

The Relationship between Oil Prices, Gold Prices, the Stock Market, and U.S. GDP

Xingyu Su*

University of Rochester, Rochester, the United States

*Corresponding author: xsu10@u.rochester.edu

Abstract. The forecast of GDP has always been a popular research topic today, and its related influencing factors are also complicated. This paper selects three key factors in economic development: gold price, crude oil price, and stock market index, explores their relationship with U.S. GDP, and then uses these three factors as predictive variables to obtain a multiple linear regression equation related to GDP. The research methods of this paper are as follows: Firstly, monthly U.S. GDP, monthly crude oil price, monthly gold price, and monthly S&P 500 index were collected. Secondly, correlation analysis was carried out on these data, including calculation of correlation coefficient and cross-correlation analysis. Correlation analysis showed that GDP had a significant positive correlation with other variables. Then, a multiple linear regression model was established with monthly U.S. GDP as the predicted variable, monthly crude oil price, monthly gold price, and monthly S&P 500 index as the predictor variable. Finally, multiple regression equations are obtained through testing. This multiple regression equation can be used to predict GDP further.

Keywords: Gross domestic product (GDP); Crude oil; Gold; Stock market.

1. Introduction

GDP is the core index of national economic accounting. GDP can reflect the overall economic strength and the speed of the economic development of a country or region, can be used for economic structure analysis, and can be combined with other indicators to calculate indicators of great significance. Therefore, GDP is often regarded as an important macroeconomic indicator to measure the overall economic situation of a country or region, and it plays a very important role in economic situation analysis and policy making. However, due to the slow updating frequency of GDP and the lag of GDP, this study hopes to find data with higher updating frequency to predict the trend of GDP, and to make decisions in advance.

With the development of financial derivatives, gold, crude oil, and stock markets became key factors in economic development, and their trading and price fluctuations also affected global financial markets. Gold has always been an important presence in global financial markets as a stable currency, seen as a hedge against economic and geopolitical risks, and a haven in times of economic, financial, and geopolitical volatility [1]. Crude oil is one of the most important energy sources in today's global economy, and its price changes have a significant impact on the global economy. Therefore, the relationship between oil price and GDP has always been a popular research topic, and oil price is also regarded as one of the predictor variables of GDP [2]. Previous studies have also shown that there is a positive correlation between the world economic growth rate and oil price fluctuations [3]. The stock market has always been regarded as closely related to economic development, and the changes in the stock market reflect the development trend of the macroeconomy to a certain extent. Stock market and macroeconomic influences are mutually reinforcing. On the one hand, the macroeconomic situation will affect the development of stocks. For example, when the economy is developing well, investors have stronger confidence and are more willing to buy stocks, thus pushing the stock market to rise. On the contrary, when the economy is not developing well, investors' confidence is weak, and they will be more willing to sell stocks to avoid risks, thus driving the stock market down. On the other hand, the situation of the stock market will also affect the situation of the macroeconomy, and the prosperity of the stock market can promote the development of the economy.

Therefore, this study selects crude oil price, gold price, and the S&P500 index as variables to explore their relationship with GDP. This study will focus on monthly data, select the GDP of the United States, and adopt correlation analysis and multiple linear regression methods. The objective of this study is to preliminarily analyze the correlation between U.S. GDP and crude oil price, gold price, and s&p500 index, and then establish a multiple regression model with U.S. GDP as the predicted variable and crude oil price, gold price, and s&p500 index as the predictor variable.

2. Method

2.1. Data

The data collected includes monthly U.S. GDP from January 2000 to May 2023, monthly crude oil prices, monthly gold prices, and monthly S&P500 index. the U.S. GDP data used for the analysis came from the Organization for Economic Co-operation and Development (OECD). Crude oil prices and gold prices come from The World Bank (<https://www.worldbank.org>). The price of crude oil is the average price of crude oil in US dollars per barrel and the price of gold is in US dollars per Troy ounce. S&P500 index data from Yahoo Finance, using data at the time of closure. Through the time plots of each data, it can be found that these data show an upward trend (Figs. 1, 2).

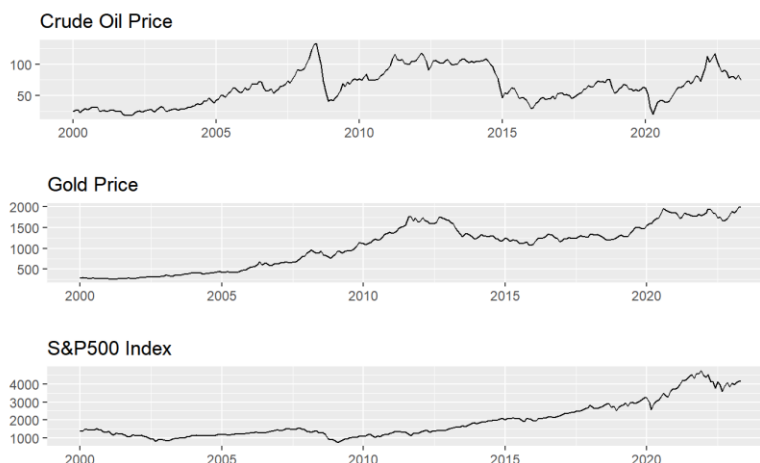


Fig. 1 Time plots of Crude Oil Price; Time plots of Gold Price; and Time plots of the S&P 500 Index (Photo/Picture credit: Original).

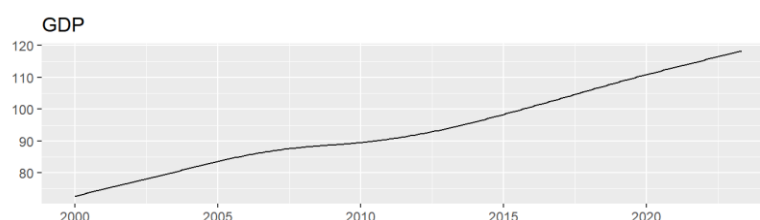


Fig. 2 Time plots of GDP (Photo/Picture credit: Original).

2.2. Correlation analysis

Correlation analysis refers to the analysis of two correlated variables to measure the closeness of correlation between two variable factors. The variables in this study include the monthly GDP of the United States, the monthly price of crude oil, the monthly price of gold, and the monthly frequency of the S&P500 index. The correlation coefficient is a statistical measure that reflects the linear correlation between two variables, and the correlation coefficient used in this study is the Pearson correlation coefficient. The formula is $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$. The pearson correlation coefficient is between -1 and 1, and the closer its absolute value is to 1, the higher the degree of

correlation between the two variables, that is, the more similar the two variables are. To make sure that the statistical conclusion obtained from the sample is valid, it is necessary to conduct a significance test on the conclusion. If the P-value is ≥ 0.05 , the result is not statistically significant. If the p-value is < 0.05 , the result is a significant correlation. If $p < 0.01$, the result is a very significant correlation.

The cross-correlation analysis of the variables was also carried out. Cross-correlation is a measure of how similar a one-time series is to a delayed version of another time series. Cross-correlation analysis was conducted between GDP data and crude oil price, gold price, and S&P500 index in turn. It is used to determine the degree to which two variables match each other and to find the point in time when the best match occurs.

2.3. Multiple linear regression model

A multiple linear regression model is a linear regression possessing one predicted variable (y) with k predictor variables (x). The general form of a multiple linear regression model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

y is the predicted variable (dependent variable); $x_1, x_2, \dots, x_k$ are the k predictor variables (independent variables); the parameter β_0 is the intercept of the regression line and the value of the predicted variable when all the predictor variables are equal to 0; the parameters $\beta_1, \beta_2, \dots, \beta_k$ re-estimated regression coefficients that measure, respectively, the extent to which the predictor variable will affect the predicted variable while holding all other predictors constant [4].

The predicted variable selected in this study is the monthly GDP of the United States, and the three predictor variables are the monthly price of crude oil, the monthly price of gold, and the monthly S&P 500 index. The purpose of using a multiple linear regression model is to explore the direct relationship between GDP and the three independent variables and to establish an expression that can be used to predict GDP based on the three dependent variables.

3. Result and Discussion

The correlation heat map shows the correlation coefficient and significance between the monthly GDP, the monthly price of crude oil (OIL), the monthly price of gold (GOLD), and the monthly S&P 500 index (SP500) (Fig. 3).

The correlation heat map shows that all variables are positively correlated, and all are significantly correlated. There is a significant moderate correlation between GDP and crude oil price, and a significant strong correlation between GDP and gold price and S&P 500 index, respectively. This may be because commodity booms and busts correspond to the economic cycle, and metal spot and futures prices are also closely linked to the economic cycle [5].

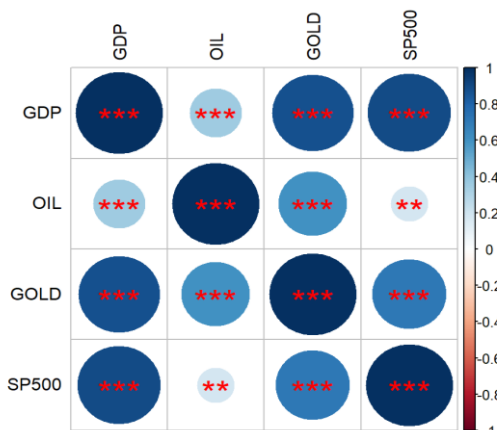


Fig. 3 The correlation heat map (Photo/Picture credit: Original).

The cross-correlation chart shows that the correlation coefficient of GDP, crude oil price, gold price, and s&p500 index reaches the maximum when the lag value is 0 respectively, indicating that there is no obvious lag between GDP and the other three variables (Fig. 4).

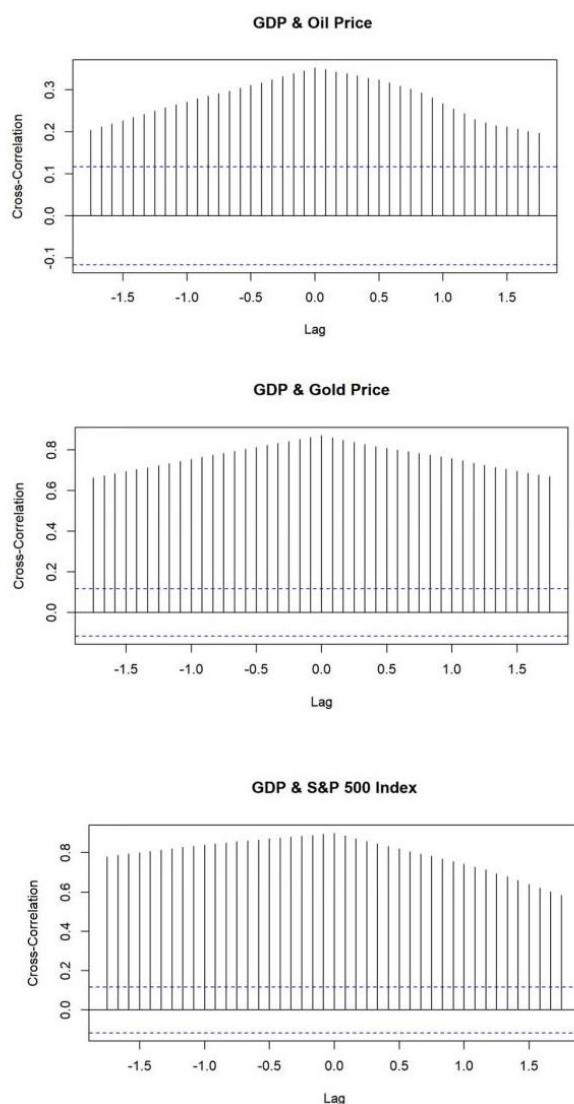


Fig. 4 Cross-correlation chart (Photo/Picture credit: Original).

With the monthly GDP of the United States as the predicted variable, a multiple linear regression model was established with the monthly price of crude oil, the monthly price of gold, and the monthly frequency of the s&p500 index as three predictor variables.

Table 1. Summary output of multiple linear regression

	Estimate	Standard Error	t value	Pr(> t)	
(Intercept)	69.956	0.642	108.951	0.0000	***
OIL	-0.031	0.011	-2.884	0.0042	**
GOLD	0.013	0.001	15.822	0.0000	***
SP500	0.007	0.000	19.462	0.0000	***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Residual standard error: 3.587 on 277 degrees of freedom

Multiple R-squared: 0.9197, Adjusted R-squared: 0.9189

F-statistic: 1058 on 277 and 3 DF, p-value: 0.0000

The R-squared of this regression model is 0.9197, indicating that the independent variables jointly explain 91.97% of the variance of the dependent variables, indicating that the model fits well (Table

1). The p-value of this regression model is about 0, indicating that the independent variables have a significant influence on the dependent variables.

Using the variance inflation factor to test the model for the presence of multicollinearity. The results show that all variance inflation factors are smaller than 10, indicating that the model is not multicollinearity.

Table 2. Goldfeld–Quandt test output

df1	df2	statistic	p.value	method	alternative
137.0	136.0	0.7	0.9935	Goldfeld-Quandt test	variance increases from segments 1 to 2

Using the Goldfeld–Quandt test to check for heteroscedasticity in the regression model. The result shows that the p-value is 0.7, so accept the null hypothesis that there is no heteroscedasticity (Table 2).

Finally, the multiple linear regression equation obtained is:

$$y = 69.956 + (-0.031)x_{oil} + 0.013x_{gold} + 0.007x_{s\&p500} \quad (2)$$

The regression coefficient and the p-value of each predicted variable show that gold price and the S&P 500 index have a significant positive correlation with GDP, which also corresponds to correlation analysis. The reason for this may come from the significant influence of futures markets on the existence of economies. For oil price, the regression coefficient is negative and the correlation coefficient is positive, which may be caused by collinearity. Although the model passes the multicollinearity test, the model still has collinearity, because the correlation between the predictor variables can be seen in the correlation analysis.

4. Conclusion

Since the updating frequency of GDP is relatively slow, this study aims to use gold price, oil price, and stock market index, three key data in economic development, which are updated with high frequency to estimate GDP, to make decisions faster. The correlation analysis of monthly GDP, monthly price of crude oil, monthly price of gold, and monthly S&P 500 index of the United States is carried out in this study, and the results show that these factors are positively correlated with each other, and the correlation is relatively significant. The results of cross-correlation analysis show that there is no significant lag between GDP and other variables. In this study, the monthly GDP of the United States is taken as the predicted variable, the monthly price of crude oil, the monthly price of gold, and the monthly s&p500 index are taken as the predicted variable, and a multiple linear regression model is established. The polynomial regression equation obtained can be used to further explore the impact of crude oil, gold, and the stock market on GDP, and can be used to further forecast GDP.

References

- [1] Arslanalp S, Eichengreen B, Simpson-Bell C. Gold as International Reserves: A Barbarous Relic No More? 2023.
- [2] Pan Z, Wang Q, Wang Y, et al. Forecasting US real GDP using oil prices: A time-varying parameter MIDAS model. *Energy Economics*, 2018, 72: 177-187.
- [3] Nan Y, Sun R, Zhen Z, et al. Measurement of international crude oil price cyclical fluctuations and correlation with the world economic cyclical changes. *Energy*, 2022, 260: 124946.
- [4] Hyndman R J, Athanasopoulos G. *Forecasting: principles and practice*. OTexts, 2018.
- [5] Ge Y, Tang K. Commodity prices and GDP growth. *International Review of Financial Analysis*, 2020, 71: 101512.