Identifying the modelling on Gold price using multiple linear Regression under Indian environment

M. Narsimulu

Assistant Professor (Contract)
Department of Applied Statistics
Telangana University
Nizamabad, Telangana State, India.

Abstract- **This Research paper explores how the price of gold will be influenced directly or indirectly by the price of crude oil, interest rates (repo rates), Indian currency prices in dollars, the Sensex (BSE), inflation rates, and the US dollar. Multivariate regression analysis refers to regression using a dependent variable and a number of independent variables. Forecasting is crucial to econometrics and aids in the creation of government policies that are optimum. Business decisions that depend on the cost of such goods might gain from reliable forecasting. We will quickly review the regression model's error mean squares, which will help us determine how well the model predicts the future .The data spans a large period of time and is accessible from the first of October 2000 to the first of August 2020. The statistical interpretations of these factors are used to create a forecast model for the gold price in India, which depends on six variables. The independent variables were the cost of crude oil, interest rates (repo rates), and the value of Indian rupee in US dollars, the Sensex (BSE), the rate of inflation, and the US dollar. The model passes many tests, including as Adjusted R Squared, with high favoring values, and the ANOVA test infers that the model is significant. In addition to Residual Analysis, which shows that the model is a good fit and acceptable, the model is passed as a perfect fit. The extended time period during which the data were collected and the absence of any missing values made the Regression model more favorable. Strong correlations between the gold price and the USD to INR, CPI, and Sensex values were found.**

*Keywords***: Multiple linear Regression, Forecasting Gold price, Residual analysis.**

1. Introduction

Since many years ago, gold has made a significant contribution to the Indian economy. It has functioned as a decoration as well as a safe doorway during periods of endless inflation. Due to a variety of factors, gold is a key component in almost every portfolio. To name a few, it has outperformed other investments in terms of returns when the global market for other goods was weak, offering extremely high returns of 13.66% over the past 15 years, just slightly less than the Sensex returns of 13.97% over the same period. Gold is also attractive because it is thought to have little connection with other assets, which lowers risk by diversifying portfolios. It is incredible that the demand for yellow metal increased by more than 120% on a yearly basis. Since India accounts for a significant portion of the global demand yet only produces 0.75 percent of the world's gold, India is the second-largest importer of gold after China. India is the second-largest importer of gold since there aren't enough indigenous gold deposits to meet demand. The investment sector is where gold is most in demand. Jewellery sector is followed by industrial demands, according to Thailand Statistician, 2019; 17(1): 125–131. Although customers value gold as an asset, it also contributes significantly to the current account deficit, which poses a threat to the nation's financial stability. As an idle asset, gold contributes to the problems of financial instability. Thus, customers really benefit from gold price modelling. The goal of the current study is to use seven financial factors, namely, to build a regression model to forecast the price of gold. Crude oil prices, interest rates (repo rates), the value of Indian rupee in US dollars, the Sensex (BSE), the US dollar's inflation rate, and gold demand. In order to avoid multicollinearity, the multicollinearity of these variables was tested, and the regression model was created using principle component analysis (PCA). It would be possible to estimate the relative changes in the regressed variables by doing a regression analysis of the price using several time-variant predictors for gold, such as demand for gold, the price of oil, and the stock index.

2. Related works

Ismail et al. (2009) forecast the price of gold in the Malaysian market while taking into account the era of US inflation. The Commodity Research Bureau future index (CRB), the USD/EUR foreign exchange rate (EUROUSD), the inflation rate (INF), the money supply (M1), the New York Stock Exchange (NYSE), the Standard and Poor 500 index (SPX), the Treasury Bill (T-BILL), and the US dollar index (USDX) have all been included in the model as explanatory variables. By taking into account the lagged variable values as well, they have expanded their research. Toraman et al.'s (2011) analysis, which employed a variety of variables as determinants, came to the conclusion that the US exchange rate and oil price had the largest correlations, but the latter is negative and follows it in importance. Principal component analysis-based multiple linear regression modelling is extensively discussed in the literature. Ul-Saufie et al. (2011) utilized a multiple linear regression model with principal component analysis to try to forecast the PM10 concentration in Seberang Prai, Pulau Pinang. By using PCA, the study intended to increase its prediction potential. Results showed that the use of principal component as inputs improved multiple linear regression models prediction by reducing their complexity and removing data collinearity. Performance indicators used in the analysis included prediction accuracy (PA), coefficient of determination (R Square), index of agreement (IA), normalized absolute error (NAE), and root mean square error (RMSE). Sopipanet al. (2012) created a multiple regression utilizing PCA forecasting model for the SET50 index (the stock

prices of the top 50 listed businesses on the SET (Stock Exchange of Thailand)). The strong correlation between the variables posed a hurdle for the regression model's development; PCA was used to solve this problem. Just 3 major components were extracted from the model's 18 explanatory variables. These elements account for 93.35% of the data's volatility. Saravanan et al. (2012) used regression analysis based on PCA to estimate the demand for energy in India. To build the model, 11 explanatory variables are employed. Principal component analysis (PCA), a novel approach based on artificial neural networks (ANNs), is also used. The ANNs were tested using data from 10 years and 29 years of training data. Multiple linear regression (based on the original data and the main components) and artificial neural networks (ANNs) with the original data as input variables are compared.

3. Methodology

One of the most effective statistical methods leverages the least squares approach for parameter estimation: multiple linear regression modelling. The proposed model is

If there are p predictor variables, then the following is the form of a multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

Where Y is the response variable, X_i is the jth predictor variable, β_i is the average effect on Y of a one unit increase in X_i, holding all other predictors fixed, ε is the error term

The values for $\beta_0, \beta_1, B_2, \ldots, \beta_p$ are chosen using the least square method, which minimizes the sum of squared residuals (RSS):

$$RSS = \Sigma(y_i - \hat{y}_i)^2$$

$$R^2 = \frac{Sum\ of\ square\ due\ to\ Residual}{Sum\ of\ square\ due\ to\ Total}$$

How well a linear regression model "fits" a dataset is gauged by its R-squared value. R^2 is also called the coefficient of determination, measures the variance in the response variable by the predictor variable. R-squared value might be between 0 and 1. A value of 0 means that the regressor variable has no effect on the response variable at all. If the predictor variable has a value of 1, it can completely and error-free explain the response variable. Regression MS / residual MS is the formula used to obtain the regression model's overall F statistic, or F. Significance The p-value for the total F statistic is represented by the letter F. It reveals the statistical significance of the regression model as a whole. In other words, it reveals if there is a statistically significant relationship between the dependent variable and the two or more explanatory factors taken together. The coefficient p-values, which are the individual p-values, indicate the statistical significance of each explanatory variable.

4. Data Analysis

Data on the price of gold during a 21-year period (2000–2020) and six explanatory factors are gathered from a variety of sources. The World Gold Council's official website and the Reserve Bank of India's official website are where the information on the price and demand for gold is gathered. Crude oil price, Interest rate (repo rate), Indian currency exchange rate in US dollars, Sensex (BSE), Inflation rate, and US Dollar are the explanatory variables. The analysis of variance (ANOVA) test is used to determine the significance of the R square value, which measures how well the model fits the data. The R square value of 0.989 shows that the models account for 98.9% of the volatility in gold price. The R square by ANOVA significance test results in a test statistic of 3722.144 and a p-value of 0.05, indicating that the model fits the data well.

Goodness of the fit of the model

R	R Square	Adjusted R Square	Standard error of the estimate	ANOVA F	p-value
0.9948	0.989	0.98516	2319.664452	3722.144	0.00

Residual Analysis

The standard deviation is rather small, and the mean value of the residuals is 0. Additionally, the residuals' frequency map exhibits normalcy. There is no relationship visible in the residuals' scatter plot when compared to the dependent variable. In light of this, we may say that the residuals are random.
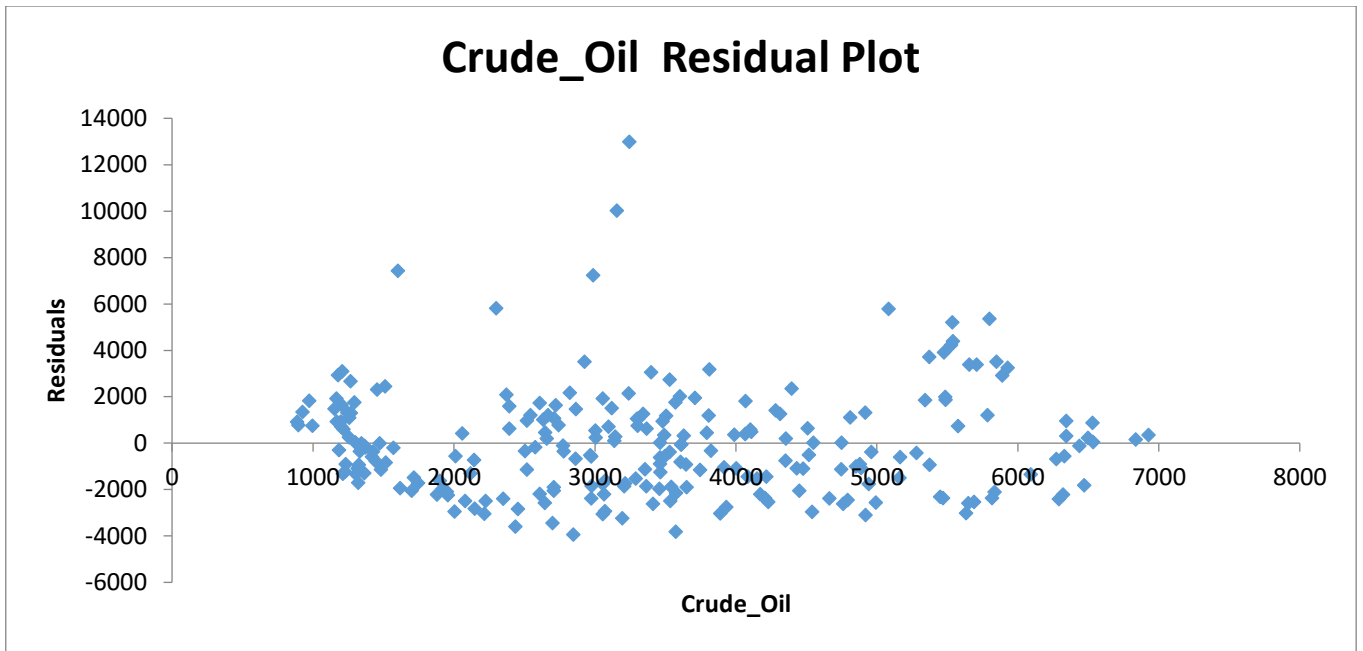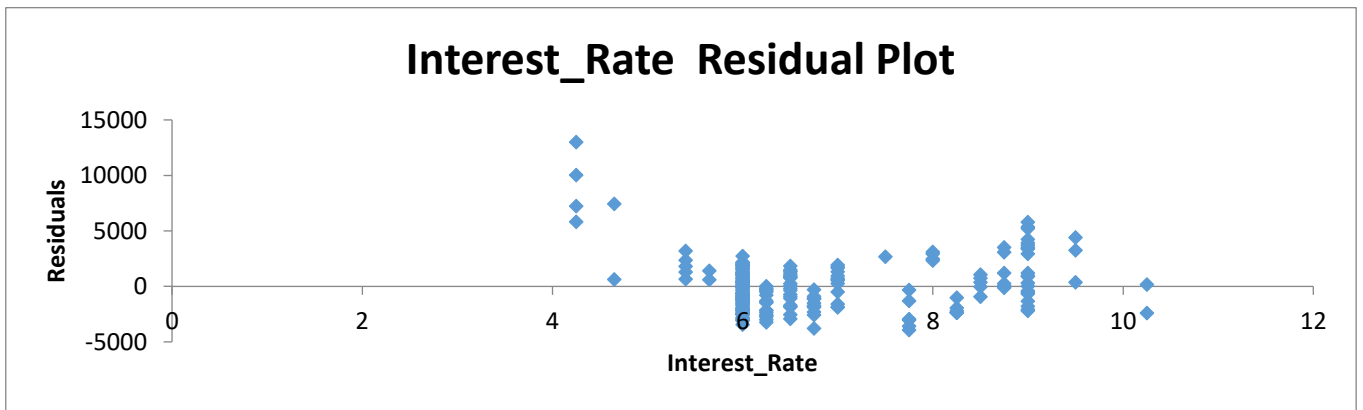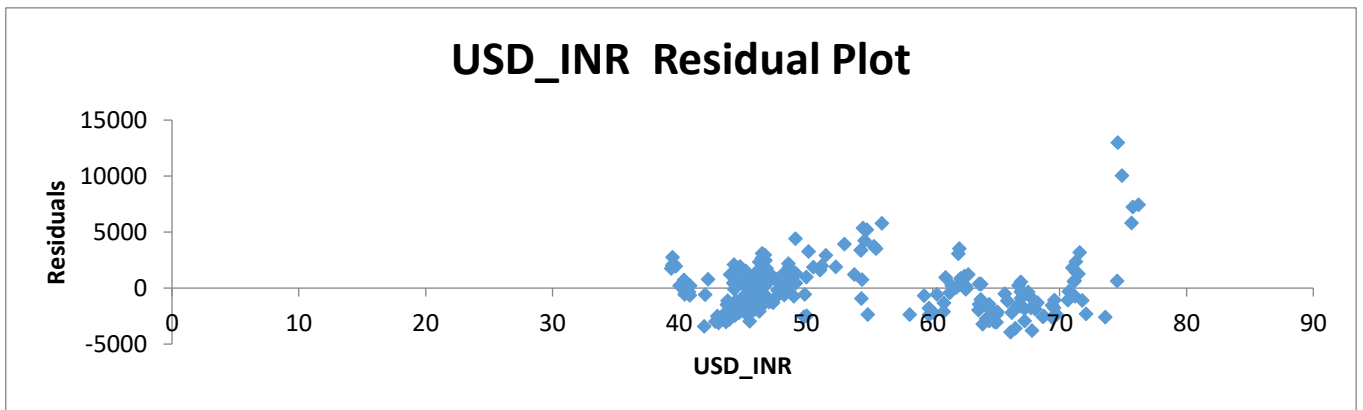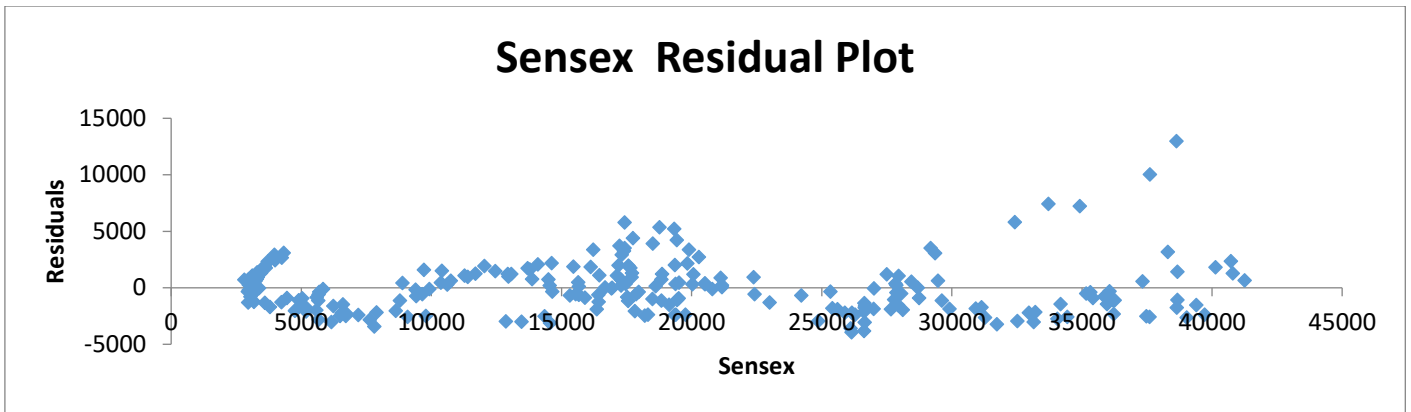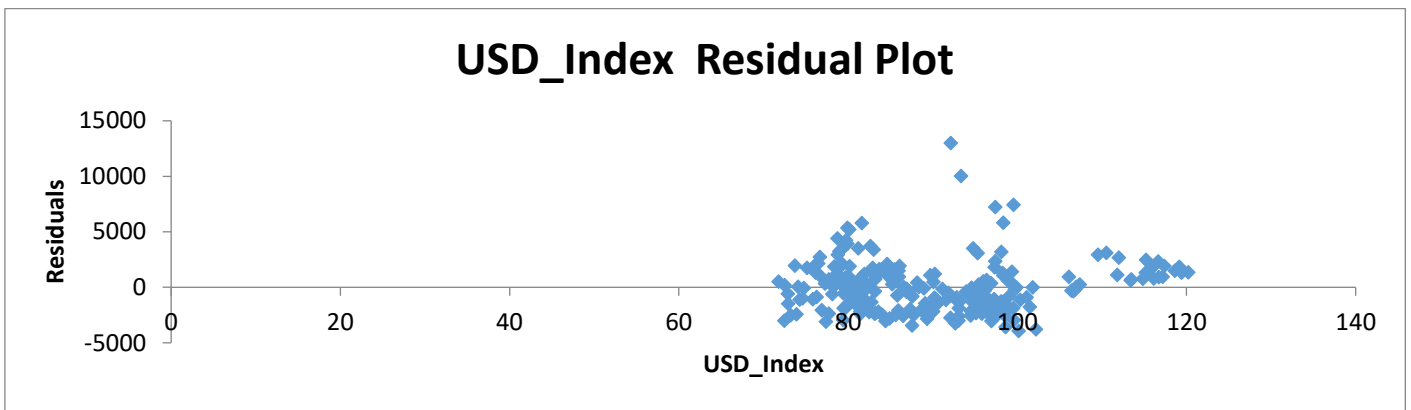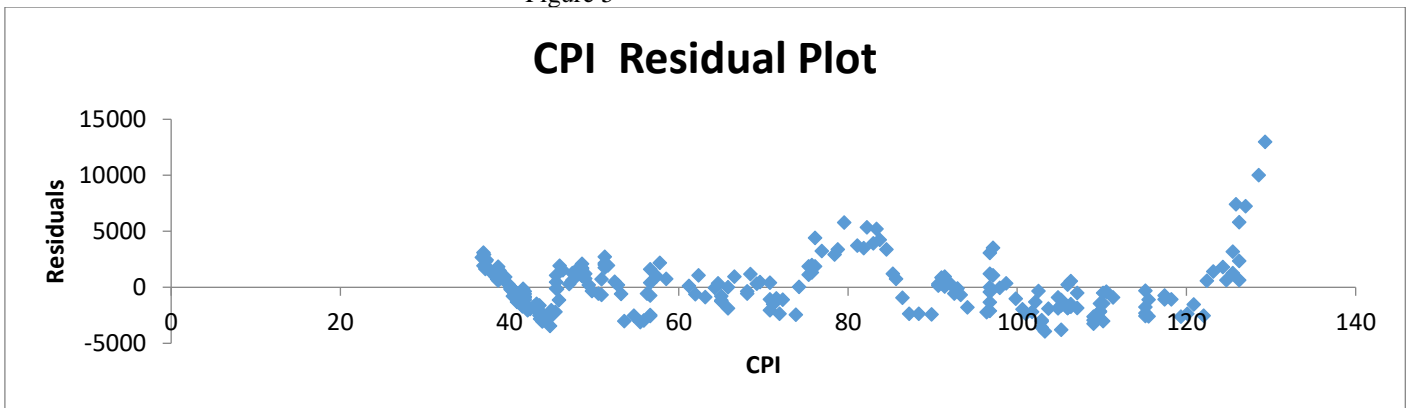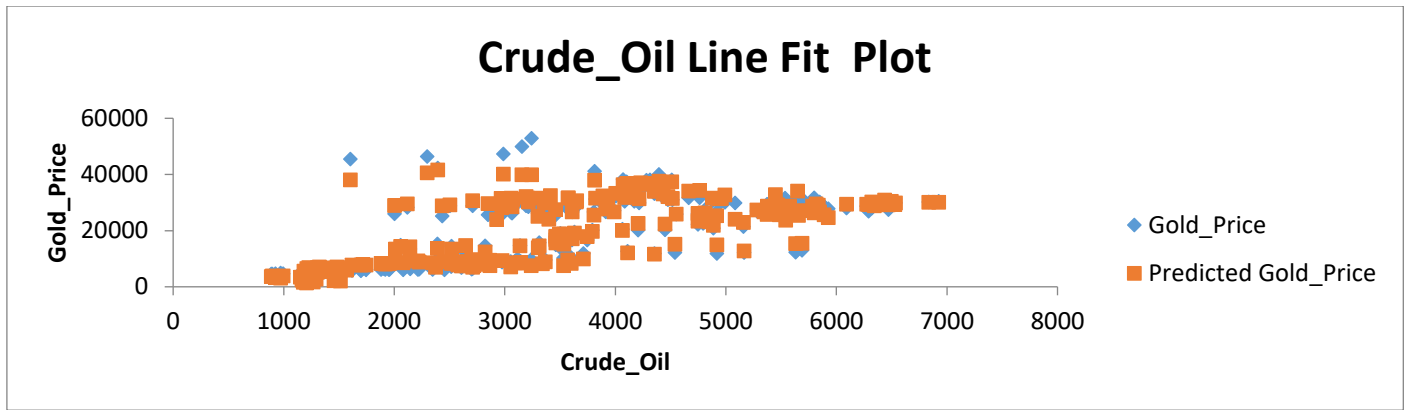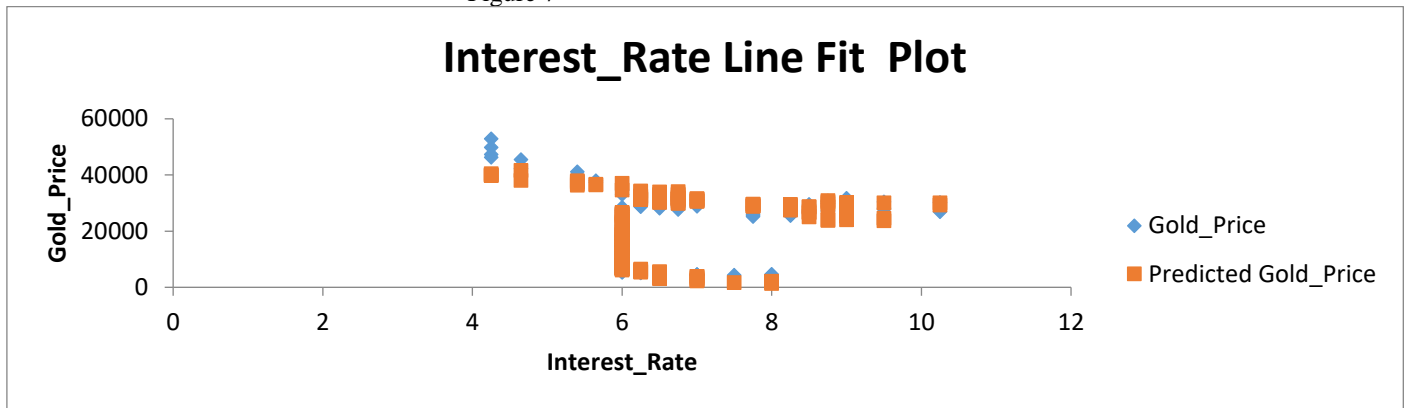
Figure 1

Figure 2

Figure 3

Figure 4

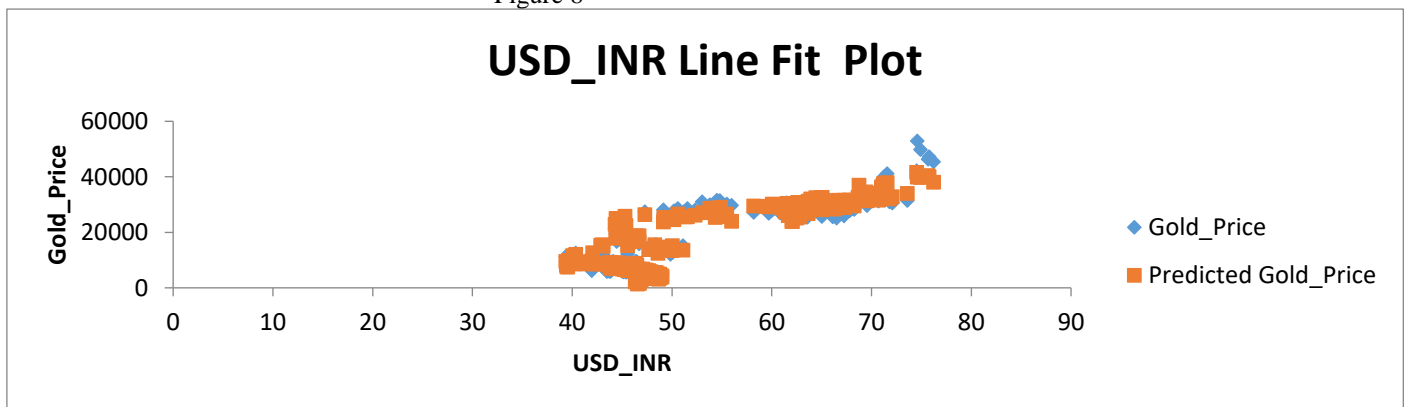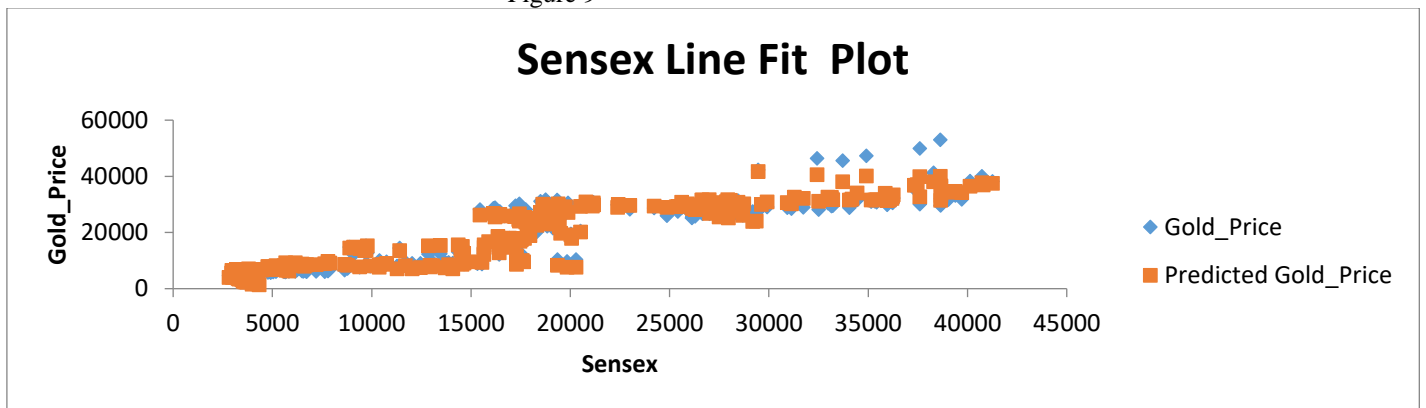Figure 5
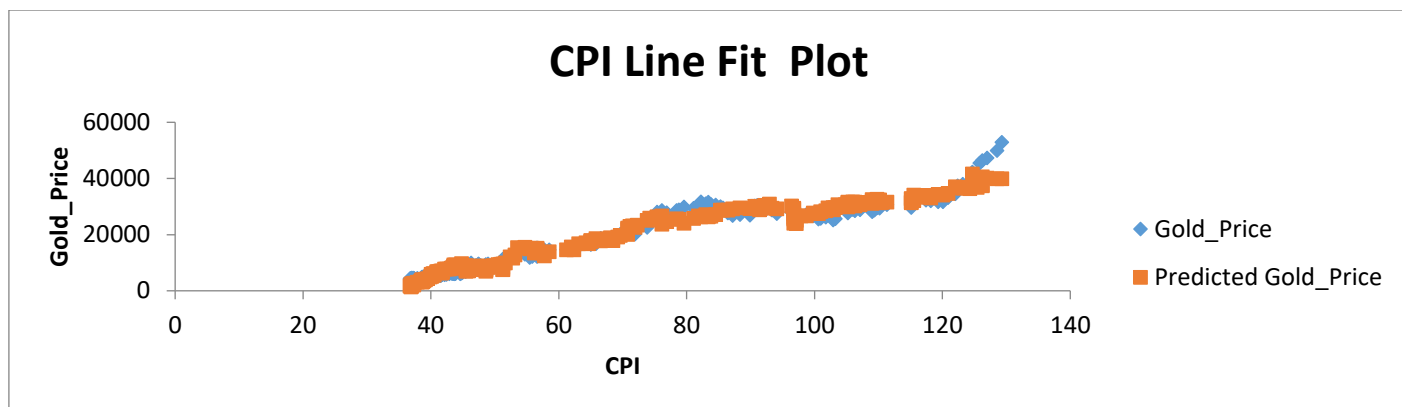

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11
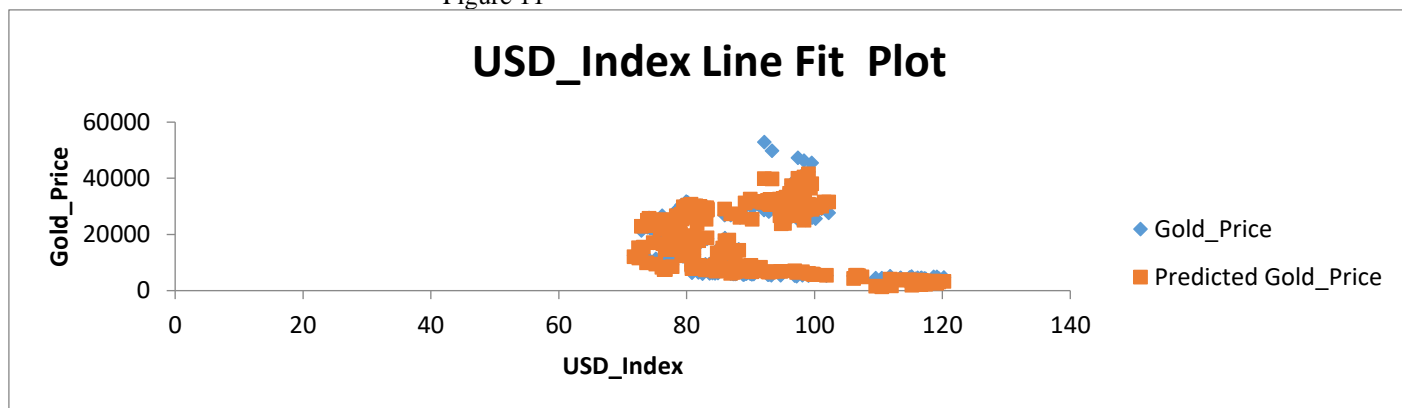

Figure 12

By observed above Residual Analysis figures, supports the good fit of the model.

Thus the model Gold Price = 1.108108346(crude_oil)-714.8842534(Interest_Rate)-315.1002258(USD_INR) - 0.699963599(Sensex) + 714.034807756597(CPI) -26.23926929(USD_Index)

5. Conclusion and future work

Using multiple linear regression from the six independent variables, a prediction model for the price of gold in India is created. The model is based on multiple linear regression, and the factors taken into account are the demand for gold, the exchange rate of the US dollar to the Indian rupee, the Sensex, interest rates, the price of crude oil in India, and the US dollar index. Gold Price = 1.108108346 (crude oil) - 714.8842534 (interest rate) -315.1002258 (USD_INR) -0.699963599 (Sensex) + 714.034807756597 (CPI) -26.23926929 (USD_Index) is the model that was generated. In India, the price of gold may be predicted using the model dependent on changes in exogenous factors. Six explanatory variables account for 98.9% of the variation in gold price in this MLR model. This MLR model has been shown to fit the data well, and the residual analysis that was done supports the validity of the model. In comparison to machine learning and deep learning techniques, multiple linear models have lower predicting accuracy. We firmly advise researchers to adopt machine learning and deep learning techniques to increase predicting accuracy. The data may be over a longer time period and other factors, such as the price of diamond and other precious metals, can be added to the model in the future, taking it to higher levels of optimality.

REFERENCES:
1. Gujarati D, Porter DC, Gunasekhar S. Basic econometrics. 5th edition, Special Indian Edition, McGraw-Hill Publications; 2012.
2. Ismail Z, Yahay A, Shabri A. Forecasting gold prices using multiple linear regression method. Am J Appl. Sci. 2009; DOI: 10.3844/ajassp.1509.1514.
3. Saravanan S, Kannan S, Tanagra C. India's electricity demand forecast using regression analysis and artificial neural networks based on principal components. ICTACT J Soft Compute. 2012; 2: 365-370.
4. Sopipan N, Kanjanavajee W, Sattayatham P. Forecasting SET50 index with multiple regression based on principal component analysis. J App Finance Bank. 2012; 2: 271-294.
5. Toraman C, Basarir C, Bayramoglu MF. Determination of factors affecting the price of gold: a study of MGARCH model. Bus Econ Res J. 2011; 2: 1309-2448.
6. Ul-Saufie AZ, Yahya AS, Ramli NA. Improving multiple linear regression model using principal component analysis for predicting PM10 concentration in Seberang Prai, Pulau Pinang. Int J Env Sci. 2011; 2: 403-410.