Asian Journal of Economic Modelling
ISSN(e): 2312-3656
ISSN(p): 2313-2884
DOI: 10.18488/journal.8.2019.73.110.120
Vol. 7, No. 3, 110-120

URL: www.aessweb.com

DETERMINANTS OF PETROL PRICES IN INDIA: A REGRESSION MODEL WITH DE-AUTOCORRELATED TIME-SERIES DATA

Betcy Theodore[1+]
Martin L William[2]

[1]*Graduate Student, Department of Statistics, Loyola College, Chennai, India.*
Email: betcy.theodore@gmail.com Tel: +91 9445457954
[2]*Associate Professor, Department of Statistics, Loyola College, Chennai, India.*
Email: martin_lw@yahoo.com Tel: +91 9884094833

(+ Corresponding author)

Article History
Received: 24 January 2019
Revised: 5 March 2019
Accepted: 10 April 2019
Published: 1 July 2019

Keywords
Box-cox transformation
De-autocorrelation
Partial auto-correlation function
Partial correlation
Multicollinearity
Normality of errors.

JEL Classification:
C30, C31, C32.

ABSTRACT

The control that the Central Government of India had for many years over the petrol prices, kept the life of the common man, who was dependent on the prices of essential commodities for day to day existence, at ease. The liberalization policy allowing determination of petrol prices by the oil marketing companies based on the international market prices is now affecting the prices of all essential commodities with cascading effects. This paper attempts to discover how strongly various manifest factors govern the petrol prices in India on a month-to-month basis. A regression model is built using time-series data on monthly petrol prices and other lagged variables including petrol consumption, import and export of petroleum products, etc. spread over the period 2010 to 2018. Various transformations are performed on the variables to get a more accurate model. The auto-correlation present in the data is suitably handled and finally a regression model is built with the de-autocorrelated data.

Contribution/ Originality: This paper's primary contribution is the identification of the main factors, among those considered, that affect petrol prices in India. This objective is achieved using a regression model with monthly time series data wherein the autocorrelation is removed for more reliable results. The path that is followed to get improvements to the model by making transformations on the dependent and independent variables, multicollinearity redressal and de-autocorrelation is an interesting and unconventional feature of this paper.

1. INTRODUCTION

When Colonel Edwin Drake drilled the first successful oil well in Titusville, Pennsylvania in 1859, few had any idea of how petroleum would change the world. But the prime position of oil in the market started to disappear when Thomas Edison invented the light bulb and brought in the era of electricity in illuminating the world. This provided alternate energy for lighting, thus displacing oil from its prime position of providing energy. However, this displacement was temporary and the invention of the automobile brought oil back into focus. Subsequently, mobility gained over light. The vehicle population grew in multitudes rapidly in the subsequent years and decades. [Refer to https://enviroliteracy.org/energy/fossil-fuels/petroleum-history/]. The prime position of oil remains ever since and it has become an essential part of the economy of nations and a necessary component of civilized life.

In today's world, there are several forces affecting everyday life of people. Some may have direct impact and some, indirect. They may affect the common man in such a way that he blames something completely different. One of the best examples in India, that often makes news, is the alarming rise in petrol prices. Swadimath *et al.* (2013) discuss the rise and impact of crude oil price in India and point out how the OPEC nations work as a cartel and exert a monopolistic approach in controlling crude oil prices.

Petroleum is an indispensable and integral part of varied areas of human activity including transportation, manufacturing, power-generation, etc., which directly or indirectly affects agriculture, business, economics, etc. Rise in petrol prices increases the cost of transportation which in turn affects the price of goods that are being transported. This affects the everyday life in the form of price increase in vegetables, food items and other basic necessities that are required to be transported on a regular basis. Even the common transport options like buses and taxis become dearer with hiked fares which affects people's daily routine. The entire economy faces a chain effect. The impact of oil price volatility on select economic indicators in India has been studied by Srithar *et al.* (2015).

The common man tends to blame the government for the rise of fuel prices and increase in taxes and judges the government as money-minded as he is much concerned about the effect on his life. But there are other factors affecting petrol prices, some of them being availability, consumption, mismatch of supply and demand, import/export costs, the country's balance of payments, etc., a few of which are considered for the present research. However, it is not only the common man but also the educated people who remain in the dark on the most important factors that determine oil prices. This research is an attempt to reveal the factors behind the changes in petroleum price and tries to find answers to the questions on the determinants of the same. An interesting work in this context is a paper by Gyagri *et al.* (2017) who presented a theoretical review on the 'Determinants of Global Pricing of Crude Oil'.

The specific objective of this research is to understand the effect of various factors on petrol price through a suitable regression model and to identify the level(s), if any, of the factor(s) at which petrol prices can be controlled at minimum levels. As the data to be used for the study is a time series data available on a monthly basis, it is likely to possess auto-correlation, which needs to be detected and corrected, if present. The regression model to be built needs to be a 'good' one in terms of inclusion of relevant variables leading to proper interpretation on the significant determinants of petrol prices.

2. DATA AND METHODOLOGY

The data required for this study are obtained from the official website of petroleum planning and analysis cell of the government of India and websites of Daily News Analysis and Autocar Professional. The dataset contains one response variable and six regressor variables. Eighty-three data points of monthly data from May, 2011 to March, 2018 are available for these variables.

India is a developing nation with a huge population of around 1.34 billion people currently. In such a country, fuel is of utmost importance for large factories and for the individuals as well. And so, the demand for petrol is always high. But the ability to produce petrol in India is limited due to the scarce availability of oil fields. Therefore, to meet the demand, petroleum products are imported from oil producing countries. The cost of petroleum being imported depends on dollar-rupee exchange rates.

Considering all the factors mentioned above and the availability constraints, the regressor variables considered for this study are:

X_1 – Consumption of petrol ('000 Metric Tonnes)

X_2 – Production of petroleum products in India ('000 Metric Tonnes)

X_3 – Production of petroleum products in OPEC ('000 Metric Tonnes)

X_4 – Imports of petroleum products to India ('000 Metric Tonnes)

X_5 – Exports of petroleum products from India ('000 Metric Tonnes)

X_6 – Dollar-Rupee exchange rate (in INR)

and the response variable is:

Y – price of petrol per litre (in INR)

It is only logical to expect that the current month's petrol price depends on the previous month's consumption, production, imports and exports. Thus, the regressor variables from X_1 to X_5 are taken with a lag of one month, i.e., the current month's petrol price is regressed on previous month's consumption, production, imports and exports. However, the exchange rate of Indian Rupee versus the US Dollar has a 'concurrent' relationship with the daily prices of petroleum products and hence, X_6 is considered concurrently with the dependent variable Y. In this context, we refer to Hidhayathulla and Rafee (2014) for a study on the relationship in a different direction between crude oil price and Rupee-Dollar exchange rate.

Table 1 gives a sample of the dataset used in the present study:

Table-1. Sample dataset.

YEAR	MONTH	Y	X_1	X_2	X_3	X_4	X_5	X_6
2011	May	68.33	12595	17011	31754.12	17438	4598	44.86
2011	June	68.62	12880	17232	31867.76	14979	5757	44.86
2011	July	68.62	12129	16957	32518.31	15555	5065	44.42
2011	August	68.62	11762	16926	32762.95	15159	5229	45.31
2011	September	71.91	11185	17189	32868.91	15588	5257	47.43
2011	October	73.81	11275	15644	32895.26	14259	5336	49.17
2011	November	71.47	12038	16188	32630.96	13536	5025	50.6

Source: Websites of Petroleum Planning and Analysis Cell (Govt. of India), Daily News Analysis and Autocar Professional.

The analysis is carried out using SAS and R software. Model building is carried out by forward, backward and step-wise algorithms and the best among these is considered. Transformation of the response variable towards improving the model is attempted using the Box-Cox approach and compared with the initial (base) model. Partial correlation is assessed to include transformed regressors in the model. Normality of errors is checked using Shapiro Wilk test. Presence of auto-correlation is diagnosed using Breushe Godfrey test and PACF (Partial Auto-Correlation Function). Multi-collinearity is diagnosed by making use of Variance Inflation Factors (VIF) and Variance Decomposition Proportions. Following these diagnostics and appropriate 'tunings', an ideal model is finally built.

3. MODEL BUILDING

In this section, we present the sequence of steps we have followed towards building an efficient model. We aim at getting a model with an adequately high value of R^2 so as to obtain reliable inferences about the factors affecting oil prices in India.

3.1. Initial Model with the Original Regressors

3.1.1. With Original Dependent Variable

The initial model that is built with the specified regressors gives the output summarized in Table 2:

Table-2. Model with original dependent variable and original regressors.

Regressors	Estimate	Std. error	t value		Pr(> | t |)
Intercept	99.1460372	17.6449975	5.616		$2.74e^{-07}$
X_2	0.0009274	0.0004896	1.894		0.0618
X_3	-0.0012691	0.0006750	-1.880		0.0637
Multiple R^2	0.04964	Adjusted R^2	0.02589	p value	0.1305

Source: Results of analyses carried out in R package.

It is seen that, the variables significantly affecting petrol price are production of petroleum products in India(X_2) and production of petroleum products in OPEC(X_3). The coefficient of determination, R^2 is only 0.04964 and the adjusted R^2 is 0.02589, which is too low to arrive at reliable conclusions.

It is well known that the Shapiro-Wilk test is a powerful omnibus test for normality against all types of alternative distributions and sample sizes. The power comparison of Shapiro-Wilk procedure with other approaches has been carried out over the past many years by several authors who have pointed out the merit of Shapiro-Wilk test over other tests. A recent work that we refer to in this context is a paper by Razali and Wah (2011).

We carry out the normality test for the residuals of the initial model using Shapiro-Wilk test which reveals the non-normality of the residuals. Thus, the initial model is unsatisfactory with a low R^2 value and non-normal errors. We therefore proceed to do modifications to the model.

3.1.2. Model with Box-Cox Transformation on the Dependent Variable

Model-building using Ordinary Least-Squares Method of estimation of regression coefficients require constancy of variance across the records and tests of significance require normality of the errors. Significant deviations from these assumptions result in undesirable Type I or Type II errors. One of the best ways to overcome these violations is to transform the data. Box Cox transformations represents a family of power transformations for variance-stabilization and normality. We refer to Box and Cox (1964) for the original ideas. The important role of Box-Cox approach in modelling has been studied by many authors over the past many years. For a primer on Box-Cox transformation and Monte Carlo investigation for small samples, we refer to Spitzer (1978;1982) and for its use in limited dependent variable models, we refer to Poirier (1978). Recently, Nwakuya and Nwabueze (2018) discussed the application of Box-Cox transformation as a corrective measure to heteroscedasticity for an economic data.

We subject the dependent variable to Box-Cox transformation and the optimum value of the power (λ) that minimizes the mean-square residual is found to be 1.59. A model is built using the transformed dependent variable and all independent variables used in the previous model. The model summary in Table 3, shows that the variables that significantly affect the petrol price are production of petroleum products in India(X_2) and production of petroleum products in OPEC(X_3).

Table-2. Model with Box-Cox transformed dependent variable and original regressors.

Regressors	Estimate	Std. error		t value	Pr(> | t |)
Intercept	1522.85056	371.85491		4.095	0.0001
X_2	0.02018	0.01031		1.957	0.0539
X_3	-0.02745	0.01422		-1.931	0.0570
Multiple R^2	0.05251	**Adjusted R^2**	0.02883	**p value**	0.1156

Source: Results of analyses carried out in R package.

The coefficient of determination, R^2 is 0.0521 and the adjusted R^2 is 0.02883 showing a marginal improvement over the initial model. But still, the performance is quite low and hence, we proceed to transform the independent variables and include those transformed variables as inputs to the model.

3.2. Model with Transformations on the Independent Variables

3.2.1. With Original Dependent Variable

Since the transformation on the dependent variable did not yield desirable results, various transformations on each independent variable is carried out. Partial correlation between two variables, say X and Y is the correlation between X and Y controlling the effect of other variables involved, on X and Y. This analysis suggests a way to decide on what transformed form / forms of an independent variable is / are highly correlated to the dependent variable and, therefore advisable to include in the model. We subject each independent variable to a number of transformations such as logX, $X^{1/2}$, X^2 and X^{-1}. Partial correlation between the dependent variable and each transformed independent variable is computed and the transformation which has maximum absolute partial correlation is chosen to be included in the model.

Table-3. Partial correlation between original dependent variable and various transformations of the regressors.

Candidate Transformation	X_1	X_2	X_3	X_4	X_5	X_6
Without transformation	-0.1351	0.1336	-0.1413	0.0203	0.0028	0.1278
Log transformation	**-0.1532**	0.1059	**-0.1478**	0.0105	-0.0200	0.1621
Square Root transformation	-0.1447	0.1198	-0.1446	0.0149	-0.0085	0.1453
Square transformation	-0.1132	**0.1601**	-0.1346	**0.0337**	**0.0243**	0.0910
Reciprocal transformation	-	-	-	-	-	**0.1930**

Source: Results of analyses carried out in R package.

Thus, from Table 4, it is noted that, log X_1, $(X_2)^2$, log X_3, $(X_4)^2$, $(X_5)^2$ and $(X_6)^{-1}$ are the transformations chosen for the independent variables. A new model is constructed with petrol price(Y) as dependent variable and the transformed independent variables. Even in this model, the increase in the values of R^2 and adjusted R^2 is not found adequate. Hence, the model is reconstructed by including the untransformed independent variables along with the transformed ones considered in the above model.

Table-4. Model with original dependent variable including linear and transformed regressors.

Regressors	Estimate	Std. error	t value	Pr(> | t |)
Intercept	$5.591e^{04}$	$8.941e^{03}$	6.253	$2.44e^{-08}$
X_1	$1.131e^{-02}$	$4.588e^{-03}$	2.464	0.0161
X_3	$1.679e^{-01}$	$2.790e^{-02}$	6.017	$6.50e^{-08}$
X_5	$-1.570e^{-02}$	$7.985e^{-03}$	-1.967	0.0530
X_6	-4.331	$6.445e^{-01}$	-6.720	$3.43e^{-09}$
Log X_1	$-4.166e^{02}$	$1.520e^{02}$	-2.741	0.0077
$(X_2)^2$	$2.717e^{-08}$	$1.741e^{-08}$	1.561	0.1229
Log X_3	$-1.312e^{04}$	$2.168e^{03}$	-6.051	$5.64e^{-08}$
$(X_5)^2$	$1.377e^{-06}$	$7.379e^{-07}$	1.867	0.0659
$1/X_6$	$-1.450e^{04}$	$2.012e^{03}$	-7.206	$4.32e^{-10}$
Multiple R^2	0.5411	**Adjusted R^2** 0.4845	**p value**	$1.768e^{-09}$

Source: Results of analyses carried out in R package.

From the model summary in Table 5, it is evidenced that X_1, X_3, X_5, X_6 and transformed variables of X_1 X_2, X_3, X_5, X_6 significantly affect the petrol price. The coefficient of determination, R^2 is found to be 0.5411 and adjusted R^2 is 0.4845, which is fairly decent. The Shapiro-Wilk test confirms the normal distribution for the residuals.

3.2.2. With Box-Cox Transformation on the Dependent Variable

With the original and transformed independent variables, we apply the Box-Cox approach and find that the power transformation required of the dependent variable is $Y^{0.4646}$ and with this as the dependent variable, the output is indicated in Table 6.

Table-5. Model with Box-Cox transformed dependent variable including linear and transformed regressors.

Regressors	Estimate	Std. error	t value	Pr(> | t |)
Intercept	$2.613e^{03}$	$4.159e^{02}$	6.282	$2.17e^{-08}$
X_1	$5.264e^{-04}$	$2.134e^{-04}$	2.466	0.01599
X_3	$7.836e^{-03}$	$1.298e^{-03}$	6.037	$5.98e^{-08}$
X_5	$-7.376e^{-03}$	$3.714e^{-04}$	-1.986	0.05081
X_6	$-2.036e^{-01}$	$2.998e^{-02}$	-6.790	$2.55e^{-09}$
Log X_1	$-1.938e^{01}$	$7.071e^{00}$	-2.741	0.00771
$(X_2)^2$	$1.233e^{-09}$	$8.097e^{-10}$	1.522	0.13225
Log X_3	$-6.121e^{02}$	$1.008e^{02}$	-6.070	$5.22e^{-08}$
$(X_5)^2$	$6.483e^{-08}$	$3.432e^{-08}$	1.889	0.06291
$1/X_6$	$-6.800e^{02}$	$9.362e^{01}$	-7.264	$3.37e^{-10}$
Multiple R^2	0.5422	**Adjusted R^2** 0.4857	**p value**	$1.633e^{-09}$

Source: Results of analyses carried out in R package.

This model does give a marginal, but not a substantial improvement over the previous model in Section 3.2.1. The transformations on the independent variables have resulted in significant gains for the model; however, the transformation on the dependent variable does not really give significant advantage. With the improvements derived thus far, the next matter of concern is the possibility of the presence of autocorrelation whose mitigation is the subject of interest in the next section.

3.3. Model with De-Autocorrelation

3.3.1. With Original Dependent Variable

Auto-correlation is the correlation between the successive data points of the same variable. It reflects the serial dependency present in time series data. A high correlation between the data points is a serious ill-conditioning and might adversely influence the interpretations drawn. Hence, the data is tested for presence of auto correlation using the Breuche Godfrey test of order up to 1. This shows the presence of a significant positive auto-correlation in the petrol price time series data.

Partial Auto Correlation Function (PACF) of a specified order gives partial correlations of a time series with its own lag of that order, controlling the effect of shorter lags. It is advantageous when compared to auto-correlation function (ACF) in the sense that it controls the effect of shorter lags while analyzing the correlation with longer lags. Thus, PACF is plotted for the petrol price (Y) from which we infer that the first order correlation alone is significant. This means that the current month petrol price is very much dependent on the previous month petrol price. To de-autocorrelate the data towards building the appropriate model, the petrol price with a lag of one month (Y_{t-1}) is also considered as an independent variable along with the other independent variables and the model is rebuilt. Table 7 summarizes the output:

Table-6. Model including lagged petrol price as regressor.

Regressors	Estimate	Std. error		t value	Pr(> | t |)
Intercept	$1.251e^{04}$	$7.040e^{03}$		1.776	0.0798
$Y_{(t-1)}$	$8.139e^{-01}$	$7.737e^{-02}$		10.520	$2.81e^{-16}$
X_2	$-7.340e^{-03}$	$3.837e^{-03}$		-1.913	0.0597
X_3	$3.854e^{-02}$	$2.193e^{-02}$		1.757	0.0831
X_4	$6.804e^{-04}$	$2.861e^{-04}$		2.379	0.0200
X_6	$-8.801e^{-01}$	$5.468e^{-01}$		-1.610	0.1118
$(X_2)^2$	$1.648e^{-07}$	$9.903e^{-08}$		1.664	0.1004
Log X_3	$-3.009e^{03}$	$1.711e^{03}$		-1.759	0.0828
$1/X_6$	$-2.858e^{03}$	$1.720e^{03}$		-1.662	0.1009
Multiple R^2	0.7884	Adjusted R^2	0.7652	p value	$<2.2e^{-16}$

Source: Results of analyses carried out in R package.

The summary indicates the petrol price with a lag of one month, X_2, X_3, X_4, X_6 and the transformed variables of X_2, X_3 and X_6 are found to affect the petrol price at various significance levels. The coefficient of determination, R^2 and the adjusted R^2, also shows marked increase. Auto-correlation is checked using Breuche Godfrey test and is not found significant. The desired de-autocorrelation is achieved in the above process with a substantial increase in the model performance.

From the above model summary, it is however observed that the intercept value is inflated, which indicates ill-conditioning of the data matrix. This might be due to the presence of collinearity among the regressors, which is investigated in the sequel.

3.3.2. With Box-Cox Transformation on the Dependent Variable

We rebuild the model by determining the Box-Cox power transformation required for the dependent variable with the same set of independent variables as in Section 3.3.1. The transformed dependent variable is $Y^{0.4242}$ and we get the output in Table 8:

Table-7. Model with Box-Cox transformed dependent varaible including its lag as regressor.

Regressors	Estimate	Std. error	t value	Pr(> | t |)
Intercept	$4.561e^{02}$	$2.511e^{02}$	1.816	0.0734
$Y_{(t-1)}$	$2.911e^{-02}$	$2.760e^{-03}$	10.548	$2.5e^{-16}$
X_2	$-2.596e^{-04}$	$1.369e^{-04}$	-1.897	0.0618
X_3	$1.397e^{-03}$	$7.825e^{-04}$	1.785	0.0784
X_4	$2.397e^{-05}$	$1.020e^{-05}$	2.349	0.0215
X_6	$-3.284e^{-02}$	$1.951e^{-02}$	-1.684	0.0965
$(X_2)^2$	$5.813e^{-09}$	$3.533e^{-09}$	1.645	0.1042
Log X_3	$-1.089e^{02}$	$6.103e^{01}$	-1.784	0.0785
$1/X_6$	$-1.057e^{02}$	$6.136e^{01}$	-1.723	0.0891
Multiple R^2	0.7893	**Adjusted** R^2 0.7662	**p value**	$<2.2e^{-16}$

Source: Results of analyses carried out in R package.

Again, this model gives only a small marginal improvement over the previous model in Section 3.3.1. It is interesting to note the striking similarity in the t-statistics values and the significances of the coefficients in this and the previous model.

3.4. Multicollinearity Diagnostics and Correction

Multi-collinearity is a phenomenon wherein two or more predictors are correlated, which is undesirable and results in inflated coefficient estimates and their standard errors. This in-turn leads to uncertainty in the predictive capacity of the regression model. There are various diagnostic tools for detecting the presence of multicollinearity, and in this study, we consider variance inflation factors (VIF) (as discussed by several authors including, for instance (Daoud, 2017)) and variance decomposition proportions for this purpose. These methods are executed in SAS package. We present only the variance decomposition proportions in Table 9 to exhibit the collinearity problem present:

Table-9. Variance decomposition proportions.

No	Eigen Value	Condition Index	Intercept	$Y_{(t-1)}$	X_1	X_2	X_3	X_4	X_5	X_6	Log X_1	$(X_2)^2$	Log X_3	$(X_4)^2$	$(X_2)^2$	$1/X_6$
1	13.82953	1	1.96E-09	7.85E-06	1.04E-07	3.12E-08	4.66E-10	1.06E-07	3.58E-07	6.15E-08	3.98E-07	1.25E-07	1.84E-09	3.62E-07	1.35E-06	2.12E-07
2	0.09105	12.32441	1.84E-07	0.00107	4.12E-06	1.24E-07	1.67E-08	2.79E-06	9.49E-05	9.18E-08	0.000191	4.12E-06	9.86E-09	0.000136	0.000723	2.64E-05
3	0.04727	17.10514	8.58E-07	0.00281	3.65E-06	6.96E-07	1.54E-07	3.46E-08	5.14E-05	5.65E-06	3.51E-05	1.6E-05	4.37E-07	0.000135	0.00259	1.02E-07
4	0.01538	29.98774	1.65E-07	0.000173	1.53E-05	5.56E-06	6.30E-08	1.82E-05	9.45E-07	0.000124	0.000177	5.49E-05	3.69E-07	0.000182	0.000119	0.00173
5	0.00905	39.08825	2.20E-08	3.91E-05	0.000194	1.27E-06	2.36E-08	0.000287	8.92E-06	3.55E-07	0.00313	3.54E-05	2.10E-07	0.00305	0.000195	7.75E-07
6	0.00422	57.21979	2.53E-06	0.15905	9.32E-05	3.74E-05	1.4E-06	2.83E-05	0.000336	1.52E-06	0.000234	0.00142	9.91E-06	7.10E-08	0.000624	1.14E-05
7	0.0025	74.40482	2.04E-08	0.16826	0.000139	0.000256	3.84E-07	7.2E-05	5.81E-06	1.77E-06	0.00277	0.00387	5.09E-06	0.00102	0.00127	7.45E-05
8	0.000873	125.8598	8.05E-05	0.09486	0.00107	0.000316	1.96E-06	0.00133	0.00277	0.0003	0.000446	6.7E-05	0.000203	3.38E-05	0.00511	0.000405
9	6.13E-05	475.1535	5.31E-05	0.02302	0.00579	0.000461	3.87E-06	0.02798	0.74295	0.00317	0.00449	2.81E-05	3.03E-06	0.02706	0.75972	0.00585
10	3.23E-05	653.9176	0.00339	0.01798	0.07917	0.00465	0.000105	0.20115	0.08556	0.00262	0.06611	0.00521	0.000194	0.19501	0.08036	0.000547
11	1.64E-05	919.5928	0.000496	0.00298	0.53988	0.00535	2.08E-05	0.25594	0.00356	0.06926	0.56506	0.00426	3.4E-05	0.26579	0.00479	0.07665
12	1.16E-05	1091.13	0.00156	0.13663	0.19364	0.00497	8.1E-06	0.02759	0.05765	0.60356	0.20087	0.00493	2.02E-05	0.02758	0.05706	0.59529
13	4.08E-06	1842.082	0.00281	2.35E-09	0.0705	0.98342	0.000105	0.38209	0.08678	0.01084	0.06457	0.9775	0.000121	0.37311	0.07367	0.01322
14	6.01E-08	15176	0.99161	0.39312	0.10951	0.000536	0.99975	0.10353	0.02023	0.31011	0.09191	0.00261	0.99941	0.10689	0.01376	0.30619

Source: Results of analyses carried out in SAS package.

It is clear from this output that, X_3 and its transformation considered in our model, *viz*, logX_3 are found to be entangled in a collinear relation with the intercept term, which is anomalous. To correct this, X_3 and log X_3 are removed and the model is rebuilt with the remaining regressors.

3.4.1. With Original Dependent Variable

Removing X_3 and log X_3 to build a revised model gives the following output:

Table-10. Model excluding X_3 and log X_3.

Regressors	Estimate	Std. error		t value	Pr(> | t |)
Intercept	$6.675e^{01}$	$3.254e^{01}$		2.051	0.04365
$Y_{(t-1)}$	$9.151e^{-01}$	$5.667e^{-02}$		16.147	$<2e^{-16}$
X_2	-6.000e-03	$3.364e^{-03}$		-1.784	0.07843
$(X_2)^2$	$1.293e^{-07}$	$8.763e^{-08}$		1.476	0.14400
$(X_4)^2$	$1.935e^{-08}$	$6.634e^{-09}$		2.917	0.00463
Multiple R^2	0.7777	**Adjusted R^2**	0.7662	**p value**	$<2.2e^{-16}$

Source: Results of analyses carried out in R package.

It is observed that the inflation in the intercept is rectified and variables significantly impacting the petrol price of a month are the petrol price of the previous month (Y_{t-1}), production in India (X_2) and its square transformation and $(X_4)^2$, where X_4 is India's import quantity of petroleum products, all these variables being the figures corresponding to the previous month.

3.4.2. With Box-Cox Transformation on the Dependent Variable

Subjecting the dependent variable to the Box-Cox approach with the reduced data set of only the latest significant regressors, we find the required power transformation as $Y^{0.6263}$ and the output with this transformed dependent variable is shown in Table 11:

Table-11. Model with Box-Cox transformed dependent variable excluding X_3 and log X_3.

Regressors	Estimate	Std. error		t value	Pr(> | t |)
Intercept	$1.401e^{01}$	$4.088e^{00}$		3.426	0.000984
$Y_{(t-1)}$	$1.151e^{-01}$	$7.119e^{-03}$		16.167	$<2e^{-16}$
X_2	$-7.653e^{-04}$	$4.226e^{-04}$		-1.811	0.074021
$(X_2)^2$	$1.653e^{-08}$	$1.101e^{-08}$		1.502	0.137269
$(X_4)^2$	$2.425e^{-09}$	$8.333e^{-10}$		2.910	0.004717
Multiple R^2	0.7782	**Adjusted R^2**	0.7667	**p value**	$<2.2e^{-16}$

Source: Results of analyses carried out in R package.

This is again a very small gain over the model in Section 3.4.1 and the t-statistics and significance values are all found to be very similar. Thus, for ease of interpretation, we decide to go ahead with the model without the Box-Cox transformation. We find that, apart from the single lag value of Y, only linear and quadratic terms of X_2 and X_4 have been found significant. Hence, in the next section, we just build a model with the linear and quadratic terms along with the lagged value of the dependent variable.

3.5. Model with Linear and Quadratic Terms of Regressors

With the objective of identifying the level(s) of the factor(s) at which petrol prices can attain minimum and to maintain a de-autocorrelated structure, we build a model with Y_{t-1} and the linear and quadratic terms of all the regressors except X_3, and obtain exactly the same model as in Section 3.4.1. This model has a decent value of R^2 and Adjusted R^2 and we therefore, decide to finalize the model obtained in Section 3.4.1. Interestingly, only X_2 is found to give an answer for the question on finding the point of minimum petrol price attainable. We refer to Table 10 for the model coefficients.

Inference:

- From the model summary, it is inferred that the petrol price of any month is found to hover around 92% of the previous month's petrol price, with the actual ups and downs in price level being determined by other factors.
- Both linear and quadratic terms of X_2 are found significant in affecting petrol prices. The convex relationship between the response variable and X_2 (fixing X_4), provides the level at which X_2 gives minimum value of the response variable. We find that this point of minimum is X_2 = 23201.86 thousand metric tonnes. That is, the price of petrol in a month would have reduced greatly when the previous month's production in India(X_2) were about 23200 thousand metric tonnes. But, from the data set, it is observed that this level of monthly production of petroleum products has not been achieved so far.
- Significance of $(X_4)^2$ shows that petroleum imports to India have a non-linear effect on prices. The incremental impact of imports on price is not 'static' at the various import levels, but varies with the actual quantum of imports. Table 12 gives a brief view of this aspect.

Table-8. Incremental Effects of Import Levels on per-litre Petrol Price.

Import Level	A 1000 Metric Tonne Incremental Effect
15000	0.295088
16000	0.314438
17000	0.333788
18000	0.353138
19000	0.372488
20000	0.391838
21000	0.411188

Source: Computations done using R package.

For example, when the actual import level in a month is about 15000 thousand metric tonnes, an increase of 1000 thousand metric tonnes of import leads to a 29 paise increment in the per-litre petrol cost in the subsequent month. Similar interpretation may be given for the figures corresponding to other import levels.

4. CONCLUDING REMARKS

The rise in petrol price is indeed a disturbing phenomenon with questions and doubts raised by the common man about the reasons for the same. The final regression model obtained in Section 3.4.1 Refer to Table 10 sheds some light on the causative factors behind the increase in petrol price. India's ability to indigenously produce petroleum is low as it stands in the 24th position when ranked according to the number of oil reserves found. (Refer to https://en.wikipedia.org/wiki/List_of_countries_by_proven_oil_reserves).

It is also evident that there has been a steady decrease in oil production in India as reported in economic review articles (For example, we refer to the 2018 May 4th edition of 'ET Energy World'). Hence, to meet the increasing demand, India's imports of petroleum products have increased to an extent that a major portion of the petroleum product requirements is met by imports. India ranks 3rd among the largest importers of petroleum products. The intriguing aspect is that, increase in imports lead to meeting higher demand for petroleum products comfortably but, there is an increase in the market price as the burden of the import costs appear to be directly imposed on the common man. The interesting conclusion that we derive from our model is that increased domestic production will help contain the petrol prices. Search for additional domestic oil sources and research towards identifying domestically available alternatives to petrol is called for to reduce the burden on the exchequer and the common man as well.

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Contributors/Acknowledgement: Both authors contributed equally to the conception and design of the study.

REFERENCES

Box, G.E. and D.R. Cox, 1964. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26(2): 211-243.

Daoud, J.I., 2017. Multicollinearity and regression analysis. Journal of Physics, 949(1): 1-6.

Gyagri, M., E. Amarfio and S. Marfo, 2017. Determinants of global pricing of crude oil-a theoretical review. International Journal of Pharmacology, Phytochemistry and Ethnomedicine, 3(3): 7-15.

Hidhayathulla, A. and M. Rafee, 2014. Relationship between crude oil price and rupee, dollar exchange rate: An analysis of preliminary evidence. IOSR Journal of Economics and Finance, 3(2): 1-4.

Nwakuya, M.T. and J.C. Nwabueze, 2018. Application of box-cox transformation as a corrective measure to heteroscedasticity using an economic data. American Journal of Mathematics and Statistics, 8(1): 8-12.

Poirier, D.J., 1978. The use of the box-cox transformation in limited dependent variable models. Journal of the American Statistical Association, 73(362): 284-287.Available at: https://doi.org/10.2307/2286653.

Razali, N.M. and Y.B. Wah, 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of Statistical Modeling and Analytics, 2(1): 21-33.

Spitzer, J.J., 1978. A Monte Carlo investigation of the box—cox transformation in small samples. Journal of the American Statistical Association, 73(363): 488-495.

Spitzer, J.J., 1982. A primer on box-cox estimation. The Review of Economics and Statistics, 64(2): 307-313. Available at https://doi.org/10.2307/1924310.

Srithar, P., N. Bairavi and G. Mariselvam, 2015. Oil price volatility and its impact on the selected economic indicators in India. International Academic Research Journal of Economics and Finance, 3(4): 10-16.

Swadimath, U.C., K.H. Anilkumar and P.B. Joshi, 2013. Rise and impact of crude oil price in India. International Journal of Marketing, Financial Services and Management Research, 2(1): 82-91.