Thailand Statistician
January 2023; 21(1): 180-195
http://statassoc.or.th
Contributed paper

Application of Time-Varying Coefficient Regression Model for Forecasting Financial Data

Jyothi Manoj*[a] and Suresh K K [b]
[a] Department of Statistics, Kristu Jayanti College, Bangalore, Karnataka, India.
[b] Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, India.
*Corresponding author; e-mail: jyothimanoj@kristujayanti.com

Received: 5 December 2019
Revised: 2 March 2021
Accepted: 21 March 2021

Abstract

Regression analysis using ordinary least squares method is a very powerful statistical technique widely used. Though this method is widely accepted, its ability to model accurately when the predictors are dynamic in time is suspected. In scenarios where we use panel data with endogenous variables possessing serial correlation, the role of time-varying coefficient regression is worth analyzing. The present study probes this aspect with price of gold in India as the dependent variable. The predictors considered are US dollar exchange rate, oil price, demand of gold and NIFTY. A comparative analysis of multiple linear regression modeling and time-varying coefficient linear regression modeling is carried out in the present study using a 7-year panel data set ranging from 2012 to 2019.

Keywords: Time-varying coefficient models, kernel smoothing, multiple linear regression, RMSE, Akaike information criteria.

1. Introduction

Time-varying parameters in linear regression and time series models have taken more importance these days than earlier days as they have proved their ability to capture dynamic structure of lengthy time series. Financial time series undergo structural changes during various policy changes and economic crisis or growth. The present study reports the semi-parametric modelling using time varying coefficients for predicting the price of gold. Linear models are most popular for their well-established properties and convenience in estimation and ease of interpretation as well. But these models are often unrealistic in applications. Other non-linear parametric models also have their own limitations. Non-parametric models pose various limitations like curse of dimensionality, greater variance for the estimators as it fails to incorporate some prior information etc. Hence an alternative approach suggested to relax the conditions of parametric models while exploring the hidden structure is varying coefficient models suggested by Hastie and Tibshirani (1993). These models are applicable when the study cannot assume the same structural parameters for all the observations. The models cannot be estimated by usual procedures as the inferences and estimation will not be able to provide the forecast accurately.

The price of gold has always been an interesting characteristic to be analyzed. Any drastic changes in financial market show an impact on the price of gold. The price of gold in India has taken significant fluctuations over the recent past, since 2010 (www.gold.org/Gold Demand Trends August 2010). In order to determine the reason for the turbulence in the price of gold in India, it is necessary to identify the factors that influence the change of price of gold. Manoj and Suresh (2019) reported that there are many factors that drive the price of gold in India of which the crucial ones are US dollar exchange rate, crude oil price, stock exchange rates, and demand of gold. All these variables are dependent on time as there are many studies that provide Autoregressive models to these variables. These variables exhibit autoregressive properties. Among the four, the influence of each one can be determined by correlation and regression analysis. Hence time-varying coefficient model is developed as a prediction model for price of gold in India. Further, the effect of these factors is also determined. The model built is analyzed whether it has better explanatory and prediction power.

In an economy, the determination of exchange rates is crucial in comprehending the links between the domestic and foreign economies. It is US dollar exchange rate that is most influential for Indian economy than others. Similarly, the price of crude oil plays a pivot role in Indian economy. Inflation rate of our country is heavily dependent on these two characters. Investment in stock/share market and gold are analyzed by many researchers. Many literature reports indicate when stock market is down, that suggests the best time for investment in gold. Demand of gold is also found to have a key role to play though both the series of prices or demand do not exhibit specific patterns.

2. Review of Literature

Varying coefficient models which was first proposed by Hastie and Tibshirani (1993) gives more flexibility than parametric linear models. This was one of the most significant advancements in general linear models and is being widely applied in finance, economics, atmospheric studies etc., where certain phenomenon tends to repeat periodically in the predictors thereby making the coefficient functions of other variables. Theoretical advancements in this concept are in progress to find the optimal/ best methods of estimation. Cai (2002) suggested a new approach to estimation of varying coefficient models which involves two-steps is illustrated in this study. They have proved that this method provides optimal estimators to the coefficients when some coefficients functions possess different degrees of smoothness. This method is claimed to have optimal convergent rate and same optimality as the ideal case where the other coefficient functions were known. Three types of variable selection problems are of practical interests in varying coefficient (VC) models: (i) separation of varying and constant coefficients (Huang et al. 2002; Wang et al. 2009); (ii) selection of variables with nonzero varying coefficients (Cai et al. 2000; Qu and Li 2006); (iii) selection of variables with nonzero constant coefficients (Fan and Huang 2005; Wang et al. 2009). Wang et al. (2009) in their research article discuss the flexibility and parsimony of a class of marginal partially linear quantile modes with varying coefficients. Their estimation of conditional quantile curves is without any specifications of the error distribution or intra-subject dependence structure. Estimation of quantile smooth coefficients using basis function approximation and the large sample properties of the estimators are presented in the paper.

Hastie and Tibshirani (1993) was the first article to bring to light the concept of varying-coefficient models. Different forms of varying coefficient models are also listed which indicates that though there are different instances, they have a common structure. The study describes the potential of varying coefficient models to describe the effect of a repressor to vary as a function of another factor reflecting the advantage of these models over regular regression models. Fan and Zhang (1999) brings out a new two-step estimation method for varying coefficients for data in which different coefficient functions admit different degrees of smoothness. Wu and Chiang (2000) developed an estimation method for longitudinal dependent variable and cross–sectional covariates. Application of the procedure is also presented on an epidemiological data. Huang et al. (2004) considers longitudinal data and develops non-parametric estimation of coefficient functions by approximating each coefficient function by a polynomial function by a polynomial spline and hence employed least squares for the estimation. Various developments in the theory and methods on VC models is discussed in Fan and Zhang (2008).

Assaf and Campostrini (2015) carried out an analysis of trends of behavior risk factor surveillance (BRFS) is carried out to comprehend the trends of health outcomes and risk factors using single effect modifier which is also called time- varying coefficient model. The estimation is carried out using non-parametric technique of P-spline estimation due to the flexibility provided and the reduced computation time that is required. This study demonstrates the use of varying coefficient models to provide the effect of possible determinants on health outcome/risk factor. Four different models were presented in Gupta (2012) to predict the demand of electricity in Delhi with predictor as apparent temperature. The models include a simple linear regression model, a semi-parametric additive model using unpenalized splines, a semi-parametric additive model with penalized splines and a variable coefficient model which will capture the time-varying impact of temperature on electricity demand. Beckmann and Czudaj (2013) attempted to analyze the ability of gold to be a hedge against inflation using data from four major economies (USA, UK, Euro area and Japan) between January 1970 to December 2011. Using Markov-switching vector error correction models (MS-VECM) the analysis includes both long- term and time-varying short run dynamics.

3. Theoretical Framework of Varying Coefficient Models

Consider a longitudinal data which are realizations of variables over continuous time stochastic process. It can be represented as $\{Y(t), \mathbf{X}(t),\ t \in \boldsymbol{T}\}$ where Y(t) represents the values of the dependent variable and $\mathbf{X}(t) = \left(X_0(t), X_1(t), X_2(t), ..., X_L(t)\right)'$ are the real valued covariates while $\boldsymbol{T}$ denotes the time interval at which the measurements are taken. The parametric modelling of the data of this type involves modelling the mean curves of $Y(t)$ based on the effect of the covariates. Non-parametric methods suffer the curse of dimensionality when there are many covariates. Hence, the appropriate method of dimensionality reduction approach for longitudinal data is time-varying coefficient model which is represented as

$$Y(t) = X'(t)\beta(t) + \varepsilon(t), \quad t \in T, \tag{1}$$

where $\mathbf{X}(t) = \left(X_0(t), X_1(t), X_2(t), ..., X_L(t)\right)'$, $\boldsymbol{\beta}(t) = \left(\beta_0(t), \beta_1(t), \beta_2(t), ..., \beta_L(t)\right)'$ and $X_0(t) = 1$, $\beta_0(t)$ represents the baseline effect, $\varepsilon(t)$ is a stochastic process with variance function $\sigma_\varepsilon^2(t)$ and covariance function $C_\varepsilon(t_1, t_2)$ for $t_1 \neq t_2$ and $\mathbf{X}(t)$ and $\boldsymbol{\varepsilon}(t)$ are independent. Equation (1) is a linear model for each fixed time **t** but allows the coefficients to vary with time. It has a meaningful interpretation while retaining the general nonparametric characteristics. In the case where the coefficient vector is subject to changes over time, $\boldsymbol{\beta}(t)$ is specified by a dynamic process, say, autoregressive process. The model can be then considered as a particular case of Gaussian state space model and the estimation process can be carried out based on the Kalman filter. Kalman filter works with a recursive algorithm for $t = 1, 2, ..., n.$ The smoothing method produces the estimates of the estimates of $\beta_t = E(\beta_t \mid y_1, y_2, ...y_n)$ recursively backwards in time for $t = n, n-1, ..., 2, 1.$ The corresponding variances are also computed by these state space models.

In the state-space model framework, the Kalman filter estimates the values of a latent, linear, stochastic, dynamic process based on possibly mismeasured observations. Given distribution assumptions on the uncertainty, the Kalman filter also estimates model parameters via maximum likelihood. Starting with initial values for states $(x_0 \mid 0)$, the initial state variance-covariance matrix $(P_0 \mid 0)$, and initial values for all unknown parameters ($\theta 0$), the simple Kalman filter:

EViews (Quantitative Micro Software 2007a, 2007b, 2007c) is a statistical software package for data analysis, regression and forecasting (Van den Bossche 2011). For state space models, an object space should be created which provides estimation of for single or multiple dynamic equations in state space form by applying the built-in Kalman filter algorithm. The model estimation is carried out by maximum likelihood estimation (Durbin and Koopman 2002). However, the starting values of the parameter are specified in the coefficient vector. The optimization of the log- likelihood can be done by Marquardt–an improvement of Gauss Newton algorithm and Berndt-Hall-Hall-Hausman which is a modification of Newton Rapson algorithm.

4. Result and Discussion

The present study is carried out to analyze the impact of four different factors on determining the price of gold in India. The four determinants under consideration are US dollar to Indian rupee exchange rate, NIFTY index, oil price and demand of gold in India to develop a model for price of gold in India. Monthly data of all these variables are collected from various official sources and used for the estimation. The time period under study is between 1st, April 2012 to March 31st, 2019 with 82 observations each for the five factors. The statistics of the four independent variables–closing price of NIFTY, price of crude oil, US dollar exchange rate and demand of gold in India and the dependent variable price of gold in India is provided in Table 1.

Table 1 Descriptive statistics of the variables-NIFTY, crude oil price, US dollar exchange rate, demand of gold and price of gold

	NIFTY	Crude Oil Price (Rs/barrel)	Demand of Gold (Tones)	Exchange rate-US Dollar (Rs)	Price of gold (Rs/Gram)
Mean	6730.76	5384.09	1108.04	57.217	2487.5
Median	6134.5	1497.94	1090.73	59.79	2519.87
Maximum	9304.05	7323.42	1293.06	68.23	3056.64
Minimum	4624.3	2161.06	919.52	44.030	1629.15
Std Dev	1389.21	23.303	90.130	8.1599	332.53
CV	20.6	29.792	8.134	14.26	13.36
Skewness	0.28	–0.486	0.31	–0.34	–0.72
Kurtosis	1.5	–1.29	2.494	1.6931	2.9794
Jarque-Bera	8.9	9.097	2.313	7.690	7.524
Probability	0.0115	0.0105	0.31	0.02	0.0232

Each variable is expressed in different units. None of the variables have very high variation, they seem to be of almost same consistency. All the variables except demand of gold exhibit negative skewness also which is an indication, that during majority of the study period the values were above average. The variables have a kurtosis coefficient less 3 indicating a flatter frequency distribution. The comparatively price and demand of gold seems to be close to normal curve. The test of normality by

Jarque-Bera test (Jarque and Bera 1987) infers that none of series follow normal distribution except for the variable demand whose probability value is 0.3145 > 0.05.

Boxplots are very efficient visualization tools that help us locate the average and comprehend the dispersion of the data. The variations in each variable is analyzed by boxplot of these variables represented in Figure 1.

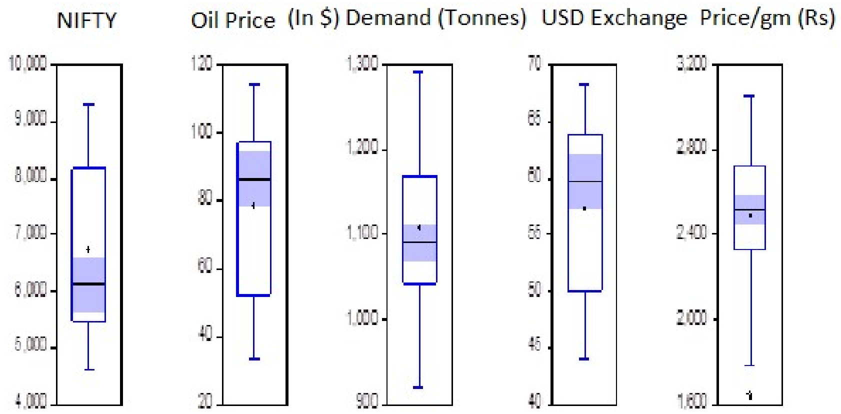

Figure 1 Boxplot of NIFTY price, price of crude oil, US dollar exchange rate, demand of gold and price of gold

Maximum variability is found in NIFTY values (Range-5000) and minimum variability was found in Exchange rate of dollar (Range-25). Each variable has a different unit. The other three variables, Price of gold, demand of gold and oil price exhibit more uniformity. But the assumption of normality is does not exist in none of the variables except demand of gold which is evident with Jarque-Bera test as well as symmetry of the box plots. Line graph of values of NIFTY price, price of crude oil, US dollar exchange rate, demand of gold and price of gold for comparison of the series is presented in Figure 2 to observe the trend and patterns of each series.

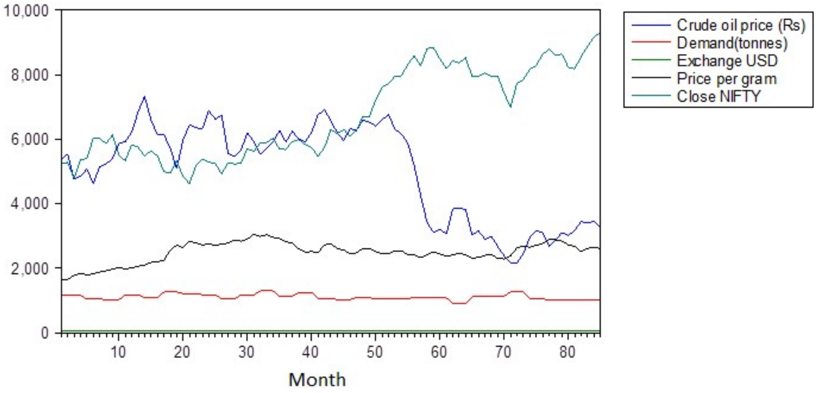

Figure 2 Line graph of values of NIFTY price, price of crude oil, US dollar exchange rate, demand of gold and price of gold for comparison of the series

Price of crude oil and closing price of NIFTY represent more fluctuation compared to other series. The variations are compared with differenced data of the five series in Figure 3 to decide on stationarity of the series.

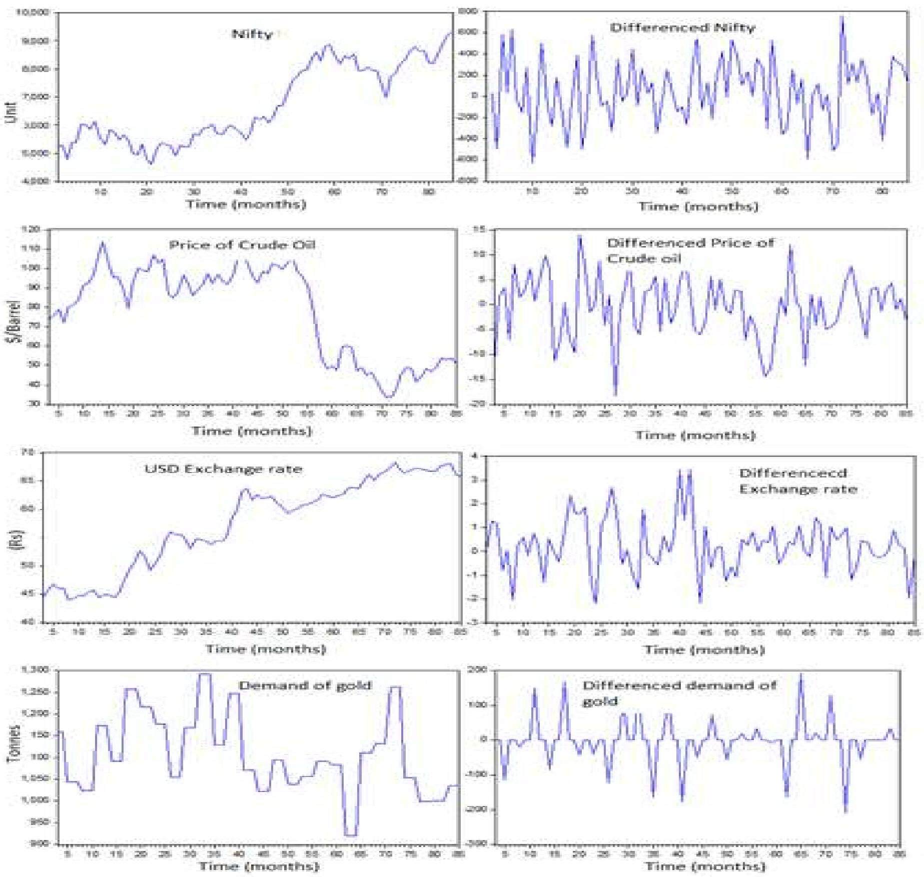

Figure 3 Line graph of raw data and differenced data of the monthly values of NIFTY, price of crude oil, US dollar exchange rate and demand of gold

4.1. Test of stationarity

A series should be stationary for using it for analysis, i.e, mean and variance and covariance are time invariant. To test if there is stationarity (trend stationary or not), Kwaitkowski-Phillips-Schmidt -Shin (KPSS) test is carried out. This test has a null hypothesis that the series under study is stationary around the mean/trend stationary.

Table 2 KPSS test for Stationarity of the variables NIFTY price, price of crude oil, US dollar exchange rate, demand of gold and price of gold

Series	Test statistic	1% critical value	5% critical value	Remark
Price of Gold	0.37707	0.7390	0.4630	Stationary at level
Demand of Gold	0.4243			Stationary at level
Oil Price	0.7273			Stationary at level
NIFTY	1.0285			Non-stationary at level
D (NIFTY)	0.0904			Stationary at first difference
USD Exchange Rate	1.1089			Stationary at level
D (USD Exchange Rate)	0.1163			Stationary at first difference

The result of KPSS test is presented in Table 2. KPSS test is preferred over unit root tests by augmented Dickey-Fuller test and Phillips-Perron test whenever there is a contradictory result provided by both the test (Katircioglu et al. 2014; Jafari et al. 2012; Kwiatkowski et al. 1992).

It can be observed that while price and demand of gold and oil price are stationary at level, USD exchange rate and NIFTY are not stationarity at level. But the non-stationary series are attaining stationarity at first difference. Hence, the two series US dollar exchange rate and NIFTY are stationary at first difference while price and demand of gold are stationary at level. Figure 3 presents a comparison of the raw data and difference series to visualize the series attaining stationarity.

4.2. Multiple linear regression (MLR) model using principle of ordinary least squares

Regression analysis is possible if the predictor variables are significantly correlated with the dependent variable. This can be checked by the simple bivariate correlation coefficient due to Karl Pearson. At the same time, there should not be multicollinearity present in the predictor variables. The multiple linear regression (MLR) model is also based on various other assumptions too. To find if there is any correlation between the dependent variable and the independent variables, bivariate correlation analysis is carried out. Initially scatterplot of the variables is viewed in Figure 4.

The matrix scatterplot indicates moderate correlation between price and demand of gold and also price of gold and exchange rate of USD. The other two variables do not exhibit any pattern of correlation. To confirm this Karl Pearson's correlation coefficient is calculated and the values are presented in Table 3.

Table 3 Correlation coefficient values of the variables NIFTY price, price of crude oil, US dollar exchange rate, demand of gold and price of gold

	Oil price Rs/barrel	NIFTY	Demand (tones)	Exchange USD (Rs)
Price of Gold (Rs per gram)	–0.007	0.153	0.125	0.480**

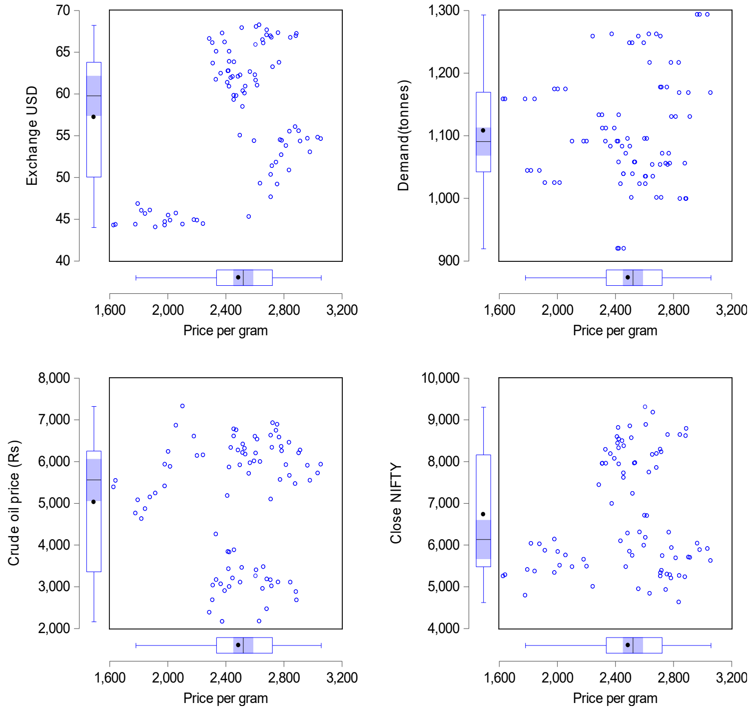

Figure 4 Scatterplot to visualize the correlation between NIFTY price, price of crude oil, US dollar exchange rate and demand of gold with price of gold

All four correlation values provided in Table 3 are low but positive, price of gold and exchange rate of US dollar is the only pair which has significant correlation. However, the multiple linear regression model is estimated by the principle of least squares with price of gold as dependent variables and the other four variables as predictors. The model summary is depicted in the following Table 4.

Table 4 MLR model for price of gold

Variable	Coefficient	Standard Error	t-statistic	Probability
C	–943.62	644.81	–1.46	0.1500
Demand	0.9646	0.3628	2.6585	0.0095
Crude Oil Price	5.3851	1.7564	3.066	0.0030
NIFTY	–0.0805	0.0448	–1.7968	0.0763
US Dollar Exchange Rate	43.31	5.747	7.5355	0.0000
Performance indicators of the model				
R-Squared		0.4945	AIC	13.908
Adjusted R-squared		0.4682	SIC	14.055
Log Likelihood		–565.2508	HQN	13.967
F-Statistic		18.8313	Durbin-Watson Statistic	0.2324
Significance Value		0.000		

The regression model hypothesized is

$$\text{Price of Gold} = B_0 + B_1 \times \text{Demand} + B_2 \times \text{Price of Crude Oil} \\ + B_3 \times NIFTY + B_4 \times \text{US Dollar Exchange Rate.} \tag{2}$$

The estimated values as given in Table 4 is

$$\text{Price of Gold} = 943.62 + 0.9646 \times \text{Demand} + 5.3851 \times \text{Price of Crude oil} \\ - 0.0805 \times NIFTY + 43.31 \times \text{US Dollar Exchange Rate.} \tag{3}$$

As suggested by correlation value, US dollar exchange rate is having the highest influence in predicting the price of the gold. None of the regression coefficients except NIFTY is insignificant according to the result of the t-test carried out shown in Table 4. The R squared value is quite low, 49.45% only is a matter of concern. Log likelihood value of –565.25. The ANOVA test with null hypothesis that the model is not suitable for acceptance provides a statistic value F =18.83 with p-value 0.00 suggesting accepting the model. The information criteria values are all close to each other around 13. The Durbin-Watson test value is 0.2324 indicate the presence of autocorrelation of the residuals. The value close to 2 is an indication of absence of autocorrelation. Values close to zero implies presence of negative autocorrelation.

The assumptions of validating the MLR models are to be verified before accepting the model. The diagnosis of Multicollinearity of the variables is found by analyzing the values of variance inflation factor (VIF), the values of VIF are given below, they indicate no significant multicollinearity between the variables as all four of the VIF values given in Table 5 are less than 5 (Mansfield and Helms 1982).

Table 5 VIF value of the predictors of MLR model

Variable	Demand	Crude Oil Price	NIFTY	Dollar Exchange Rate
VIF	1.48	2.316	4.98	3.011

Multicollinearity is suspected if the VIF values exceed 5. Here, all four of the independent variables have multicollinearity below 5. Hence, the presence of significant multicollinearity is ruled out. Residual analysis is another important aspect. The residuals are supposed to have normal distribution with mean zero and constant variance in the case of MLR. To check this the graph of the actual, fitted and residual values of the variable is given in Figure 5.

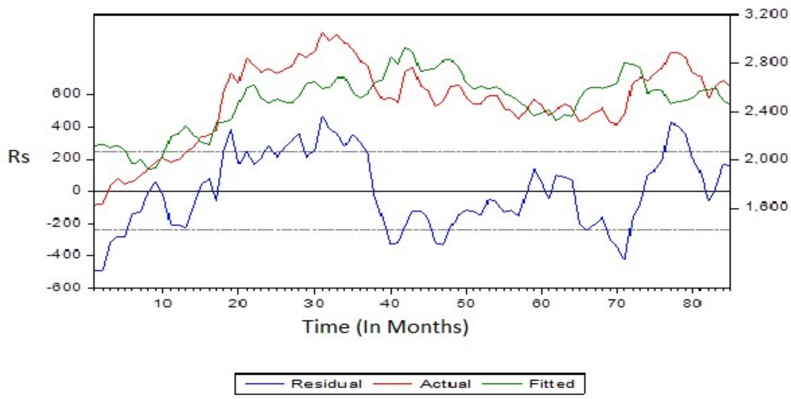

Figure 5 Line graph of predicted values and residuals along with actual values of the MLR model developed by OLS method

From Figure 5, it can be observed that the residuals show large variations. Also, there is wide gap between the actual and predicted lines. The efficiency or prediction power of the model explained by R-squared. This value indicates the percent variation in the dependent variable that can be explained by the model. A good model must explain at least 60% of variations in the dependent variable. The model estimated in Table 6 is only 0.491, with adjusted R-squared 0.401 implying that the model explains only 49% variation of price of gold. The standard error of the estimate is also very high. However, ANOVA test (Table 6) is accepting R-squared as significant which satisfies the model as acceptable.

Table 6 Model diagnostic for the MLR model estimated

R	R-squared	Adjusted R-squared	Standard Error	F Statistic	p-value
0.701	0.491	0.466	234.014	19.321	0.000

The performance indicators of the model are root mean square (RMSE), mean absolute percent error (MAPE) and mean square error (MSE) values. These values should be as low as possible for a model to be accurate. For the present model RMSE of the model 238.4987, MAPE is 8.379 and MAE is 205.024. These values along with the graph of residuals with the confidence interval is given in Figure 6.

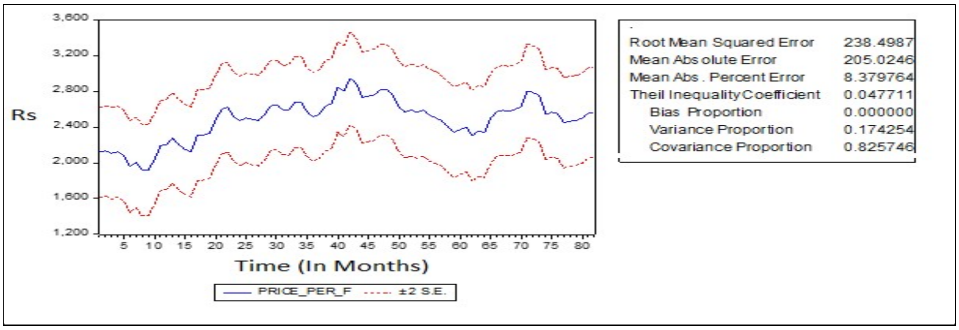

Figure 6 Line graph of the forecast values and 95**%** confidence band of the MLR model developed by OLS method

Though the RMSE, MAE and MAPE values are high, the mean bias and variance bias are low. These values indicate the difference between the actual and fitted mean and variance, respectively.

The residuals of a model fit are assumed to follow normal distribution with mean 0 and constant variance. The following Figure 7 gives the histogram of the residuals along with Jarque-Bera test result. The hypothesis of normal distribution of residuals is accepted (3.16, 0.205) but the standard deviation of the residuals is very large (SD = 237.15).

The various R-squared value, heteroscedasticity value and the residual diagnosis of the model do not favor the model much. Hence an attempt to estimate a better model is considered. R-squared value is just 0.49 which indicates the percent variation captured by the model is a very low percent of 49 only. The MSE and MAPE values also should be reduced to improve prediction accuracy.

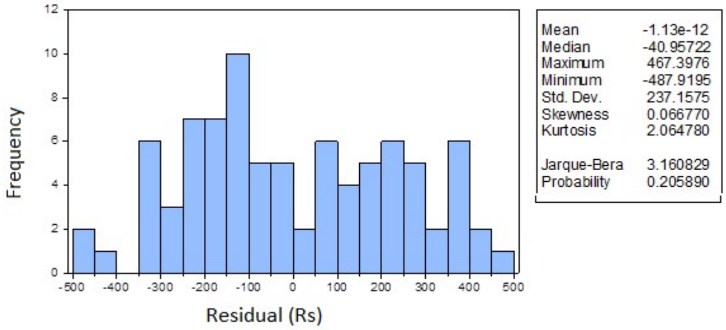

Figure 7 Histogram and normality check of residuals of the MLR model developed by OLS method

4.3. Need for time-varying coefficient model

Varying coefficient models are not mere mathematical extensions, rather they are derived out of the need in various situations. They capture the dynamics features that may exist in the data, which mostly are ignored by the parametric models. They are better in the sense, the parametric models are not discarded, and rather the coefficients are made to evolve with certain characteristics. They remain to be locally parametric models. Price of gold is determined in India mainly by 4 determinants-US dollar to Indian rupee exchange rate, NIFTY index, oil price and demand of gold in India. When each variable separately analyzed, they exhibit serial correlation or autocorrelation which is presented in the result and discussion of the same chapter. Moreover, all four series are dynamic and they are not static with time. These factors are time dependent; hence it would be appropriate to capture the properties of these variables too to incorporate into the prediction model of price of gold.

The four factors used as predictors are very dynamic variables with inherent autoregressive property. Most appropriate ARIMA model for each variable is determined to comprehend if the variables are time dependent.

The parameters are finalized based on the correlogram and also those which gives the least AIC values, the estimation of the models is carried out by Box-Jenkins method. The values of the coefficients of the relevant ARIMA models is presented in the following Tables 7 and 8.

Table 7 Identifying the parameters of AR/MA model for the variables NIFTY price, Price of crude oil, US Dollar Exchange rate, Demand of gold and Price of Gold

Parameter	Price of Gold (Rs/gms)	US Dollar Exchange Rate	Crude oil price (Rs)	NIFTY	Demand of gold (Tones)
P	1	1	1	1	1
D	0	1	0	1	0
Q	0	0	0	0	0

4.4. Estimation of time varying coefficient regression model for price of gold

From the results in Tables 7 and 8, the various ARIMA models for the variables under study are as follows-oil price and price and demand of gold are AR(1) while USD exchange rate and NIFTY index are ARIMA(1,1,0) Hence, it is imperative for the study to model the price of gold by considering the autoregressive nature of these of the variables, US dollar exchange rate, NIFTY, crude oil price and demand of gold.

Table 8 Estimation of the suitable ARIMA models for the variables NIFTY price, Price of Crude oil, US dollar exchange rate, demand of gold and price of gold

Variable	Price of Gold (rRs/Gms) ARIMA(1,0,0)	NIFTY Closing Price ARIMA(1,1,0)	Crude Oil Price (Rs) ARIMA(1,0,0)	US Dollar Exchange Rate ARIMA(1,1,0)	Demand of Gold (Tones) ARIMA(1,0,0)
C	2637.34	48.42	4119.19	0.2588	1102.81
AR(1)	0.9212	–0.069	0.9727	0.1998	0.75992
AIC	11.8633	14.335	14.8354	3.0463	11.0312

A time-varying regression model is developed where the regression coefficients are time dependent or autoregressive in nature. NIFTY is not included in the model as it was not showing a significant correlation and regression coefficient, and excluding NIFTY brought the values of AIC lower in the present model.

The hypothesized model has the signal equation and state equations are as follows:

$$\begin{aligned}@\text{Signal Price of Gold} &= sv1\times\text{Demand} + sv2\times\text{US Dollar Exchange Rate}\\ &\quad + sv3\times\text{Price of Oil} + sv4 + c(1)\times sv5\end{aligned} \tag{4}$$

$$\begin{aligned}&@\text{state } sv1(\text{Demand}) = c(3) + [\text{var} = \exp(c(2))]\\ &@\text{state } sv2(\text{US Dollar Excahnge Rate}) = c(4) + c(5) * sv(-1) + [\text{var} = \exp(c(2))]\\ &@\text{state } sv3(\text{Price of Oil}) = c(6) + c(7) * sv3(-1) + [\text{var} = \exp(c(2))]\\ &@\text{state } sv4 = c(8) * sv4(-1) + [\text{var} = \exp(c(2))]\\ &@\text{state } sv5 = sv4(-1).\end{aligned} \tag{5}$$

A constant structural covariance is assumed in the estimation. Maximum likelihood estimation procedure with Marquandt method is used for the estimation. Kalman filter estimates parameters via maximum likelihood method. It gives an initial value for states and variances-covariance matrix and for all unknown parameters also. The model developed is given in the following Table 9.

Table 9 Estimation of time varying coefficient model for price of gold in India with predictor variables price of crude oil, US dollar exchange rate and demand of gold

	Coefficient	Std Error	Z-statistic	Probability
C(1)	2.084743	330940.2	6.30E-06	0.0980
C(2)	–7.648572	0.194293	–39.36622	0.0000
C(3)	0.146889	0.204239	0.719200	0.4720
C(4)	8.584976	9600.907	0.000894	0.3993
C(5)	0.752929	276.2983	0.002725	0.0978
C(6)	0.004148	0.004248	0.976441	0.0288
C(7)	0.920139	0.059724	15.40664	0.0000
C(8)	0.969928	30335.61	3.20E-05	0.0900

Table 9 (Continued**)**

	Final State	RMSE	Z-Statistic	Probability
SV1	0.146889	0.021834	6.727512	0.0000
SV2	34.74702	0.033175	1047.374	0.0000
SV3	0.051561	0.022680	2.273407	0.0230
SV4	1.50E-05	0.089707	0.000167	0.1999
SV5	1.55E-05	0.089707	0.000172	0.1999
Performance Indicators of Estimators				
Log likelihood	−521.1736	Akaike info criterion		12.45114
Hannan-Quinn criterion	12.54361	Schwarz criterion		12.68104

The model can be represented as follows:
The signal equation is

$$\begin{aligned}@\text{Signal Price of Gold} &= 0.1468\times\text{Demand}+34.74447\times\text{US Dollar Exchange Rate}\\ &\quad +0.0515\times\text{Price of Oil}+0.00001+2.0847\times0.0000155.\end{aligned} \tag{6}$$

The state equations are

$$\begin{aligned}&@\text{Signal Price of Gold} = 0.1468\times\text{Demand}+34.74447\times\text{US Dollar Exchange Rate}\\ &\qquad +0.0515\times\text{Price of Oil}+0.00001+2.0847\times0.0000155\\ &@\text{state } sv1(\text{Demand}) = 0.1468+\exp(-7.6485)\\ &@\text{state } sv2(\text{US Dollar Excahnge Rate}) = 8.5849+0.7529\times sv2(-1)+\exp(-7.6485)\\ &@\text{state } sv3(\text{Price of Oil}) = 0.004148+0.9201\times sv3(-1)+[\exp(-7.6485)\\ &@\text{state } sv4 = 0.9699\times sv4(-1)+\exp(-7.6485)\\ &@\text{state } sv5 = sv4(-1)+\exp(-7.6485).\end{aligned} \tag{7}$$

The model gives the relation between the coefficients of the predictors to their past values due to auto-regression. A constant variance and covariance is imposed in the model while estimation. The initial values to the unknown parameters are assigned and the estimation is carried out by Maximum likelihood method. The estimated model indicates the huge impact of US dollar exchange price on gold. No other variable seems to have such huge influence.

The indicators of performance of the model developed are log likelihood, AIC, SIC and HQN. This model seems to outperform the one estimated by OLS method. This inference is made by comparing the information criterion values and the log likelihood values of the model estimated. Moreover, the coefficients are found to be significant in all the variables in both state and signal equations. The value of log likelihood is −521.3724 which less than that of OLS method is. The values of all three information criteria are also close to 12 which is less than 13, the corresponding values in MLR.

The graphs displayed in Figure 8 is the one-step ahead forecast values of the signal state equation. The confidence interval is also provided which is very close to the predicted line indicating low variation.

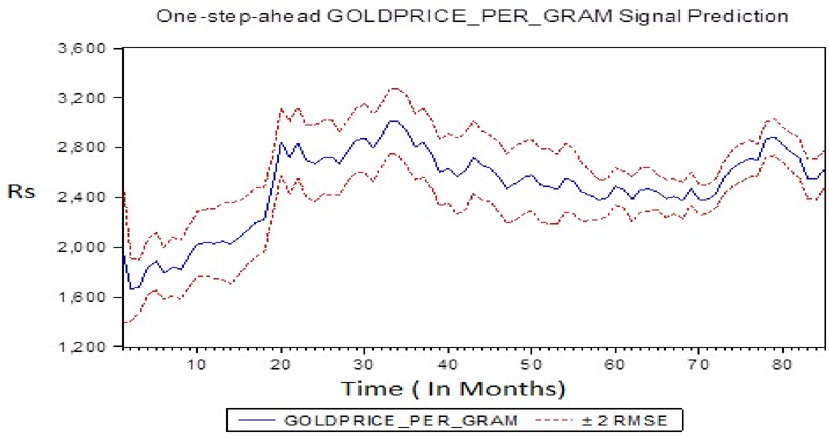

Figure 8 One-step ahead forecast values of the signal state equation

The prediction of gold price in the signal equation presented in Figure 8 shows high accuracy in the estimation. The 2SE values are very close to the line of prediction indicating less SE. this can be confirmed by considering the State equations too. Figure 9 gives the scatter plot of the residuals of the estimated model given in Table 8.

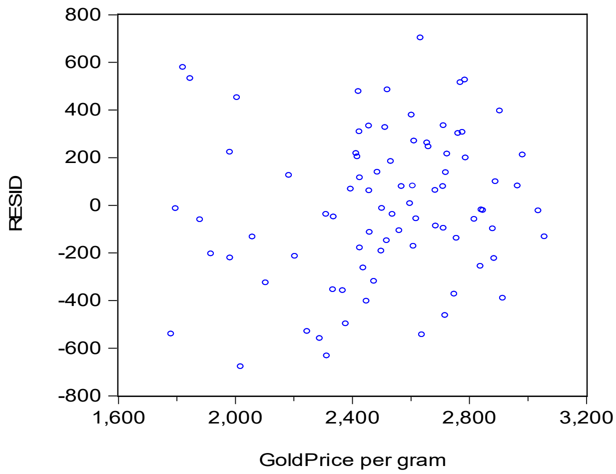

Figure 9 Scatter plot of residuals exhibit random pattern

The estimation of parameters of the model is resented gives the coefficients of the variables their autoregressive values too along with the indicators of AIC as 12.45 which is lower than the linear regression model. The coefficients are all significant according the t-test except for few. The forecast graph using the estimated model is displayed in Figure 10.

The forecast graph shows that the gap between actual and predicted values of the price of gold has reduced significantly. This itself proves that the time-varying coefficient model outperforms OLS method. This must be because the method has the ability to capture the dynamic nature of the time series of independent variables.

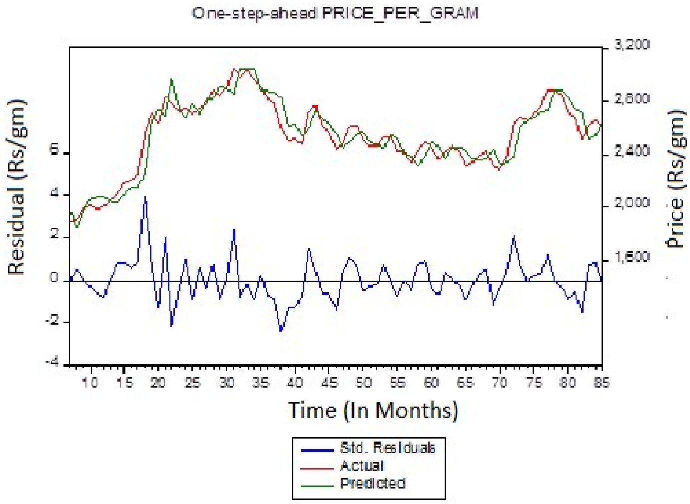

Figure 10 Line graph of the actual observations and predicted values and residuals of price of gold

The residuals in the figure are very close to zero. The actual and estimated values also are very close to each other. Apart from the closeness of actual and predicted values of the dependent variable, the residuals also exhibit interesting properties. The residuals also have a random pattern when plotted along with the dependent variable. This is an indication that the model fits in well on the data. To confirm this the graph of the actual observations and predicted values is considered in the line graph and their correlation is also calculated. It is found to be 0.7302 (t = 4.85, p-value = 0.0002). The varying coefficient model developed seems to outperform the regular regression model. This can be attributed to the dynamic nature of the predictor variables.

5. Conclusions

This study attempts to estimate and compare the linear regression model and time-varying regression model to predict price of gold. Unlike linear regression models, the time varying coefficient models not only provides the signal model but also the impact of time on each variable by the state coefficients. It is also observed that the latter method is more effective as the all the three information criteria values (AIC, SIC and HQN) values are lesser. The log likelihood values which are supposed to be small for a better model when two or more models are compared is also low in time-varying coefficient models. The residual analysis also supports the inferences. Thus, it can be concluded that dynamic models have better prediction accuracy than the routine linear models. The model also highlights the role of past values of the predictor variables in forecasting the values of the dependent variables. It is an indication how to reduce the gap between the actual and predicted values by incorporating the dynamic and autoregressive characteristics of the predictor variables.

References

Assaf S, Campostrini S. Application of the varying coefficient model to the behaviour risk factor surveillance data in Italy: A study of changing smoking prevalence among sub-populations. BMC Pub Health. 2015; 15(1): 1-12.

Beckmann J, Czudaj R. Gold as an inflation hedge in a time-varying coefficient framework. North Am J Econ Finance. 2013; 24: 208-222.

Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. J Am Stat Assoc. 2000; 95(451): 888-902.

Cai Z. Two-step likelihood estimation procedure for varying-coefficient models. J Multivar Anal. 2002; 82(1): 189-209.

Durbin J, Koopman SJ. A simple and efficient simulation smoother for state space time series analysis. Biometrika. 2002; 89(3): 603-616.

Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli. 2005; 11(6): 1031-1057.

Fan J, Zhang W. Statistical estimation in varying coefficient models. Ann Stat. 1999; 27(5): 1491-1518.

Fan J, Zhang W. Statistical methods with varying coefficient models. Stat Interface. 2008; 1(1): 179.

Gupta E. Global warming and electricity demand in the rapidly growing city of Delhi: A semi-parametric variable coefficient approach. Energy Econ. 2012; 34(5): 1407-1421.

Hastie T, Tibshirani R. Varying coefficient models. R Stat Soc Series B Stat Methodol. 1993; 55(4): 757-779.

Huang JZ, Wu CO, Zhou L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Stat Sin. 2004; 14: 763-788.

Huang JZ, Wu CO, Zhou L. Varying coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika. 2002; 89(1): 111-128.

Jafari Y, Othman J, Nor AH. Energy consumption, economic growth and environmental pollutants in Indonesia. J Policy Model. 2012; 34(6): 879-889.

Katircioglu ST, Feridun M, Kilinc C. Estimating tourism-induced energy consumption and CO_2 emissions: The case of Cyprus. Renewable Sustainable Energy Rev. 2014; 29: 634-640.

Kwiatkowski D, Phillips PC, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. J Econom. 1992; 54(1-3): 159-178.

Manoj J, Suresh KK. Forecast model for price of gold: multiple linear regression with principal component analysis. Thail Stat. 2019; 17(1): 125-132.

Mansfield ER, Helms BP. Detecting multicollinearity. Amer Stat. 1982; 36(3a):158-60.

Qu A, Li R. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics. 2006; 62(2): 379-391.

Street L, Gopaul K, Hewitt A, Grubb M. Gold demand trends full year 2014. World Gold Council: Technical Report 2015. Available from https://www.gold.org/goldhub/research/gold-demand-trends/gold-demand-trends-full-year-2014.

Tang Y, Wang HJ, Zhu Z, Song X. A unified variable selection approach for varying coefficient models. Stat Sin. 2012; 22: 601-628.

Van den Bossche FA. Fitting state space models with EViews. J Stat Softw. 2011; 41(8): 1-6.

Wang HJ, Zhu Z, Zhou J. Quantile regression in partially linear varying coefficient models. Ann Stat. 2009; 37(6B): 3841-3866.