IDR : Vol. 15, No. 2, (July-December, 2017) : 77-92 | ISSN : 0972-9437 |

USING MULTIPLE REGRESSION MODEL FOR GOLD PRICE FORECASTING: ISSUES AND THEIR SOLUTIONS

Arijit Mitra*

ABSTRACT

The constant fluctuation of the gold price draws the attention of researchers and different models are used to predict the future price of this precious metal. This study attempts to use a multiple regression model to forecast the gold price and while doing so, the issue of multicollinearity, heteroscedasticity and auto-correlation issue of the historical data are addressed and solved with the help of suitable techniques. After that, the final model is proposed with the assessment of overall fit of the same. Using this proposed model, further, the future price of gold is forecasted and the errors are estimated. The paper ends with the limitation of the research and the way forward is discussed.

INTRODUCTION

The continuous fluctuation in gold price in the recent time significantly draws an attraction of the investors towards the precious metal. While individuals see it as an inevitable phenomenon, as gold is used both as a financial asset and in the jewelry sector, research should look into the economic rationale which is present in the behavior of the price gold. Although, There are several ways for predicting the price of any commodity, one of the most common way to find the economic logic in the fluctuation of the price is to build a causal model in which some (one or more) factors, based on existing literature, may be considered as the causes of the price fluctuation of the commodity in hand. In other words, the gold price can be treated as a dependent variable which is dependent on some factors (which are the cause). These causal variables, which are termed as independent variables, can then be used to predict the future price of the dependent variable (in this case, gold price) using multiple regression model. The model essentially is an equation which can be used to forecast the price of the gold in future, depending on the input values of the independent variables. The statistical significance of each of the causal/independent variable in the

* Fellow Student, Production, Operations and Decision Science, XLRI, Jamshedpur, India, *E-mail: fb10001@astra.xlri.ac.in*

equation can also be tested in order to find out the importance of each of the independent variable as far the gold price increase is concerned.

The primary objective of the study is, to investigate the prevailing literature to find out some relevant factors which can be used as independent variables to forecast the gold price in future with the help of multiple regression technique. In endeavoring so, three main issues namely, multicollinearity, heteroscedasticity and auto-correlation which are prevalent while using regression technique are taken care for error-free results. To solve those issues the help of statistical techniques are taken and lastly, the finalized model is used for the prediction of the gold price at a certain time period and compared with the actual data of the gold price at the same time period to estimate the percentage error in prediction.

LITERATURE REVIEW

The review of literature covers the parameters that are prevalently used for predicting gold price with the help of regression technique. In the past some studies have used the multiple regression to build up predictive models for the gold price (Abken, 1980; Koutsoyiannis , 1983; Vural, 2003; Ismail *et. al.,* 2009 etc.). Abken (1980) assumed that the increase in gold price is not a frenzy one and there is economic rationale present behind this phenomenon. Accordingly he attempted to find out the factors behind the phenomenon and came out with four broad factors namely, political and economic uncertainty, Flow supply and demand of gold, inflation and government auction policy. He developed a regression model of gold price as a function of interest rate and lagged value of gold price. In this work, he used the monthly data of gold price from January 1973 to December 1979. Although the explanatory power of the equation was quite low, a similar relation was developed for the future spot price with interest rate and future price lagged by three month and explanatory power of this equation was higher. So, finally he concluded that the estimation of gold price, future spot price and even other storable commodities are similar. Only the degrees by which various economic factors influence their price are different.

Koutsoyiannis (1983) found that gold price is dependent on the US economy rather than the worldwide economic situation. He found that gold price are expressed as a function of US dollar and ultimately a negative relationship between US dollar and gold price were found. This research also shows that regression is used for estimating gold price. Dooley *et. al.,* (1992) similarly studied the relationship between the exchange rate and the gold rate and built a VAR model, it is found out that the relation of USD and other currencies explain the change in gold price. This model is also a causal model which is developed for gold price forecasting.

In another study, Ghosh *et. al.,* (2002), attempted to investigate the effects of worldwide inflation level, USA inflation level, world-wide income, value of USD and random shock on gold price with the help of VAR model. In this study it is concluded that the gold price has a relationship with USA inflation level, interest rate and USD exchange rate. Also a long term relation has been found between the gold price and US Consumer Price Index (CPI) as a result of a cointegration analysis.

Vural (2003) tested the relation of gold price with different causal variables namely, USD/Euro parity, Dow Jones industrial production index, oil prices, interest rates, silver and copper prices and developed a model where he found the negative relation of gold price with interest rates, USD/ Euro parity and Dow Jones industrial production index and positive relation of gold price with silver, oil and copper prices. This is an example of typical multiple regression which is used for forecasting gold price.

In another similar study, Topçu (2010) examined the relationship of gold price with several variables Dow Jones industrial production index, US Dollar exchange rate, oil prices, US Inflation rate, Global Money Supply (M3) with the help of data from 1995 to 2009. He also developed a multivariate regression model where gold price was taken as a dependent variable while other all variables were taken as causal or independent variables. He found the positive relationship of gold price with Jones industrial production index, US Dollar exchange rate and oil prices while it was found that US Inflation rate and Global Money Supply (M3) negatively affect gold price. Among all the relations, the importance of oil prices and US Inflation rate were found to be statistically insignificant.

Apart from these studies, there are several studies which used regression method not only to predict the future gold price but to forecast any kind of commodities and stock markets also. In fact, Toraman et. al., 2011 has listed a lot literature where different types of regression models have been used to predict gold prices. Also they used a MGARCH model just to deal with the heteroscedasticity problem which are very common for the time-series or longitudinal data. According to Toraman *et. al.,* 2011, mainly Engle (1982)'s ARCH (Autoregressive conditional heteroscedasticity) model and Bollerslev (1986)'s GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models are the models which are pioneers as far as dealing with the heteroscedasticity problem in regression is concerned. Toraman *et. al.,* 2011 also argued that different models such as Exponential GARCH (EGARCH), ARCH-M (ARCH in mean), T-ARCH (Threshold ARCH), C-ARCH (Junction ARCH), PARCH (Asymmetric ARCH) are derived afterwards. The present study is inspired by the study and similar attempt to correct the three main problems. The details of those methods are discussed in the later section of the paper.

METHODOLOGY

The paper attempts the prediction of the future gold price with the help of a multiple regression model (Gujarati, 2004). As discussed earlier, effort is also made to figure out how much the model is affected by a) Multicollinearity b) Heteroscedasticity and c) Auto-correlation and finally those issues are handled with the help of standard methods described in the textbooks which deals with the time series data. Finally, the corrected model is used to forecast the gold price with the help of the independent variables data and errors with the actual price data of gold price were checked.

For the multiple regression model, the gold price (in dollars per ounce) is chosen as the dependent variable and three independent variables namely, Per capita income (in dollars per year), Returns on 90 days T-bills (in percentage) and M2 money Stock (in billions of dollars). The variable details are given in Table 1.

Table 1
The Details of the Multiple Regression used

Gold Price (In dollars per ounce)	Dependent variable
Per capita income (In dollars per year)	Independent variable
Returns on 90 days T-Bills (%)	Independent variable
M2 Money stock (Billions of Dollars)	Independent variable
Data Type	Time series
No. of observations	30

The paper covers the average yearly gold price (in dollars per ounce) from 1981 to 2010. The data is taken only for the period from 1981 to 2010 mainly because of two reasons. First, before that the historical value of gold, if used, may give significant forecasting error because the abrupt change in the value of money. Secondly, the returns on 90 days T-bills (in percentage) are available only for the period from 1981 to 2010. Again the rationale of taking the yearly gold price is same i.e. the availability of T-bill return data on a yearly basis. The data sources of various variables are indicated in the Table 2.

Then the multiple linear regression with Gold price (in dollars per ounce) as dependent variable and yearly Per capita income (In dollars), Returns on 90 days T-Bills (%) and M2 Money stock (Billions of Dollars) as independent variables is conducted. Minitab 15.0 version software is used for conducting the regression. The equation used can be given as,

$$Y_t = \beta_0 + \beta_1 * X_{1t} + \beta_2 * X_{2t} + \beta_3 * X_{3t} + \varepsilon_t \quad (1)$$

where,

Table 2
The Details of the Data Source

Parameter	*Data Source*
Gold Price (In dollars per ounce)	*http://www.kitco.com/charts/historicalgold.html*
Per capita income (In dollars)	Source: ERS International Macroeconomic Data Set and Reference: *http://www.ers.usda.gov/Data/Macroeconomics/*
Returns on 90 days T-Bills in percentage	*http://w4.stern.nyu.edu/~adamodar/New_Home_Page/datafile/histret.html*
M2 Money stock (Billions of Dollars)	Source: Board of Governors of the Federal Reserve System and reference: *http://research.stlouisfed.org/fred2/series/M2*

Y_t = Average yearly Gold price (in dollars per ounce) at time period t

X_{1t} = Yearly Per capita income (In dollars) at time period t

X_{2t} = Returns on 90 days T-Bills (%)at time period t

X_{3t} = M2 Money stock (Billions of Dollars) at time period t

ε_t = Error or unexplained part of the equation at time period t

β_0, β_1, β_2 and β_3 are regression co-efficients; β_0 being the constant terms and β_1, β_2 and β_3 are the co-efficients corresponding to X_{1t}, X_{2t} and X_{3t} respectively.

The result of the multiple regression as found from the Minitab 15.0 software is given in Figure 1a.

The software also returns some additional information which are used for the further analysis. These are shown in the Figure 1b.

INTERPRETATION OF THE RESULTS

The t-test and F-test of the model are also conducted to test the statistical significance of each of the regression co-efficient i.e. β_0, β_1, β_2 and β_3. Result shows that while β_0, β_1, and β_3 are significant at $p = 0.05$, β_2 is not significant at $p = 0.05$ level [Figure 1(a)].

For multiple regression, the F-test is also required to test whether all the co-efficient values i.e. β_0, β_1, β_2 and β_3 are same or different. An F value of 76.22 is obtained and the corresponding p-value is found to be 0.000, which indicates that the coefficients are different from each other [Refer Figure 1(a)].

The R2 value is found to be 0.898 i.e. 89.8% of the variation in the variable Y is explained by the three predictor variables. In case of multiple regression R2 value can be increased with the increase in the no. of predictor variables even if all the variable(s) are not important. So, calculation of the Adjusted

R2 value, which is the value of R2 adjusted for the number of terms in the regression function here, gives better understanding about the degree of fit of the model in hand. In this case, adjusted R2 is found to be 0.886 [Figure 1(b)].

Figure 1a: The Output of Preliminary Regression Model

The regression equation is
Gold Price per ounce (Y) = 1463 - 0.0637 Per capita income (X1)
+ 14.0 Returns on 90 days T-Bills(%)
+ 0.265 Money stock (M2)

Predictor	Coef	SE Coef	T	P	VIF
Constant	1462.8	196.3	7.45	0.000	
Per capita income (X1)	-0.063665	0.006772	-9.40	0.000	8.317
Returns on 90 days T-Bills(%)	13.959	7.531	1.85	0.075	3.020
Money stock (M2)	0.26526	0.01979	13.40	0.000	8.477

S = 73.6271 R-Sq = 89.8% R-Sq(adj) = 88.6%

Figure 1b: Some additional information for the Preliminary regression model

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	1239585	413195	76.22	0.000
Residual Error	26	140945	5421		
Total	29	1380530			

Source	DF	Seq SS
Per capita income (X1)	1	248256
Returns on 90 days T-Bills(%)	1	17413
Money stock (M2)	1	973916

Unusual Observations

Obs	Per capita income (X1)	Gold Price per ounce (Y)	Fit	SE Fit	Residual	St Resid
1	26070	459.7	444.4	46.6	15.3	0.27 X
30	42189	1224.5	1065.4	43.9	159.1	2.69R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Durbin-Watson statistic = 1.13148

Multicollinearity Issues and the Solution

The high R square value with insignificant β_2 value in the model raised a question of multicollinearity present in the independent variables. So, the multicollinearity problem of the multiple regression model is tested with the Variance VIF method. In this method, X_{1t} regressed against X_{2t} and X_{3t}, then X_{2t} and X_{3t} are also regressed against the other two independent variables in the same way. The R2 value of each of the regression is used to calculate the VIF using the formula:

$$VIF_j = \frac{1}{(1 - R_j^2)}$$

Where, j = 1, 2, 3….k and R_j^2 is the co-efficient of determination from the regression of jth independent variable on the remaining k-1 variables. If VIF for any variable is greater than 5, high multicollinearity is there and if VIF for any variable is greater than 10, the multicollinearity is considered severe [For details please refer book by Gujarati, (2004)].

The result of the VIF test for multicollinearity is calculated by Minitab software and the value is given in Figure 1a. According to this value, it is found that multicollinearity exists between the variables X_{1t} and X_{3t} because corresponding VIF values are found to be 8.317 and 8.477 respectively, both of which are greater than 5. The variable X_{2t} is found to be free from multicollinearity because the corresponding VIF value is found to be 3.020 which is less than 5.

To remove the multicollinearity between X_{1t} and X_{3t}, at first the variables are transformed by first difference method i.e. in case of each variable, the differences between two consecutive observations the time-series data are computed [e.g. In case of Y_t, the difference between the observation at time period t and the observation at time period (t -1) is taken and so on.]. The result of the corresponding VIF test and the multiple regression with first difference of each of the variable are provided in Figure 2.

The results in Figure 2 indicate that although the VIF is within limit, the regression co-efficient are insignificant. It was found that in the modified model, R2 value is 23 percent and adjusted R2 value is 13.7 percent, both of which are very poor. So, the multicollinearity may be solved with the first difference method but the regression model is not a good fit with the same. Secondly one of the variables between X_{1t} and X_{3t} (those which are having high VIF) is omitted and the regression is performed again with only X_{2t} and X_{3t}. The regression result is found comparatively well with an R^2 value of 0.55 and multicollinearity problem is also solved. The result is shown in Figure 3:

Figure 2: Result of Regression with First Difference of all the variables

```
The regression equation is
Y(t) - Y(t-1) = - 16.6 - 0.0138 FD(X1)= X1(t) - X1(t-1)
                + 11.0 FD(X2) = X2(t) - X2(t-1) + 0.234 FD(X3) = X3(t) - X3(t-1)


29 cases used, 1 cases contain missing values


Predictor                    Coef  SE Coef      T      P    VIF
Constant                   -16.59    36.33  -0.46  0.652
FD(X1)= X1(t) - X1(t-1)  -0.01383  0.02413  -0.57  0.572  1.704
FD(X2) = X2(t) - X2(t-1)    11.05    11.05   1.00  0.327  1.461
FD(X3) = X3(t) - X3(t-1)   0.2337   0.1038   2.25  0.033  1.245


S = 70.1349   R-Sq = 23.0%   R-Sq(adj) = 13.7%

PRESS = 162185   R-Sq(pred) = 0.00%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       3   36712  12237  2.49  0.084
Residual Error  25  122972   4919
Total           28  159684


Durbin-Watson statistic = 1.47583
```

Figure 3: Result of Regression after omitting X_1 Variable

```
The regression equation is
Gold Price per ounce (Y) = - 203 + 30.3 Returns on 90 days T-Bills(%)
                           + 0.114 Money stock (M2)


Predictor                         Coef  SE Coef      T      P    VIF
Constant                        -203.0    173.9  -1.17  0.253
Returns on 90 days T-Bills(%)    30.34    15.08   2.01  0.054  2.858
Money stock (M2)               0.11379  0.02365   4.81  0.000  2.858


S = 151.540   R-Sq = 55.1%   R-Sq(adj) = 51.8%

PRESS = 809721   R-Sq(pred) = 41.35%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2   760493  380246  16.56  0.000
Residual Error  27   620037   22964
Total           29  1380530


Durbin-Watson statistic = 0.291660
```

So, by using X_{2t} and X_{3t} as predictor variable, the multicollinearity problem can be solved and the R^2 value = 0.55; which is a medium fit model. Or else, all the three predictor variables can be taken (i.e. do nothing technique) because in the original equation given in Figure 1(a), R^2 value is 0.898 i.e. a good fit. To do nothing is decided in this case as because in that case the R2 value is excellent.

Heteroscedasticity Problem of the data and its solution

Then the heteroscedasticity problem has been analyzed. In this case, White test for heteroscedasticity, which follows a χ^2 distribution, has been conducted [For details please refer book by Gujarati, (2004)], where the auxiliary equation is:

$$\varepsilon^2 = \alpha_0 + \alpha_1 * X_{1t} + \alpha_2 * X_{2t} + \alpha_3 * X_{3t} + \alpha_4 * X_{1t}^2 + \alpha_5 * X_{2t}^2 + \alpha_6 * X_{3t}^2 + \alpha_7 * X_{1t} * X_{2t} + \alpha_8 * X_{2t} * X_{3t} + \alpha_8 * X_{3t} * X_{1t} + V_i \quad (2)$$

Where, V_i = Error or unexplained part of the equation

For White test, the test statistic is Lagrange multiplier (LM) where LM = (no. of observations)*R^2 of the auxiliary equation and if LM = n* R^2 χ^2 (p -1) (where p is the no. of estimated parameters used in the auxiliary equation, it is considered that heteroscedasticity is present in the original regression equation.

In this case, the estimated parameters are 10 and so, in case if LM is χ^2_9, it is evident that the data is having heteroscedasticity. White test is conducted using the excel sheet and the results found are given in Figure 4:

The result clearly shows that there is heteroscedasticity present in the data. In order to remove the same, we use White-Heteroscedasticity consistent standard errors (S.E) as because here the variance (α_i^2) is not known. In order to find the same, we find the S.E of ß0, ß1, ß2 and ß3 using corresponding error square (α_i^2) instead of the corresponding variance (α_i^2). These S.Es are called robust S.Es as these S.Es are full proof against heteroscedasticity still the corresponding t-values are also significant and so is the case in our model (Figure 5). It is therefore concluded that there is no problem with the heteroscedasticity.

The Auto-Correlation Issues and the Solution

The auto-correlation problem is first investigated by using Minitab. The Figure 6 indicates that the spikes are dying out gradually which indicates that autocorrelation is present in the data and there is a chance that the data has the auto-correlation problem where the error terms of two consecutive observations are dependent to each other which means it is likely that AR (1) model can handle this issue [For details please refer book by Hanke and Wichern, (2013)]. To investigate the same, Durbin-

Figure 4: Results of White's Test

Regression Statistics	
Multiple R	0.788634129
R Square	0.62194379
Adjusted R Square	0.451818496
Standard Error	4005.754712
Observations	30

LM = n*(R square)	18.6583137
χ Square critical value at P=0.05 for df 9	16.919
Remark	Heteroscedastic

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	527950809.1	58661201	3.655798463	0.007550769
Residual	20	320921416.2	16046070.8		
Total	29	848872225.3			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-278591.7972	204403.1914	-1.3629523	0.188047141	-704969.3818	147785.7874
X Variable 1	9.704098385	12.50085649	0.77627468	0.446670573	-16.37223125	35.78042802
X Variable 2	33083.49078	14826.36845	2.23139543	0.037259997	2156.228224	64010.75334
X Variable 3	13.5696404	31.26619465	0.43400358	0.668931625	-51.65049858	78.78977938
X Variable 4	6.86601E-05	0.00021696	0.31646432	0.754929393	-0.00038391	0.00052123
X Variable 5	-311.4809537	171.7559995	-1.8135084	0.084788577	-669.7576894	46.79578209
X Variable 6	0.004791457	0.002916934	1.64263476	0.11608897	-0.001293161	0.010876075
X Variable 7	-1.185122407	0.620003259	-1.9114777	0.070379994	-2.478426537	0.108181724
X Variable 8	-0.001951406	0.001393728	-1.4001337	0.176795589	-0.004858671	0.00095586
X Variable 9	2.531566757	2.007716051	1.26091872	0.221849465	-1.656455526	6.71958904

Figure 5: Standard Errors and Corresponding t Statistics values and their Significance

Variables	*Coefficients*	*Standard Error*	*t Stat*	P-Value	White-Heteroscedasticity consistent standard errors	*t Stat*	P-Value
Intercept	1462.824	196.3159788	7.45137363	6.5267E-08	NR	NR	NR
X Variable 1	-0.064	0.00677221	-9.4009535	7.55136E-10	0.002	-27.839	Significant
X Variable 2	13.959	7.531172686	1.85344	0.075200276	4.689	2.977	Significant
X Variable 3	0.265	0.019790344	13.4036493	3.47676E-13	0.008	32.450	Significant

Watson test is conducted with AR (1) model and the test statistics is D-W test statistics (d) was found to be 1.13. As $0 < d < d_L$ i.e. 1.21 (where dL is found out from the standard table (Figure 7), it is obvious that there is positive autocorrelation (ρ) in the data (Hanke and Wichern, 2013).

From the relation, $d = 2*(1 - \rho)$, the estimated value of ρ is found out to be 0.4342 which means any two consecutive error terms in the time series data are related to each other by the equation,

$$\varepsilon_t = \rho * \varepsilon_{(t-1)} \tag{3}$$

In order to solve the problem of auto-correlation, General Least Square (GLS) method is used [For details please refer book by Hanke and Wichern, (2013)]. As the value of ρ is not known, we take the estimate of ? (which is

Figure 6: Visual observation for auto-correlation in the data of Gold Price

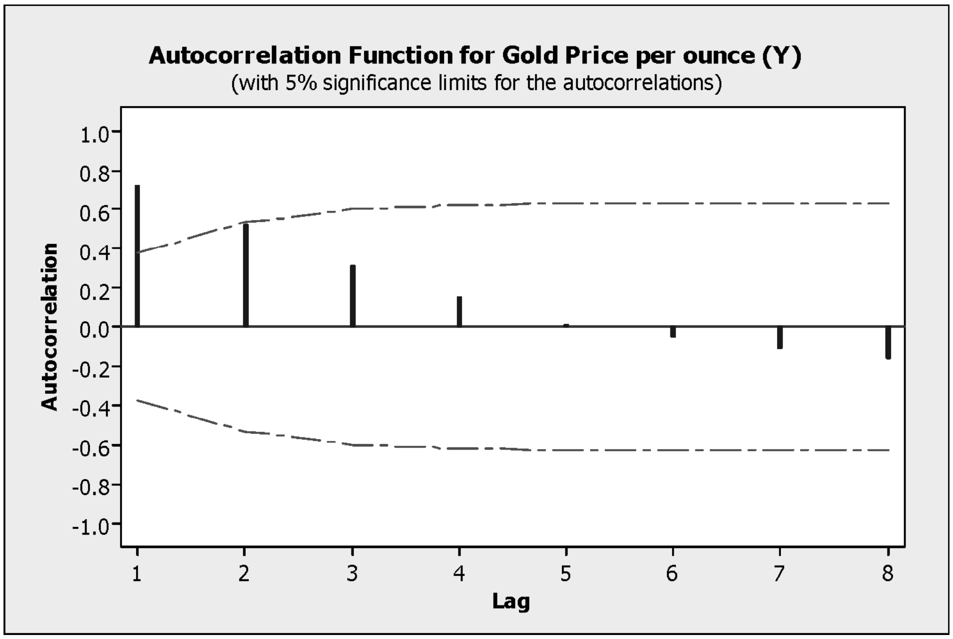

Figure 7: Results of D-W Test

D-W Statistics	1.131479597	
D(L)	1.21	For n=30 and K=3 at α = 0.05, from table
D(U)	1.65	
Conclusion	positive auto-correlation	
ρ	0.434260202	

calculated from D-W Test). Now considering the observation at time period t, the original regression equation of the time series data is,

$$Y_t = \beta_0 + \beta_1 * X_{1t} + \beta_2 * X_{2t} + \beta_3 * X_{3t} + \varepsilon_t \qquad (1)$$

And we have, $\varepsilon_t = \rho * \varepsilon_{(t-1)}$ (3)

Taking 1 lag and multiplying the equation (1) by ρ, the new equation obtained is,

$$\rho * Y_{(t-1)} = \beta_0 * \rho + \beta_1 * \rho * X_{1(t-1)} + \beta_2 * \rho * X_{2(t-1)} + \beta_3 * \rho * X_{3(t-1)} + \rho * \varepsilon_{(t-1)} \qquad (4)$$

(1) - (4) gives,

$$Y' = \beta_0 * (1 - \rho) + \beta_1 * X'_1 + \beta_2 * X'_2 + \beta_3 * X'_3 + e \text{ [As, } (\varepsilon_t - \rho * \varepsilon_{(t-1)}) = 0 \text{ by (3)]} \qquad (5)$$

Where,

$Y' = Y_t - \rho * Y_{(t-1)}$
$X'_1 = X_{1t} - \rho * X_{1(t-1)}$
$X'_2 = X_{2t} - \rho * X_{2(t-1)}$
$X'_3 = X_{3t} - \rho * X_{3(t-1)}$ and,
e = The error term of the regression equation (5)

The above method is called quasi difference method for solving auto-correlation. The result of the corresponding regression is given in Figure 8.

The R square value of the equation (5) is found to be 0.827, while all the other co-efficients, except β_2 are significant at $p = 0.05$ level.

The co-efficient β_2 is found to be insignificant while using the equation (1) and the same is found to be insignificant after all the corrections. So, the variable i.e. the Returns on 90 days T-Bills (%) at time period t is removed from the model. The final model which is considered for prediction of the gold price is,

$$Y' = \beta_0 * (1 - \rho) + \beta_1 * X'_1 + \beta_3 * X'_3 + e \quad (6)$$

Forecasting Gold Price with the help of the Final Model

Multiple regression is conducted with the proposed model. The details of the same can be seen in APPENDIX A. The result of the same is given in Figure 9. The Multiple R is found to be 0.897 with an R square value of 0.805 and adjusted R square of 0.79 which indicates an overall good fit. Also, all the co-efficient are found to be significant at p = 0.05. The value of different co-efficient which are obtained from the regression model are:

$$\beta_0 * (1 - \rho) = 793.552,\ \beta_1 = -0.0566 \text{ and } \beta_3 = 0.238$$

So, the equation (6) becomes,

$$Y' = 793.552\text{-}0.0566 * X'_1 + 0.238 * X'_3 + e$$

With this, Y' is forecasted and then it is transformed to Y (forecasted) with the help of the reverse equation i.e. $Y_t = Y' + \rho * Y_{(t-1)}$. Then the error from the actual gold price data is calculated. The Mean absolute Deviation (MAD), Mean Squared Error (MSE) and Mean absolute percentage error (MAPE) were calculated. From MSE, then the Root Mean Squared Error (RMSE) or Standard Deviation (α) is calculated. The results obtained are as follows.

MAD = 50.676 \$, MSE = 4463.87 and α = 66.81 \$, MAPE = 11.425%

Then the co-efficient of variation is also calculated by dividing the ? with mean of the actual Gold Price data and it is found to be 14.82%.

So, the final model proposed after the correction for multicollinearity, heteroscedasticity and auto-correlation is found to be a good fit for predicting Gold Price.

Figure 8: The Results of the Regression Equation after using Quasi Difference Method

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.907007979					
R Square	0.822663475					
Adjusted R Square	0.801383092					
Standard Error	68.59293866					
Observations	29					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	3	545660.9171	181886.9724	38.65830207	0.000000	
Residual	25	117624.7809	4704.991235			
Total	28	663285.698				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	731.6362733	141.4411022	5.172727462	0.00002	440.3328729	1022.939674
X Variable 1	-0.059009567	0.009630383	-6.127437279	0.00000	-0.078843712	-0.039175423
X Variable 2	18.15874111	11.45299712	1.585501237	0.12542	-5.429147814	41.74663003
X Variable 3	0.262376382	0.030215076	8.683624691	0.00000	0.200147268	0.324605495

Figure 9: The Summary Output of Multiple Regression with the Proposed Model

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.897124201							
R Square	0.804831832							
Adjusted R Square	0.789818896							
Standard Error	70.56155374							
Observations	29							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	2	533833.4435	266917	53.6092	6E-10			
Residual	26	129452.2545	4978.93					
Total	28	663285.698						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	793.5519434	139.8449102	5.67451	5.7E-06	506.097	1081.01	506.097	1081.01
X1(t) - ρ*X1(t-1)	-0.05665002	0.009787772	-5.78784	4.3E-06	-0.07677	-0.03653	-0.07677	-0.03653
X3(t) - ρ*X3(t-1)	0.237845407	0.026698007	8.90873	2.2E-09	0.18297	0.29272	0.18297	0.29272

LIMITATION OF THE STUDY

Although the study succeeds to recognize the errors in forecasting due to multicollinearity, heteroscedasticity and auto correlation present in the data set and attempts to solve those problems with proper methods, there are certain limitations in the present study. They are as follows:

Only 30 data points are considered for each of the variables while conducting the regression, which is too small. Actually the normality is an important assumption in regression and 30 no. is the lower limit for the normality assumption. So, the same study can be conducted with large no. of data to ensure more accuracy in the forecasted values. One possible way to conduct the same experiment is to take daily or monthly gold price with all the daily or monthly independent variables which could not be taken because of the non-availability of authentic database. Also, yearly data for gold price, sometimes, is not a proper representative because of the fluctuation of gold price during the year. For this reason also, daily or monthly data of gold price should be used.

Per capita income (in dollars per year), Returns on 90 days T-bills (in percentage) and M2 money Stock (in billions of dollars) are considered as independent variables. Although some literature support are presented as the rationale of collecting these parameters, a more careful choice of independent variables may give better results in terms of forecasted values of gold price. Also more than three variables can be introduced if they are all relevant according to previous literature or with any other justification and the multicollinearity, heteroscedasticity and auto correlation present in the data set may be identified following to the proper solution. In this case, the model will be more robust.

The same study could be done with the help of other advanced model of regression. Even a comparison between ANN technique and regression technique in terms of the output could be planned just to investigate the efficiency of regression and ANN in forecasting. Also, the multicollinearity, heteroscedasticity and auto correlation present in the data set could be studied with the help of more advanced technique which are described in various literatures and text books.

CONCLUSION

Forecasting is always erroneous and it is the job of a researcher to minimize this error or random variation. It is always difficult to predict a very dynamic market like gold, stock or future market and so, researchers have employed several techniques for predicting these markets. Still, research in this field needs to be more in number because of its uncertainty. Detailed study is required for main two purposes. First, the factors or causal variables for the gold or stock price should be chosen carefully and for that empirical surveys and economics related theoretical studies are necessary before choosing the parameters. Secondly all the methods of forecasting can be conducted simultaneously just to compare the best one or one with minimum error. Then only one can understand which causal variables should be used for gold or stock price prediction and which method should be used for predicting those markets. This paper is an attempt to use regression method with all possible corrections in the data set and the two models which we obtained as a result may be refined more to use for gold prediction theoretically.

References

Abken, P. A. (1980), The economics of gold price movements.

Dooley, M.P., Isard, P. and Taylor, M.P. (1992), Exchange Rates, Country Preferences and Gold. IMF Working Paper.

Engle, R. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, vol. 50: 987–1007.

Ghosh, D., Levin, E.J., Macmillan, P. and Wright, R.E. (2002), Gold As An Inflation Hedge? University of St. Andrews, Department of Economics, Discussion Paper Series.

Grudnilski, G and Osburn, L. (1993), Forecasting S&P and Gold Futures Prices: An Application of Neural Networks, *The Journal of Futures Markets*, vol. 13 (6): 631-643.

Gujarati. (2004), Basic Econometrics, Fourth Edition, The McGraw-Hill Companies.

Hanke, J. E. and Wichern, D. (2013), Business Forecasting, 9th Edition, Pearson Education, Inc.

Koutsoyiannis, A. (1983), A Short-Run Pricing Model for a Speculative Asset, Tested with Data from the Gold Bullion Market. *Applied Economics*, vol. 15: 563–581.

Mendelsohn, L., and Stein, J. (1991), "Fundamental Analysis Meets the Neural Network". Futures, vol. 20(10):22-24.

Topçu, A. (2010), Altın Fiyatlarını Etkileyen Faktörler. SPK Araþtırma Raporu, Ankara

Toraman, C., Başarır, C. and Bayramoðlu. M. F. (2011), Determination of Factors Affecting the Price of Gold: A Study of MGARCH Model. Business and Economics Research Journal, vol. 2 (4): 37-50

Vural, M.G. (2003), Altınpiyasasıvealtınfiyatlarınıetkileyenfaktörler (The gold market and the factors that affect the price of gold). (Unpublished dissertation). Central Bank of the Republic of Turkey.

Z.ismail, *et.al.* (2009), Forecasting Gold Prices Using Multiple Linear Regression Method. *American Journal of Applied Sciences*, vol. 6 (8): 1509-1514.

Zaremba, T. (1990), Case Study III: Technology in Search of a Buck. Neural Network PC Tools. A Practical Guide. (Eberhart, R.C., and Dobbins, R.W. Eds.). Academic Press, San Diego.

APPENDIX A
The Table Containing Regression Results

Gold Price (Y)	*Y(t)- ρ*Y(t-1)*	*X1(t) - ρ*X1(t-1)*	*X3(t) - ρ*X3(t-1)*	*Forecasted Y(t)-ρ* Y(t-1)*	*Forecasted Y*	*MAD*	*MSE*	*MAPE (%)*
459.7								
375.8	176.2	13999.6	1102.5	262.7	462.3	86.5	7486.0	23.02
424.2	261.0	15228.0	1257.6	230.0	393.2	31.0	960.4	7.31
360.4	176.2	16478.1	1328.8	176.1	360.3	0.1	0.0	0.03
317.2	160.7	16661.9	1452.2	195.1	351.6	34.4	1180.2	10.83
367.5	229.8	16995.1	1568.9	203.9	341.7	25.9	668.3	7.03
446.5	286.9	17353.7	1646.0	202.0	361.6	84.9	7210.5	19.02
437.0	243.1	18017.7	1728.4	183.9	377.8	59.2	3500.3	13.54
381.4	191.7	18411.3	1778.5	173.6	363.3	18.1	328.0	4.75
383.5	217.9	18294.0	1898.6	208.8	374.4	9.1	82.7	2.37
362.1	195.6	17691.3	1944.7	253.9	420.4	58.3	3400.2	16.10
343.8	186.6	18532.1	1953.9	208.4	365.7	21.9	477.6	6.36
359.8	210.5	18747.4	1962.7	198.3	347.6	12.1	147.3	3.37
384.0	227.8	19455.4	1993.3	165.5	321.7	62.3	3877.8	16.22
383.8	217.0	19494.6	2043.3	175.2	341.9	41.9	1752.5	10.91
387.8	221.1	20171.0	2185.5	170.7	337.3	50.5	2546.9	13.01
331.0	162.6	20918.0	2299.6	155.5	323.9	7.1	50.4	2.15
294.2	150.5	21564.7	2495.5	165.5	309.2	15.0	224.0	5.09
279.0	151.2	22424.3	2683.9	161.6	289.3	10.4	107.5	3.72
279.1	158.0	22988.2	2818.0	161.5	282.7	3.6	12.6	1.27
271.0	149.8	22507.4	3118.6	260.2	381.5	110.4	12190.6	40.74
309.7	192.0	22837.3	3332.4	292.4	410.1	100.4	10079.2	32.41
363.4	228.9	23352.2	3545.4	313.9	448.4	85.0	7230.3	23.40
409.7	251.9	24087.3	3659.3	299.4	457.2	47.4	2250.8	11.58
444.7	266.8	24532.4	3803.4	308.4	486.3	41.6	1731.4	9.36
603.5	410.3	24797.3	4028.8	347.0	540.1	63.3	4008.9	10.49
696.4	434.3	24908.8	4316.4	409.1	671.2	25.2	636.0	3.62
872.0	569.5	24189.3	4646.4	528.4	830.8	41.2	1696.0	4.72
972.4	593.7	22562.8	5038.2	713.7	1092.3	120.0	14396.4	12.34
1224.5	802.3	24248.7	4958.6	599.3	1021.5	203.0	41219.4	16.58
450.8				**793.6**	**1325.3**	**50.7**	**4463.9**	**11.43**
					Std. Dev.	66.81		
					COV	0.15		