

Empirical Analysis and Forecasting of Gold Prices: Based on VAR Model

Yixuan Fang *

Department of Finance and Economics, Shandong University of Science and Technology,
Shandong, China

* Corresponding Author Email: sharonfang278@gmail.com

Abstract. To achieve accurate forecasting of future gold prices and thereby provide effective references for investors' decisions, this study selects time-series data of gold prices (gold), the U.S. Dollar Index (DXY), and WTI crude oil futures prices (WTI) from 2015 to 2020. Employing a comprehensive empirical framework that integrates stationarity tests, lag order selection, VAR model estimation, Wald tests, unit circle tests, Granger causality tests, impulse response analysis, and out-of-sample forecasting, this research systematically examines the dynamic interactive mechanisms among the three variables. The findings reveal a distinct asymmetric linkage effect: DXY exerts a more pronounced impact on gold prices than WTI, with a significant short-term negative correlation between first-order lagged DXY and gold. Granger causality tests confirm that DXY is a Granger cause of gold prices, while gold serves as a Granger cause for both DXY and WTI. Impulse response analysis further demonstrates that gold prices exhibit high short-term sensitivity to DXY shocks, with a negative inhibitory effect persisting for the first three periods before gradually weakening. Moreover, a 6-period ahead forecast of gold price first-order differences shows that the VAR model's predicted values align closely with actual values in terms of trend and magnitude, verifying its reliability in forecasting. This study offers actionable insights for investors to mitigate risks and for policymakers to monitor the stability of the gold market.

Keywords: Enter key words or phrases in alphabetical order, separated by commas.

1. Introduction

As a special asset with commodity, monetary, and hedging attributes, gold price fluctuations are affected by multiple intertwined factors such as the U.S. dollar exchange rate, energy prices, and macroeconomic policies. Accurately capturing the dynamic driving mechanism and future trends of gold prices is of great significance for investors' asset allocation, risk management, and policymakers' macro-control. The Vector Autoregression (VAR) model, which can effectively describe the dynamic linkage relationships among multiple variables, has become one of the core tools for analyzing price fluctuations in financial markets. Many scholars have used this model to explore the impact mechanisms of capital factors and macroeconomic factors on gold prices and have achieved rich research results.

A review of relevant literature shows that some studies focus on the relationships between single or multiple macro-variables and gold prices by constructing VAR models. Ambarita and Djamaluddin employed Multiple Linear Regression (MLR) analysis to examine the impacts of the Federal Funds Rate, US Inflation Rate, DJIA Index, and WTI Price on the gold price over the period 1971-2019. Their findings revealed that the WTI Price exerts a significantly positive influence on the gold price [1]. Mamcarz utilized VAR/VECM models to analyze the interrelationships between stock, gold, crude oil, and bond returns. The results indicated that gold prices have a negative impact on the stock market, while crude oil and bonds have a positive impact on the stock market [2]. Robinson applied a structural VAR model to analyze the relationship between the gold price and variables such as the GDP of OECD countries, oil prices, and the US dollar's real effective exchange rate. It was found that a short-term shock to oil prices drives up the gold price, while shocks to the US dollar exchange rate have a relatively moderate impact on the gold price [3]. Lili and Chengmei classified the factors influencing the gold price into three categories—gold reserves and energy prices, financial market indices, and global macroeconomic indicators—by employing a Factor-Augmented Vector

Autoregression (FAVAR) model. They confirmed a positive correlation between energy prices and the gold price [4]. However, this study failed to conduct an in-depth exploration of the differences in the intensity of the impact between variables under different lag orders, nor did it test the model's ability to forecast the gold price. As a result, it cannot provide investors with actionable basis for price prediction. Pachiyappan and Chandrakala focused on the Indian market and used the Ordinary Least Squares (OLS) method to find that variables such as the exchange rate and Wholesale Price Index (WPI) have a positive correlation with the gold price [5]. Although the study involves multivariate analysis, it is confined to a single national market. Consequently, their reference value for understanding the linkage logic of global gold prices is limited. Sujit and Kumar analyzed the dynamic relationships between the gold price, oil price, exchange rate, and stock market returns using VAR and cointegration techniques. They found that the exchange rate is greatly affected by other variables [6]. Nevertheless, the study period ended in 2011, which undermines the timeliness of its conclusions. Tran and Nguyen used a Panel Vector Autoregression (PVAR) model to analyze the correlations between the stock market, gold price, and US dollar exchange rate in Asia and Europe during the COVID-19 pandemic. The results showed that there is a bidirectional causality between the stock market and gold price in Asian markets, and in European markets, a causal relationship exists among the three variables at the 1% significance level [7]. However, this study focuses on the special period of the COVID-19 pandemic, an exogenous shock. This may disrupt the original linkage logic between variables, leading the model to overfit short-term abnormal fluctuations and fail to capture core correlations. Thus, the reference value of its conclusions for normal market conditions is limited.

Existing studies on gold prices based on the VAR model have laid a theoretical foundation, but there are still three shortcomings: First, most studies only focus on a single variable (such as DXY) and ignore the interaction effects between variables, making it difficult to systematically reveal the linkage essence of gold, DXY, and WTI. Second, the research periods are mostly concentrated around 2008 or 2021-2024. The samples in these two periods will introduce multiple sudden and disruptive exogenous shocks (such as the global financial crisis, the COVID-19 pandemic, and the Russia-Ukraine conflict). These shocks will break the original linkage logic between variables, leading the model to overfit short-term abnormal fluctuations and fail to capture core correlations, so the conclusions have limited reference value for the current market. Third, some studies only stay at the level of verifying causal relationships, and the test of the forecasting ability of the VAR model is insufficient, failing to form a complete research chain of "mechanism analysis-model verification-forecasting application".

Based on this, the research value and innovations of this study are mainly reflected in three aspects: First, it focuses on the key period from 2015 to 2020 to ensure that the core driving factors of gold are single, avoiding the complex and changeable driving logic caused by the superposition of multiple factors (pandemic hedging, excessive currency issuance, geopolitical conflicts, and interest rate hikes to suppress), making the research conclusions more in line with the fluctuation characteristics of the gold market in recent years. Second, it incorporates gold, DXY, and WTI into the same VAR framework, and quantifies the asymmetric linkage effects among the three through model estimation, impulse response analysis, and other methods, filling the gap in the details of the interactive effects of multiple variables in existing studies. Third, it not only verifies the causal relationships between variables but also tests the forecasting reliability of the VAR model by forecasting 6 periods ahead and comparing the forecast values with the actual values. It provides investors with operable price prediction tools and at the same time provides empirical support for policymakers to monitor gold market fluctuations and prevent financial risks. This study will systematically explore the inter-influential relationships among the three through stationarity test, lag order selection, model estimation, Wald test, unit circle test, Granger causality test, and impulse response analysis, in order to provide a scientific reference for gold investment decisions [8-12].

2. Research Foundation

To explore the effects of DXY and WTI on gold and analyze the dynamic linkage among the three, this study incorporates the three variables into the same model for analysis, while considering the endogenous relationships caused by variable lags.

This study selects time-series data from 2015 to 2020 and constructs a VAR model to verify the factors affecting gold price changes. The constructed VAR model is as follows:

$$Y_t = A + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_k Y_{t-k} + U_t, U_t \sim \text{IID}(0, \Omega) \quad (1)$$

$$Y_t = (\text{gold}_t, \text{DXY}_t, \text{WTI}_t)' ; A = (\alpha_1, \alpha_2, \alpha_3)' ; U_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' ; \quad (2)$$

$$B_j = \begin{bmatrix} b_{11,j} & \dots & b_{13,j} \\ \vdots & & \vdots \\ b_{31,j} & \dots & b_{33,j} \end{bmatrix}, j = 1, 2, \dots, k. \quad (3)$$

In Equation (1) (2) (3), A is the equation constant matrix, U_t is the white noise vector, t is time, B_j is the matrix of to-be-estimated coefficients of variables, and k is the lag order of the model.

3. Variable Selection and Data Analysis

A total of 72 monthly data points from January 2015 to December 2020 are selected as the research objects, and descriptive statistical analysis is conducted on gold, DXY, and WTI. The results are shown in Table 1.

Table 1. Descriptive Statistical Information of gold, DXY, and WTI.

Variable	Minimum Value	Maximum Value	Mean	Standard Deviation	Skewness	Kurtosis
gold	1068.3000	1968.6000	1349.4208	216.4096	1.4185	4.1573
DXY	89.1300	102.2100	96.0775	2.8171	-0.2860	2.7875
WTI	18.8400	74.1500	50.9800	10.8085	-0.3200	3.6557

Data source: Investing.com

It can be seen from Table 1 that the means and standard deviations of the three variables are all positive, but the degrees of dispersion are quite different. Among them, the standard deviations of DXY and WTI are small, while that of gold is very large, indicating that gold has a large degree of dispersion and volatility. In terms of skewness, the skewness values of DXY and WTI are negative, indicating that their time series are left-skewed, while the skewness value of gold is positive, indicating that its time series is right-skewed. By taking the absolute values of the skewness of the three variables, it is found that the skewness value of gold is the largest, and that of DXY is the smallest, indicating that the left-skewness of WTI is more serious than that of DXY, and the right-skewness of gold is more obvious than the left-skewness of DXY and WTI. In terms of kurtosis, the kurtosis values of gold and WTI are 4.1573 and 3.6557, respectively, which are greater than 3, indicating that both are leptokurtic and fat-tailed distributions.

4. Empirical Analysis

4.1. Data Stationarity Test

To deeply explore the inter-influential relationships among the three variables, this study conducts an empirical analysis by constructing a VAR model. Since the selected data are all time-series data, there is a possibility of spurious regression. The establishment of a VAR model is based on stable time-series data. Therefore, before model estimation, the Augmented Dickey-Fuller (ADF) test is conducted to judge the stationarity of each variable. The relevant ADF test results are shown in Table 2.

Table 2. Variable Unit Root Test Results.

Variable	T-statistic	1% Level Critical Value	5% Level Critical Value	10% Level Critical Value	P
gold	-1.707	-4.040	-3.450	-3.150	0.0169
DXY	-2.252	-4.040	-3.450	-3.150	0.1212
WTI	-2.802	-4.040	-3.450	-3.150	0.0373
gold1	-5.696	-4.040	-3.450	-3.150	0.0000
DXY1	-5.535	-4.040	-3.450	-3.150	0.0000
WTI1	-6.920	-4.040	-3.450	-3.150	0.0000

It can be seen from Table 2 that among the three variables of gold, DXY, and WTI, the P-values of the original sequences of gold and WTI are less than 0.05, while the P-value of the original sequence of DXY is greater than 0.05, so DXY is not a stationary sequence. Therefore, the natural logarithm of the original data of the three variables is first taken, and then the first-order difference processing is carried out to obtain gold1, DXY1, and WTI1. The unit root test is conducted again on them. The results show that after processing, all three reject the null hypothesis of the existence of a unit root. Therefore, the variables pass the stationarity test and can be used for VAR model construction and analysis.

Before estimating the VAR model for the linkage relationships among the three variables of gold, DXY, and WTI, it is also necessary to determine the lag order of the model. Therefore, the values of evaluation indicators such as the Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQIC), Schwarz Bayesian Information Criterion (SBIC), and Final Prediction Error (FPE) are calculated, and the results are shown in Table 3.

Table 3. Lag Order Test Results

Lag	AIC	HQ	SC	FPE
1	-18.91072	-18.75233*	-18.50929*	6.1289E-09
2	-18.87504	-18.59786	-18.17254	6.36455E-09
3	-18.97478*	-18.57881	-17.97121	5.78852E-09*
4	-18.88938	-18.37462	-17.58475	6.3616E-09
5	-18.74166	-18.10811	-17.13596	7.48253E-09
6	-18.60943	-17.85709	-16.70266	8.72781E-09

It can be seen from Table 3 that the AIC and FPE evaluation indicators support building a model with a lag of 3 periods, and the HQ and SC evaluation indicators support building a model with a lag of 1 period. The AIC criterion tends to select a more complex model, which is particularly suitable when the sample size is large, because it can better capture the dynamic relationships in the data; the SC and HQ criteria tend to select a simpler model, which may oversimplify the complex relationships between economic variables.

For financial time series, the transmission of market information takes time, and the response of financial markets to information is usually not immediate but requires a certain amount of time for digestion. Therefore, based on the principle of minimizing AIC and considering economic significance, it is more reasonable to select a lag of 3 periods to build the VAR model.

4.2. Model Estimation

After conducting empirical modeling and analysis on the three variables, the estimation results of the VAR model (i.e., the estimated values of the coefficients of each variable in the model) are obtained, and the relevant results are shown in Table 4.

Table 4. VAR Model Estimation Results.
(Variable Coefficient Estimates and P-values)

Variable	Gold Equation		
	<i>1st-order Lag</i>	<i>2nd-order Lag</i>	<i>3rd-order Lag</i>
gold	0.1639	-0.0186	-0.2829*
DXY	-0.6609**	0.0037	0.0130
WTI	-0.0781**	0.0463	-0.0439
Constant		-0.0061	
R ²		0.3078	

Variable	DXY Equation		
	<i>1st-order Lag</i>	<i>2nd-order Lag</i>	<i>3rd-order Lag</i>
gold	0.1011	-0.1471	0.0445
DXY	0.1735	-0.0233	-0.2748
WTI	0.0151	-0.0116	0.0037
Constant		0.0034	
R ²		0.1712	

Variable	WTI Equation		
	<i>1st-order Lag</i>	<i>2nd-order Lag</i>	<i>3rd-order Lag</i>
gold	0.4860	-1.2459	2.0665**
DXY	0.6803	-2.3608*	2.5569*
WTI	0.2690*	-0.3469**	-0.0515
Constant		0.0036	
R ²		0.3123	

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% significance levels respectively.

It can be seen from Table 4 that all lag terms of the gold equation are jointly significant ($P=0.0133<0.05$), indicating that these lag variables have explanatory power for the gold price as a whole.

At the same time, the 1st-order lag effect of DXY on gold is the most significant (with a coefficient of -0.6609, showing a strong negative correlation). The 3rd-order lag effect of gold on WTI is the most prominent (with a coefficient of 2.0665, indicating an extremely strong positive correlation). DXY has a significant negative impact on WTI at the 2nd-order lag (with a coefficient of -2.3608) and an extremely significant positive impact at the 3rd-order lag (with a coefficient of -2.5569).

The impact of DXY on gold is greater than that of WTI on gold, and the impact of DXY on WTI is greater than its impact on gold. DXY has a significant short-term impact on gold and a medium-term impact on WTI, while the impacts of WTI on gold and DXY are very weak and almost negligible. Therefore, there is an asymmetric linkage relationship among the three variables.

4.3. Wald Test

To determine the joint significance of the equations in the VAR model and the coefficients of each order in the equations, a Wald test is conducted on the model. The relevant results are shown in Table 5. The test results of the three statistics all satisfy $P<0.05$, which indicates that the model meets the conditions for conducting subsequent tests.

Table 5. Wald Test Results.

Variable	df	chi2	P
gold	2	6.1723	0.046
DXY	2	11.0530	0.004
WTI	2	7.3472	0.025

4.4. Unit Circle Test

A unit circle test is performed on the estimation results of the VAR model, and the test results are shown in Figure 1. It can be seen from the test results that all the eigenvalues of the model are located inside the unit circle, which indicates that the model is stable as a whole, and the estimation results of the constructed VAR model are stable and reliable. Therefore, the estimation results of this VAR model can be used to explain the linkage relationships among gold, DXY, and WTI.

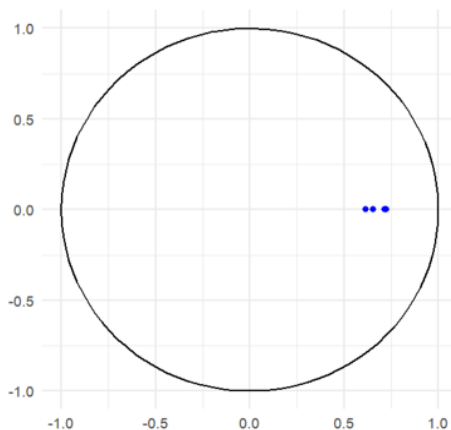


Fig 1. Unit Circle Test (Picture credit : Original).

4.5. Granger Causality Test

To analyze the statistical causal relationships between variables, a Granger causality test is conducted on the model, and the test results are listed in Table 6. At the 5% significance level, DXY is the Granger cause of gold ($P < 0.05$); WTI is not the Granger cause of gold ($P > 0.05$). In the reverse test, gold is the Granger cause of both DXY and WTI ($P < 0.05$). Therefore, the past values of DXY can be used to predict the future values of gold.

In addition, the table shows that there is a bidirectional causal relationship between gold and DXY, indicating that their fluctuations affect each other, and there may be a complex market linkage; both DXY and gold can predict WTI, but WTI cannot predict the former two. This further verifies the existence of an asymmetric linkage relationship among the three variables.

Table 6. Causality Test Results.

Equation	Exclude	F-Test	P
gold	DXY	3.1933	0.005
gold	WTI	1.8019	0.101
gold	All	2.4549	0.027
DXY	gold	3.3860	0.004
DXY	WTI	1.8019	0.101
DXY	All	0.9015	0.495
WTI	gold	3.3860	0.004
WTI	DXY	3.1933	0.005
WTI	All	2.7613	0.014

4.6. Impulse Response Analysis

From the above empirical analysis, it can be found that there is a relatively stable asymmetric linkage relationship among gold, DXY, and WTI, and DXY is the Granger cause of gold. Based on this, on the basis of the VAR model estimation, an impulse response analysis of DXY on gold is conducted, and the analysis results are shown in Figure 2.

The initial value of the impulse response of gold to DXY shocks is 0. From the initial period to the 7th period, the impulse response is relatively significant and reaches its peak in the 2nd period. Since the impulse response value is negative from the initial period to the 3rd period, it means that in

the short term, gold is highly sensitive to DXY fluctuations, and an increase in DXY will have a negative inhibitory effect on gold. From the 4th period onwards, the negative shock effect rebounds and turns positive. From the 7th period onwards, the shock effect gradually weakens and finally approaches 0. Therefore, in the long run, the impact of DXY on gold is relatively large in the early stage; in the short term, an increase in DXY will affect investors' investment intentions towards gold. When the U.S. dollar strengthens, gold priced in U.S. dollars becomes "more expensive", thereby suppressing the demand for gold purchases and leading to a decline in gold prices.

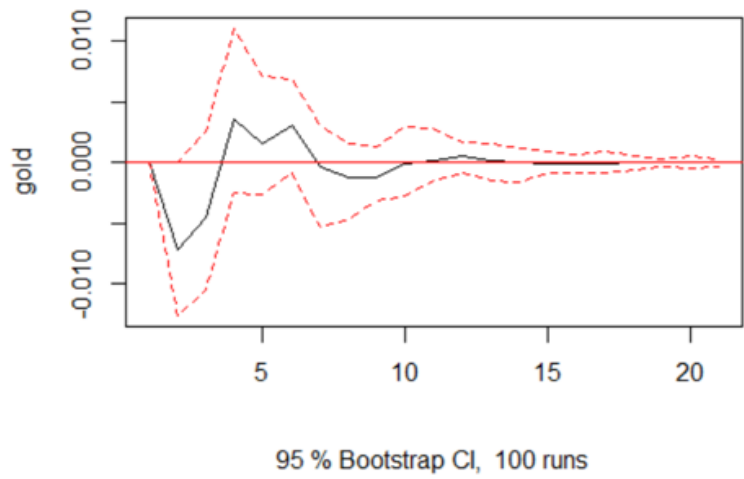


Fig 2. Impulse Response Analysis of the U.S. Dollar Index on Gold Price. (Picture credit : Original)

4.7. Forecasting of Future Values

The results of the Granger causality test show that the future values of gold can be forecasted. Based on this, a VAR model of DXY and WTI is constructed to forecast the first-order difference sequence of gold for 6 periods ahead (i.e., 6 months). The forecasting results are shown in Figure 3. The changing trends of the forecast values and the actual values are relatively consistent, with an overall downward trend and the values being relatively close. However, as time goes by, the forecasting accuracy will gradually decrease. Through comparative analysis, it can be verified that the fitted VAR model has reliability in forecasting gold prices.

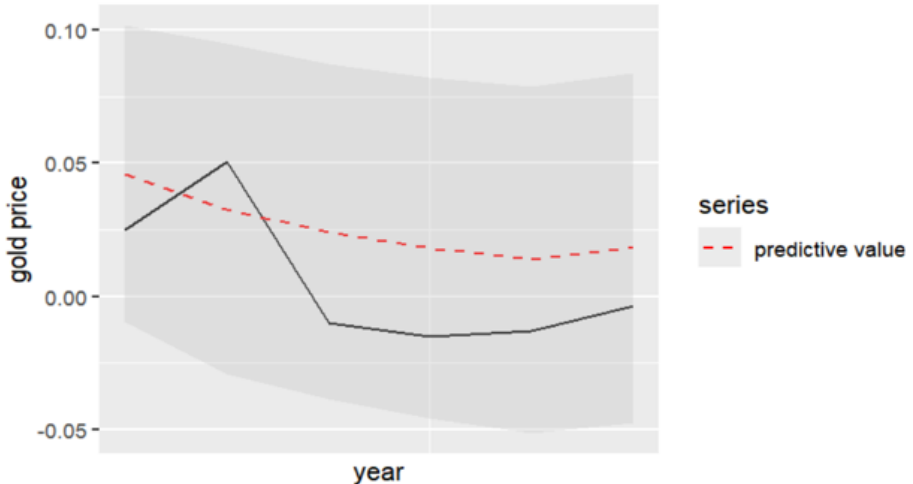


Fig 3. Comparison of Trends between Forecasted and Actual Gold Prices (Picture credit : Original).

5. Conclusion

Based on the time-series data from 2015 to 2020, an empirical analysis of gold prices is conducted using the VAR model, and the following conclusions are drawn:

There is an asymmetric linkage relationship among gold, DXY, and WTI. The impact of DXY on gold is greater than that of WTI on gold, and an increase in DXY has a negative inhibitory effect on gold in the short term.

The Granger causality test shows that DXY is the Granger cause of gold.

The forecast of the first-order difference sequence of gold for 6 periods ahead based on the VAR model shows that the changing trends of the forecast values and the actual values are consistent, and the values are close, which verifies the reliability of the model in forecasting future gold prices.

Since DXY is affected by both domestic and foreign factors such as adjustments in the Federal Reserve's monetary policy (e.g., the interest rate hike cycle from 2015 to 2018 and the loose policy during the 2020 pandemic period) and the global economic prosperity (e.g., the economic downturn in the early stage of the pandemic), it can be inferred from the asymmetric linkage relationship among the three variables that when the above-mentioned influencing factors change, the volatility of DXY will increase significantly, which in turn will lead to a simultaneous increase in its linkage with gold and WTI. Fluctuations in DXY will directly affect the trend of gold prices through channels such as the pricing effect and the transmission of hedging sentiment. It is suggested that when conducting gold investments, investors should comprehensively pay attention to the policy trends of the Federal Reserve and the international energy supply-demand pattern, rather than only tracking the gold price itself. By grasping the linkage rules among the three, they can timely predict the fluctuation trend of gold prices, provide comprehensive references for asset allocation decisions, and effectively avoid market risks.

References

- [1] Ambarita, S, Djameluddin, S. Analysis of Macroeconomic Effects on Gold Price. *IOSR Journal of Business and Management (IOSR-JBM)*, 2020, 22(9): 47-57.
- [2] Mamcarz, K. The Causal Relationship Between Stocks, Gold, Crude Oil, and Bond Returns in Poland. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H Oeconomia*, 2024, 58(4), 127-147.
- [3] Robinson, Z. Revisiting gold price behaviour: a structural VAR. *Mineral Economics*, 2019, 32(3), 365-372.
- [4] Lili, L., & Chengmei, D. Research on the influence of macro-economic factors on the price of gold. *Procedia Computer Science*, 2013, 17, 737-743.
- [5] Pachiyappan, S., & Chandrakala, G. Do the macroeconomic factors influence the volatility of gold price? An empirical study. *Journal of Commerce & Accounting Research*, 2022, 11(2), 37-44.
- [6] Sujit, K. S., & Kumar, B. R. Study on dynamic relationship between gold price, oil price, exchange rate and stock market returns. *International journal of applied business and economic research*, 2011, 9(2), 145-165.
- [7] Tran, O., & Nguyen, H. The interdependence of gold, US dollar and stock market in the context of COVID-19 pandemic: an insight into analysis in Asia and Europe. *Cogent Economics & Finance*, 2022, 10(1), 2127483.
- [8] Adewole, A. I., Bodunwa, O. K., & Akinyanju, M. M. Structural vector autoregressive modeling of some factors that affect the economic growth in Nigeria. *Science World Journal*, 2020, 15(2), 11-15.
- [9] Berlinger, E., Illés, F., Badics, M., Banai, Á., Daróczi, G., Dömötör, B., ... & Vidovics-Dancs, Á. *Mastering R for quantitative finance*. Packt Publishing Ltd, 2015.
- [10] Konya, L. Unit-root, cointegration and Granger causality test results for export and growth in OECD countries. *International Journal of applied econometrics and quantitative studies*, 2004, 1(2), 67-94.
- [11] Ivanov, V., & Kilian, L. A practitioner's guide to lag order selection for VAR impulse response analysis. *Studies in Nonlinear Dynamics & Econometrics*, 2005, 9(1).
- [12] Hassani, H., Silva, E. S., Gupta, R., & Segnon, M. K. Forecasting the price of gold. *Applied Economics*, 2015, 47(39), 4141-4152.