M. Tripathi, "K-means clustering based photovoltaic power forecasting using artificial neural network, particle swarm optimization and support vector regression," *J. Inf. Optim. Sci.*, vol. 40, no. 2, pp. 309-328 (2019). doi: 10.1080/02522667.2019.1578091.

[16] B. Parveen and S. Bhardwaj, "Estimation of solar radiation using support vector regression," *J. Inf. Optim. Sci.*, vol. 40, no. 2, pp. 339-350 (2019). doi: 10.1080/02522667.2019.1578093.

[17] K. Abdullah, Machine Learning for Financial Risk Management with Python, O'Reilly Media, Inc. (2021).

[18] T. Cheng, J. Liu, W. Yao, and A. B. Zhao, "The impact of COVID-19 pandemic on the volatility connectedness network of global stock market," *Pacific-Basin Finance Journal.*, vol. 71, p. 101678 (2022).

[19] H. Hong, Z. Bian, and C. C. Lee, "COVID-19 and instability of stock market performance: evidence from the U.S.," *Financial Innovation*, vol. 7, p. 12 (2021).

[20] S. S. Sharma, "A note on the Asian market volatility during the COVID-19 pandemic," *Asian Economics Letters*, vol. 1, no. 2 (2020).

Received February, 2023