

Relationship between Crude Oil Price and S&P BSE Stock Index - An Integration of Binary Decision Tree and Logistic Regression Approach

Rahul Kumar Si

Tobacco Board, Ministry of Commerce & Industry, Govt. of India, Guntur – 522004 (AP), India.

ARTICLE INFO

Article history:

Received: 3 March 2020;

Received in revised form:
29 March 2020;

Accepted: 9 April 2020;

Keywords

Crude Oil,
Stock Market,
Financial Ratio,
Binary Classification,
Logistic Regression and
Hosmer-Lemeshow Test.

ABSTRACT

The relationship between crude oil prices and stock market indices has always been discordant. The article examines the performance of stock market with the help of different financial ratios used in oil and natural gas sector. Seventeen distinct companies involved in S&P BSE index are taken for our study. Initially 39 financial ratios are taken for analysis and a normality test was conducted on all these explanatory variables to find whether these variables are normal or not? A logistic regression was constructed with help of the selected nine financial ratios as independent variables for determining the performance of S&P BSE index as dependent variable. The classification result showed that, about 75.3 % accurate result for prediction 'GOOD' as well as 'NOT GOOD' performance of stock. Hosmer-Lemeshow test is applied to find out the overall model evaluation. The study also showed that logistic regression can be used by investors, fund's managers, investment companies and researchers for academic interest as well as policy implications.

© 2020 Elixir All rights reserved.

1. Introduction

Crude oil is gaining its importance as a lifeline to the world's economy. Now it is the most actively traded commodity in the world. It is influenced by many factors like socio-economic, political and status of financial market of respective country. From medium to long run it is influenced by the fundamentals of demand and supply and highly related to the economic activities of that country. Global crude oil fluctuations affect the economy of the nations in a positive or negative manner depending whether the country is net importer or exporter of crude oil. Hence increase in crude oil prices definitely impact on world's economy through employment, rising inflation, decrease in exchange value all of which combine to economic slowdown, whereas decline in crude oil prices helps the government to manage the financial sector better as it translates into lower subsidies on petroleum product, thereby resulting in lower fiscal deficit. It helps the government to remain committed to fiscal consolidation road map without compromising on economic growth.

India's oil and gas sector is of strategic importance and plays a predominantly pivotal role in influencing decision in all other sphere of the economy. As a developing country, it is committed to excel its economy in the upcoming years. The India's crude oil and gas sector is one of the six core industries in India and has very significant foreword linkages with the entire economy. India being a net importer, the decrease in crude oil prices is a welcome incentive and provides an opportunity to strengthen the fiscal position. Even, government has taken this energy turmoil as an advantage to reduce the subsidies on fuel consumption and thereby strengthen the fiscal position.

2. Petroleum Industry in India – At A Glance

Like many other Indian prominent industry, the development of the oil and gas industry began slowly. The origin of this industry can be seen from 19th century, where oil

exploration started in Digboi in the state of Assam in 1889. Later the government realise the significance of the oil and gas sector for economic growth. Therefore, government under the Industrial Policy Resolution of 1954 (IPR), declared that oil and gas sector would be the crux sector industry. Consequential to IPR 1954, the entire sector was controlled by the state-owned companies. With the discovery of the Cambay and the Bombay offshore basin, the domestic oil production increased remarkably. Therefore, in the 1970s, almost 70% of the country's crude requirement was met through domestic production. However, in 1973 the OPEC has decreased crude production and declared an embargo on oil exports to the United States and the Netherlands (the supporters of Israel). After this oil shock, the government of India nationalised this sector. This act of the Indian government force the major international players exit the Indian oil and gas industry. Apart from this, the Indian government imposed lot of restriction on the pricing and distribution mechanism of oil and gas products in India. Later major determinants like technology, distribution etc. increases the problem of crude sector in India. In the early 90's the government allowed the MNC's to take part in the bidding process. In 1995, the Government declared the joint venture program with private players. Subsequent to the various reforms taken by the government, the area under oil exploration has increased to about 50%. In this movement, RIL was made world's largest gas discoveries in Jamnagar, Gujarat. Further, the sector is witnessing the entry of various multinational companies into India.

India is at present the fastest growing major economy of the world. The robust growth in the economy has triggered the energy demand and India is positioned to drive the incremental demand growth in global energy arena. GDP growth at constant (2011-12) prices has averaged 7.3 % for the period from 2015-16 to 2017-18, which is highest among

the major economics of the world. GDP at constant prices is estimated at Rs.130.11 lakh crore showing a growth rate of 6.7 % over the year 2016-17 of Rs. 121.96 lakh crore. The growth is against a backdrop of resurgence in exports coupled with global tensions on trade front and volatile global oil market with a distinct hardening crude oil prices. On the domestic front, the strong growth is underpinned by robust private consumption, expectation of greater stability in GST and public investment as well as ongoing structural reforms.

The crude oil production during the year 2017-18 is at 35.68 MMT as against production of 36.01 MMT in 2016-17, showing a decline of 0.90 %. Similarly, in natural gas production, 2.36 % higher increase in 2017-18 as compared to 2016-17. The trends in the production of crude oil and natural gas for the year 2011-12 to 2017-18 have been depicted in Table 1.1 & Figure 1.1.

Table 1.1. Crude Oil & Natural Gas Production

Year	Crude Oil Production (MMT)	% Growth	Natural Gas Prod. (BCM)	% Growth
2011-12	38.09	1.08	47.56	-8.92
2012-13	37.86	-0.60	40.68	-14.46
2013-14	37.79	-0.19	35.41	-12.96
2014-15	37.46	-0.87	33.66	-4.94
2015-16	36.94	-1.39	32.25	-4.18
2016-17	36.01	-2.53	31.90	-1.09
2017-18	35.68	-0.90	32.65	2.36

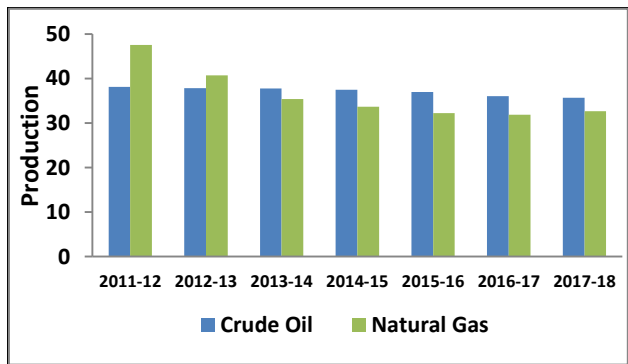


Figure 1.1. Crude Oil & Natural Gas Production

Source: Report from Ministry of Petroleum and Natural Gas, Govt of India.

Note: Crude oil and Natural Gas Production are in MMT and BCM respectively in Figure – 1.1.

Indian refinery industry, which is second largest refiner in Asia after China, is merging as a refinery hub with refining capacity exceeding demand. Crude oil processed for the year 2017-18 is 251.93 MMT, showing an increase about 2.68 % as compared to 2016-17. Similarly refinery capacity utilization is 107.67 for the year 2017-18. The production of petroleum product is at 254.40 MMT in 2017-18 against 243.55 MMT achieved in 2016-17. During 2017-18 the consumption of petroleum product was 204.92 MMT with a growth of 5.31 % as compared to 2016-17. Import of crude oil during 2017-18 was 220.43 MMT valued at Rs. 5,65,951/- crore as against import of 213.93 MMT valued at Rs. 4,70,159/- crore in 2016-17, which marked an increase of 3.04 % in quantity terms and 20.37 % in value terms as compared to 2016-17. Similarly average international crude oil price was USD 56.43 / bbl during April-March, 2017-18 as compared to USD 47.56 / bbl in the same period of 2016-17. The trends in growth of crude oil imports and crude oil international (Indian Basket) prices are shown in Table 1.2 & Figure 1.2.

Table 1.2. Import and Avg. Price of Crude Oil

Year	Import of Crude Oil (MMT)	% Growth	Avg. Crude Oil Prices ^a	% Growth
2011-12	171.73	7.97	111.89	31.50
2012-13	184.80	7.61	107.97	-3.50
2013-14	189.24	2.40	105.52	-2.27
2014-15	189.43	0.10	84.16	-20.25
2015-16	202.85	7.08	46.17	-45.14
2016-17	213.93	5.46	47.56	3.02
2017-18	220.43	3.04	56.43	18.65

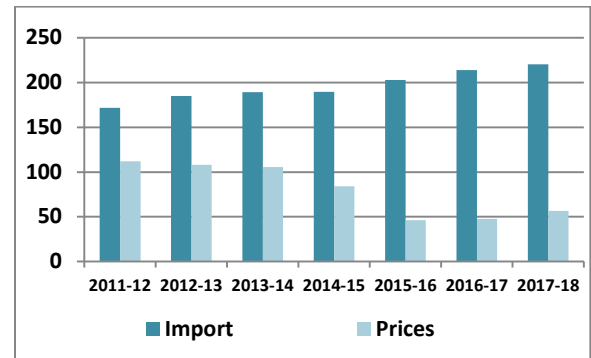


Figure 1.2. Import and Avg. Price of Crude Oil

Source: Report from Ministry of Petroleum and Natural Gas, Govt of India a: USD/bbl

Note: Import of crude oil in MMT and Avg. Price in USD/bbl.

During 2017-18 imports of petroleum products were at 35.89 MMT valued at Rs. 86,946/- crore which shows decrease of 1.09 % in quantity terms and 21.49 % increase in value terms against 36.29 MMT imports of petroleum products valued at Rs. 71,566 crore during 2016-17. Similarly exports of petroleum product were 66.76 MMT and 65.51 MMT in 2017-18 and 2016-17 respectively. Import of LNG was 19.87 % which registered an increase of 6.65 % in quantity terms and 22.39 % increase in value terms as compared to 2016-17.

Since crude oil plays an important role, many initiatives have been taken by government to increase production and exploitation of all domestic petroleum resources to address the priorities like Energy Access, Energy Efficiency, Energy Sustainability and Energy Security. From the perspective of developing economy like India crude prices play a prominent role on the inflation and current account deficit. The recent decline in crude prices and its significant impact on the economy needs investigation from the perspective of developing economy. In addition, the stock markets are the barometers of any economy and they reflect a small change in the macroeconomic factor. It is very essential in understanding the basic causes of drop in crude prices and its impact on the stock market.

3. Review of Literature

Various researchers did a remarkable work to evaluate the association between changes in crude oil price and stock market returns using different tools and techniques as it played an important role in the stock returns (Hamilton, 1983). Various researchers established the nexus relationship between oil price and macroeconomic variables (Aydogan & Berk, 2015; Cüppers & Smeets, 2015; Bass, 2017; Ulusoy & Özdurak, 2018; Ojikutu et al. 2017; Hooker, 1996; Burbridge & Harrison, 1984; Gisser & Goodwin, 1986; Cobo-Reyes & Quiros, 2005 and Basher & Sadorsky, 2006).

Jones & Kaul, (1996) studied the impact of stock market crude oil shocks by analysing present and near future fluctuations in cash flow in expected returns of the market and found that the crude prices allow predicting stock returns in developed and regulatory environment economies except for England. but Huang et al. (1996) found that oil future returns have some impact on the individual oil company returns but don't have any significant impact on the market indices. Chen et al. (1986) and Apergis & Miller, (2009) reported the similar findings, whereas Sadorsky, (1999) found the real stock returns is due to changes and volatility in the oil prices and crude stock have a negative and significant initial impact on the stock return (Papapetru, 2001).

Basher et al. (2012) studied the positive oil price shocks tend to lower emerging markets stock prices and US dollar exchange rates in the short term. Farzanegan & Markwardt, (2009) emphasizes on "Dutch Disease" syndrome through significant real effective exchange rate appreciation in the Iranian economy to be highly vulnerable to oil price fluctuations. El-Sharif et al. (2005) investigated the relationship between the crude oil price and equity values in the oil and gas sector and found that the relationship was always positive, often highly significant and reflects the direct impact of volatility in the price of crude oil on share values within the sector.

Ciner, (2001) studied a nonlinear feedback relation between oil and the stock market and the linkage between oil prices and the stock index movements. The impact of oil price shocks and oil price volatility on the real stock returns studied by Park & Ratti, (2008) using multivariate VAR analysis and an error variance decomposition method to establish relationship between oil price shocks and real stock returns and between oil price shocks and interest rate, but Miller & Ratti (2009), investigated the long-run association between the crude prices and stock markets by using VECM to find a long-run association between crude oil price and stock returns.

Bildirici, Melike & Essin, (2009) and Arouri et al. (2011) used AP-GARCH and VAR-GARCH to study the returns and volatility transmission between oil prices and stock markets and find improved forecasting result as compared to traditional method, but Bharn & Nikolova, (2010) used bivariate E-GARCH to examined the dynamic correlation between stock market and oil prices when the incidences like 9/11 terrorist attack, 2003 Iraq war and 2006 civil war in Iraq.

Mork, (1989) investigated the connection between the oil prices changing effect the inflation rate and found a significant relationship between the oil prices and inflation rate. Dhaoui & Khraief, (2014) found a strong negative association in selected seven countries between oil price and stock market returns. Cong et al. (2008) tried to explore the relationship between crude oil price shock and Chinese stock market; they did not find statistical evidence for the stated objectives in the Chinese stock market. Aloui et al. (2008) found that crude prices play a crucial role in forecasting the stock market behaviour.

A statistical support for weak form of efficiency over a wide range of time-scales was studied by Alvarez & Solis, (2010). Inayat, (2010) investigated the relationship between crude prices and stock performance of European automobile companies and found that the oil is not having a significantly adverse impact on auto returns, but Maghyereh, (2004) contradicted the view of Jungwook & Ronald, (2008) by using VAR models and concluded that the crude price shocks

have no significant impact on index returns. Ono, (2011) examined the influence of crude oil prices on BRICS stock markets. He concluded that the stock returns of China, Russia and India have a positive impact and the Brazil stock returns do not show any statistical significance.

Ready, (2013) concluded that the oil demand shocks positively correlated with stock returns; however, oil supply in the counties has a negative correlation with stock returns. Co-integration and VECM methods used by Akomolafe, Jonathan & Danladi, (2014) to investigated the association between company's stock returns and variations in crude oil returns to analyse whether the banking sector fluctuate with change in oil prices or not? Subarna & Ali, (2012) examined the co-movements of macroeconomic variables and find there is a co-integration association between the variables and crude oil prices. Chen et al. (2010) investigated the relationship between high oil prices and its impact on stock market returns by taking S&P 500 Price Index as proxy to find out high probability of a bear market emergence as a result of increase in crude prices and Nandha & Faff, (2008) studied the short-term relationship between crude prices and thirty-five prominent global industries. The findings of the study revealed that the crude oil prices have a negative impact on all of studied sectors except the oil and gas sector.

Awerbuch et al. (2006) reported that increase in crude price and volatility in crude decreases economic growth of an economy by increasing inflation, but Somoye & Ilo, (2008) found that crude oil, inflation and exchange rate helps to determine the stock returns. In an empirical study by Chaudhuri & Daniel, (1998), claimed that oil price impacts the stock market in long-run (Bhunia, 2013). Ojebiyi & Wilson, (2011) found a negative relationship between crude prices and exchange rates and Papapetrou, (2001) found the economic activity and employment were affected by changes in crude oil moment. Hidhayathulla & Rafee, (2012) found Continuous import of crude leads to increase in demand for dollar and in turn this leads to weaken Rupee value against dollar.

The above studies do not find the conclusive evidence on influence of crude price on the stock returns. Current study differs due to various following reasons. First, bulk of the literature focus on crude oil prices fluctuations and its impact on macroeconomic variables such as inflation, interest rate, employment rate, forex rate, growth rate etc. However, couple of studies undertaken to examine the association between crude oil returns and the stock returns, which have been studied from the perspective of emerging stock market and oil rich nation's stock market, only a few studies have concentrated on emerging markets like India. Not much empirical studies have been conducted from the Indian stock markets perspective. Therefore, the current study tries to analyse the association between crude returns and its impact on Indian stock returns.

4. Objective of the Study

The objective of this study are to determine the relationship between crude oil price and stock market through selected financial ratio and build a model using these ratios to predicting the Indian stock market by employing logistic regression. The present study is focused on the following objectives.

- ❖ Check whether the explanatory variables (Selected Financial Ratios) are normal or not.
- ❖ Analyse the stocks with the help of financial ratios and its movement.

- ❖ Analyze stock yields using a logistic regression model.
- ❖ Test how well the model performs with Hosmer-Lemeshow test.

5. Methodology

The study examines the impact of crude oil price on S&P BSE stock index through selected financial ratios while other are assumed to be constant. A normality test is applied to find out whether the selected financial ratios are normal or not? Data have been analysed by applying binary classification and logistic regression and Hosmer-Lemeshow test is applied to find out overall model evaluation.

6. Financial Ratios

Financial ratios have played an important part in evaluating the performance and financial condition of an entity. Over the years, empirical studies have repeatedly demonstrated the usefulness of financial ratios. Beaver, (1966) matched a sample of failed firms with a sample of non failed firms and studied their financial ratios for a period of up to five years before failure and found that they had high predictive ability. Altman, (1968) used a well known multivariate statistical technique in the social sciences called multiple discriminant analysis (MDA). This was popularised as the Z-score model and was successfully marketed for credit analysis, investment analysis and going-concern evaluation.

Statistical models using financial ratios have been used to identify the financial characteristics of problem in banks (Sinkey, 1975 and Pettway & Sinkey, 1980) lending decisions and capital adequacy (Dince & Fortson, 1972). Both Pinches & Mingo, (1973) and Eisenbeis, (1977) identified a number of difficulties arising from the statistical assumptions made in the application of the technique which researchers did not usually address. Dombolena & Khoury, (1980) found a substantial amount of instability in the financial ratios (as measured by their standard deviations and their coefficients of variation) in the ratios of firms which went bankrupt compared with those that did not. This instability increased over time as the firm neared failure.

Norton & Smith, (1979) compared the performance of a MDA bankruptcy prediction model using traditional historical cost data and using data adjusted for changes in general price-levels (GPL). They found these were similar, although Solomon & Beck, (1980) showed how the model was biased against a finding of predictive GPL data, and Ketz, (1978) found that GPL data slightly improved performance. Mensah, (1983) came to a similar conclusion as Norton & Smith, (1979) concerning specific price-level data, and Bazley, (1976) using a simulation approach, found that both were slightly inferior to historical cost. Short, (1980) used factor analysis to test whether empirical classifications were similar under historical and price-level accounting. He found that they were unaffected, suggesting that the meaning of a ratio is not altered by a price-level adjustment.

Financial ratios have been used to assess and forecast company risk in other contexts. Falk & Heintz, (1975) used industry financial ratios in what is called a partial order scalogram technique to scale industries according to their degree of risk. To a similar end, Gupta & Huefner, (1972) used cluster analysis to relate ratios to established economic characteristics of the industries involved. However, the biggest development has been the prediction of betas as measures of risk using financial ratios. Early work was by Thompson, (1976) and Bilderssec, (1975) who examined such correlations and now commercial services are available for

practitioners (Foster, 1986). There have also been studies investigating the statistical relationship between financial ratios and rates of return on common stocks (O'Connor, 1973 and Roenfeldt & Cooley, 1978), in which the inference is that ratios are useful in forecasting future rates of return.

In this context, the companies dealing with crude oil and natural gas are taken into consideration, of which, most of these companies are part of the S&P BSE index. The financial data used in this model were collected from the Web sites of respective company from 2014-15 to 2018-19. Nine financial ratios were taken for analysis and summarized in Table 2.

Table 2. Selected Financial Ratios and their Abbreviations

Sl No.	Name of the Financial Ratio	Abbreviation
1	Dividend per Share	DPS
2	Net Assets Value	BNAV
3	Earnings Before Interest, Taxes, Depreciation and Amortization	EBITDA
4	Earnings Before Interest and Taxes	EBIT
5	Asset Turnover	AT
6	Price Earning Ratio	PER
7	Cash Earning per Share	CEPS
8	Dividend Yield	DY
9	Enterprise Value – to - Sale	EVS

A normality test was conducted on all these explanatory variables and summarized in Table 3, which shows that these variables are normal and were also tested using Q-Q plot technique and were not considered for further analysis.

7. Logistic Regression

Logistic regression which is helpful for prediction of the presence or absence of a characteristic or outcome based on values of a set of predictive variables is a multivariate analysis model (Lee, 2004 and Pardo et al. 2005). They also confirmed that through the accumulation of suitable association function to the standard linear regression model.

Altman, (1968) who is considered as the pioneer of this area, advocated that in order to evaluate its likely impact thoroughly, while Ohlsen, (1980) constructed the default prediction model to access credit risk information and indicated that these model are highly efficient in forecasting financial distress and bankruptcy for probit analysis (Zimijewski, 1984 and Zavgren, 1985).

Abdel-Khalik, (1974) advocated upon analyzing the strictures of the regression of the principal sample to anticipate the rate of return of the holdout sample and observed that trustworthiness of the prediction is a direct function of the reliability of the strictures and the regression equation itself. Mc Connell et al. (1986) applied MLP design through logistic map and the Glass-Mackey equation, which have the ability to intimate and forecast ever changing non-linear system.

Aminian et al. (2006), Yu, Wang & Lai, (2009) and Aiken & Bsai, (1999) applied MLP technique to predict advertising and marketing trends, macroeconomic data, financial time series forecasting and stock market trends respectively. Jaffe & Wasterfield, (1985) and Kato et al. (1990) found that these models are effective enough to determine at least a significant behavior of index return. Lee, Ryu & Kim, (2007) emphasized that logistic regression can come to handy in conditions where prediction of the existence or deficiency of an outcome or feature is dependent on values of a set of predictor variable. Logistic regression technique yields coefficients for each independent variable based on a

sample of data (Huang, Cai & Peng, 2007). Logistic regression model with two or more explanatory variables are widely used in practice and parameters are commonly estimated by maximum likelihood (Pardo, Pardo and Pardo et al. 2005).

7.1. Analysis of Model

Binary logistic regression deals with situations in which the observed outcome for a dependent variable can have only two possible types. The outcomes are usually coded as zero or one as this leads to most straightforward interpretations.

The explanation of logistic regression can be expressed by a standard logit function, which is also a sigmoid function, takes any real input 't', ($t \in \mathbb{R}$) and output between zero and one. The standard logistic function $\sigma: \mathbb{R} \rightarrow (0, 1)$ is defined as

$$\sigma(t) = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}}$$

Let us assume that 't' is a linear function of single explanatory variable 'x' and expressed as

$$t = \beta_0 + \beta_1 x$$

And the general logistic function $p: \mathbb{R} \rightarrow (0, 1)$ can be written as

$$p(x) = \sigma(t) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Where, $p(x)$ = Probability of the dependent variable Y equaling a GOOD rather than NOT GOOD. Here the response variable Y_i are not identically distributed: $P(Y_i = 1|X)$ differs from one data point X_i to another. The logit (log odds) function as the inverse $g = \sigma^{-1}$ of the standard logistic function

$$g(p(x)) = \sigma^{-1}(p(x)) = \text{logit } p(x) = \beta_0 + \beta_1 x$$

And equivalently, after exponentiating both sides, the odds is

$$\frac{p(x)}{1 - p(x)} = e^{\beta_0 + \beta_1 x}$$

A generalized linear model function parameterized by θ ,

$$h_\theta(X) = \frac{1}{1 + e^{-\theta^T X}} = \text{Pr}(Y = 1|X; \theta)$$

Therefore,

$$\text{Pr}(Y = 0|X; \theta) = 1 - h_\theta(X)$$

And since

$Y \in \{0,1\}$, $\text{Pr}(Y|X; \theta) = h_\theta(X)^y (1 - h_\theta(X))^{(1-y)}$. Now the likelihood function (subject to condition – all the observations in the sample are independently Bernoulli distributed)

$$L(\theta|x) = \text{Pr}(Y|X; \theta)$$

$$= \prod_i h_\theta(x_i)^{y_i} (1 - h_\theta(x_i))^{(1-y_i)}$$

And the log likelihood maximized by **gradient descent** as

$$N^{-1} \log L(\theta|x) = N^{-1} \sum \log \text{Pr}(v_i|x_i; \theta)$$

$$\begin{aligned} \lim_{n \rightarrow +\infty} N^{-1} \sum_{i=1}^N \log \text{Pr}(y_i|x_i; \theta) \\ = \sum_{x \in X} \sum_{y \in Y} \text{Pr}(X = x, Y = y) \log \text{Pr}(Y = y|X = x; \theta) \\ = \sum_{x \in X} \sum_{y \in Y} \text{Pr}(X = x, Y = y) \left(-\log \frac{\text{Pr}(Y = y|X = x)}{\text{Pr}(Y = y|X = x; \theta)} \right. \\ \left. + \log \text{Pr}(Y = y|X = x) \right) \\ = -D_{KL}(Y||Y_0) - H(Y|X) \end{aligned}$$

Where, $H(Y|X)$ is conditional entropy and D_{KL} is the Kullback-Leibler divergence. So minimize the log likelihood, the K-L divergence is minimized from the maximal entropy distribution.

Logistic regression models are frequently used to predict a dependent variable from a set of independent variables. An important question is whether results of the logistic regression analysis on the sample can be extended to the population the sample has been chosen from. This question is referred as model validation. In practice, a model can be validated by deriving a model and estimating its coefficients in one data set, and then using this model to predict the outcome variable from the second data set, then checks the residuals and so on. When a model is validated using the data on which the model was developed, it is likely to be over estimated. Thus, the validity of model should be assessed by carrying out tests of goodness of fit and discrimination on a different data set (Giancristofaro & Salmaso, 2003). If the model is developed with a sub sample of observations and validated with the remaining sample, it is called internal validation. The most widely used methods for obtaining a good internal validation are data splitting, repeated data splitting, Jackknife technique and Bootstrapping (Harrell, Lee & Mark, 1996). If the validity is tested with a new independent data set from the same population or from a similar population, it is called external validation. Obtaining a new data set allows us to check the model in a different context. If the first model fits the second data set, there is some assurance of generalizability of the model. However, if the model does not fit the second data, the lack of fit can be either due to the different contexts of the two data sets, or true lack of fit of the first model.

The estimated results of the logistic regression model of the stock price return performance, along with the whole sample are summarized in Table 3.

The final logistic regression equation is estimated by using the maximum likelihood estimation for classifying a company:

Table 3. Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	DPS	0.070	0.069	1.035	1	0.309	1.072
	BNAV	-0.008	0.004	4.585	1	0.032	0.992
	EBITDA	-0.018	0.034	0.287	1	0.592	0.982
	EBIT	-0.018	0.033	0.291	1	0.589	0.982
	AT	-0.628	0.477	1.731	1	0.188	0.534
	PER	-0.090	0.046	3.757	1	0.053	0.914
	CEPS	0.082	0.060	1.856	1	0.173	1.085
	DY	-0.180	0.205	0.772	1	0.380	0.835
	EVS	-0.180	0.222	0.658	1	0.417	0.836
Constant	3.013	1.497	4.050	1	0.044	20.346	

a. Variable(s) entered on step 1: DPS, BNAV, EBITDA, EBIT, AT, PER, CEPS, DY, EVS.

$$Y = 0.070DPS - 0.008BNAV - 0.018EBITDA - 0.018EBIT - 0.628AT - 0.090PER + 0.082CEPS - 0.180DY - 0.180EVS + 3.013$$

Where $Y = \log(p/1-p)$, and

'p' is the probability that the outcome is GOOD.

In the above equation, it is possible to classify a company by calculating Y values. If the p value obtained from Y is higher than 0.42, then the stock was classified as GOOD; and, if it is lower than 0.42, then the stock was classified as NOT GOOD (Neter et al. 1996).

8. Binary Classification

For the purpose of carrying out logistic regression analysis, it is required for classifying a company as a "GOOD" or "NOT GOOD" investment choice for a given year. Although there is no such method for defining a market investment as "GOOD" or "NOT GOOD". In this study we use a method that is simple and objective – namely, if the value of a company's stock over a given year is positive, it is classified as a "GOOD" investment option; otherwise, it is classified as a "NOT GOOD" investment option. Here the S&P BSE SENSEX return has been taken as proxy for market return. To obtain the return at the end of each financial year, the March ending prices were used for each year. The return was calculated using the following formula:

$$\text{Return of Stock} = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100$$

Where, P_t = Price at 't' Year and P_{t-1} = Price at 't-1' Year

Table 4. Dependent Variables and their Encoding

Type of Company (Based on stock market return)	Classification	Internal Value
GOOD	Return above Market return (S&P BSE Index)	0
NOT GOOD	Return below Market return (S&P BSE Index)	1

8.1 Classification Accuracy

The following classification table helps to assess the performance of the model by cross-tabulating the observed response categories with the predicted response categories. For each case, the predicted response is the category treated as 1, if that category's predicted probability is greater than the user-specified cut off. The cut off value is taken at 0.5.

The Table 5 shows the comparison of the observed and the predicted performance of the companies and the degree of their prediction accuracy. It also shows the degree of success of the classification for this sample. The number and percentage of cases correctly classified and misclassified are displayed. It is clear from this table that the NOT GOOD companies have a 94.5% correct classification rate, whereas GOOD companies have a 40.0% correct classification rate. Overall, correct classification was observed in 75.3% of original grouped cases.

9. Evaluation of the Model and Empirical Result

A logistic regression model with the k independent variables is said to provide a better fit to the data if it demonstrates an improvement over the model with no independent variables (the null model). The overall fit of the

model with k coefficients can be examined via a Hosmer-Lemeshow test which tests the null hypothesis

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots \dots \dots \beta_k = 0$$

The Hosmer-Lemeshow goodness of fit test involves grouping the observations based on the expected probabilities and then testing the hypothesis that the difference between expected and observed events is approximately zero for all the groups. It is distributed as chi-square when there is no replication in the sub-populations (Hosmer & Lemeshow, 1989).

Goodness-of-fit statistic χ^2 is obtained by calculating the Pearson chi-square statistic from $g \times 2$ table of observed and estimated expected frequencies. A formula defining the calculation of \hat{C} is:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Here g is number of groups, n'_k is total number of subjects in the k-th group, c_k denotes the number of covariate patterns in the k-th decile,

$$O_k = \sum_{j=1}^{c_k} y_j$$

is the number of responses among the c_k covariate pattern and

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

is the average estimated probability.

Main pre assumptions for this test are:

- Sample is divided on two separate subpopulations corresponding to cases of presence and absence of some property.
- Probabilities for covariance pattern, unique combination of values of predictor variables, are π_k and $1 - \pi_k$ for presence and absence of some property, respectively; their sum is 1 for k-th decile.
- Estimate of expected frequencies are $m_j \hat{\pi}_j$ and $m_j (1 - \hat{\pi}_j)$ respectively, for the cell corresponding to $y = 1$ and $y = 0$ rows.
- Sum of observed and expected frequencies for k-th decile are the same

$$O_{1k} + O_{0k} = E_{1k} + E_{0k} = N_g$$

Here O_{1g} , E_{1g} , O_{0g} , E_{0g} and N_g denote sample $Y = 1$ values, expected $Y = 1$ values, sample $Y = 0$ values, expected $Y = 0$ values, number of observations in group g respectively.

To make groups of values, Hosmer and Lemeshow proposed two strategies. With the first method, percentiles of risk, use of $g = 10$ groups result in the first group containing the $n'_1 = n/10$ subjects having the smallest estimated probabilities and the last group containing the $n'_{10} = n/10$ subjects having the largest estimated probabilities. With the second method, use of $g = 10$ groups results in cut points defined at the values $k / 10$, $k = 1, 2, \dots, 9$ and the groups contain all subjects whose estimated probability between adjacent cut points.

Table 5. Classification Table.

Observed			Predicted		Percentage
			Performance		
STEP -1	Performance	NOT GOOD	POOR	GOOD	94.5
				GOOD	
		Overall Percentage	18	12	75.3

The Hosmer-Lemeshow statistic which provides the useful information about the robustness of the model illustrated in Table 6.

Table 6. Hosmer -Lemeshow Test

Step	Chi-square	df	Sig.
1	13.556	7	0.060

The observed significance level for Chi-square value is 0.06 which indicates acceptance of null hypothesis of the model (there is not much difference between observe and predicted value). The Chi-square value 13.556 of this model indicates that logistic regression is very much meaningful in accordance with the dependent variable relating to each specified independent variable.

10. Conclusion

This study used the binary logistic regression model to determine the factors that significantly affect the performance of crude oil and gas company in the stock market. The binary logistic regression method helps the investor to form an opinion about the shares to be invested. It may be observed that nine financial ratios can classify companies up to a 75.3% level of accuracy into two categories (“GOOD” or “NOT GOOD”), based on their rate of return. When evaluated from the investors’ point of view, we conclude that it is possible to predict out-performing shares by examining these ratios. Various methods are available for data processing for analysis, but in this study, we conclude that ratio methods have the capability to reveal maximum information content, if variables are chosen very carefully with regard to the purpose at hand. Ratios enjoy remarkable simplicity and in spite of the problem of multicollinearity the information revealed by them is so direct to a particular decision-control situation that movements of ratio give a picturesque representation of the movement of an actual business process.

11. References

- Hamilton J.D. (1983), Oil and the Macroeconomy since World War II, *Journal of Political Economy*, 91(2), 228-248.
- Aydogan B. & Berk İ. (2015), Crude oil price shocks and stock returns: Evidence from Turkish stock Market under global liquidity conditions, *International Journal of Energy Economics and Policy*, 5(1), 54-68.
- Cüppers L. & Smeets D. (2015), How do oil price changes affect German stock returns? *International Journal of Energy Economics and Policy*, 5(1), 321-334.
- Bass A. (2017), Does oil prices uncertainty affect Stock Returns in Russia: A bivariate GARCH-in-mean approach, *International Journal of Energy Economics and Policy*, 7(4), 224-230.
- Ulusoy V. & Özdurak C. (2018), The impact of oil price volatility to oil and gas company stock returns and emerging economies, *International Journal of Energy Economics and Policy*, 8(1), 144-158.
- Ojikutu O.T. et al. (2017), Crude oil price volatility and its impact on Nigerian stock market performance from 1985 to 2014, *International Journal of Energy Economics and Policy*, 7(5), 302-311.
- Hooker M. A. (1996), What happened to the Oil Price - Macroeconomy relationship, *Journal of Monetary Economics*, 38, 195-213.
- Burbridge J. & Harrison A. (1984), Testing for the effects of oil price rises using vector auto-regressions, *International Economic Review*, 25, 459-484.

- Gisser M. & Goodwin T. H. (1986), Crude oil and the macro economy: Tests of some popular notions, *Journal of Money, Credit and Banking*, 18(1), 95-103.
- Cobo-Reyes R. & Quiros G. P. (2005), The effect of oil price on industrial production and on stock returns, *Working Paper*, Departamento de Teoria e Historia Economica, Universidad de Granada.
- Basher S. A. & Sadorsky P. (2006), Oil price risk and emerging stock markets, *Global Finance Journal*, 17, 224-251.
- Jones C.M. & Kaul G. (1996), Oil and the stock markets, *The Journal of Finance*, 51(2), 463-491.
- Huang R.D. et al. (1996), Energy shocks and financial markets, *Journal of Futures Markets*, 16(1), 1-27.
- Chen N., (1986), Economic forces and the stock market, *The Journal of Business*, 59(3), 383-403.
- Apergis N. & Miller S. M. (2009), Do structural oil market shocks affect stock prices? *Energy Economics*, 31(4), 569-575.
- Sadorsky P. (1999), Oil price shocks and stock market activity, *Energy Economics*, 21(5), 449-469.
- Papapetrou E. (2001), Oil price shocks, stock market, economic activity and employment in Greece, *Energy Economics*, 23, 511-32.
- Basher S.A. et al. (2012), Oil prices, exchange rates and emerging stock markets. *Energy Economics*, 34(1), 227-240.
- Farzanegan M.R. & Markwardt G. (2009), The effects of oil price shocks on the Iranian economy, *Energy Economics*, 31(1), 134-151.
- El-Sharif I et al. (2005), Evidence on the nature and extent of the relationship between oil prices and equity values in the UK, *Energy Economics*, 27(6), 819-830.
- Ciner C. (2001), Energy shocks and financial markets: Nonlinear linkages, *Studies in Nonlinear Dynamics and Econometrics*, 5(3), 203-212.
- Park J. & Ratti R.A. (2008), Oil price shocks and stock markets in the US and 13 European countries, *Energy Economics*, 30(5), 2587-2608.
- Miller J. I. & Ratti R. A. (2009), Crude oil and stock markets: Stability, instability and bubbles, *Energy Economics*, 31(4), 559-568.
- Bildirici, Melike & Ersin Ö. Ö. (2009), Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange, *Expert Systems with Applications* 36(4), 7355-7362.
- Arouri M.E.H. et al. (2011), Return and volatility transmission between world oil prices and stock markets of the GCC countries, *Economic Modelling*, 28(4), 1815-1825.
- Bharn R. & Nikolova B. (2010), Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries, *MPRA Paper No. 96299*, Posted 05, Nov 2019 16:54 UTC, Department of Economics, University of Portsmouth, Portsmouth Business School, Portsmouth, Portland Street, PO1 3DE, United Kingdom.
- Mork K. A. (1989), Oil and the macroeconomy. When prices go up and down: An extension of Hamilton’s results, *The Journal of Political Economy*, 97(3), 740-744.
- Dhaoui A. & Khraief N. (2014), Empirical Linkage between oil price and stock market returns and volatility: Evidence from international developed markets, *Economics E-Journal, Discussion Paper*, 1-28.

29. Cong R.G et al. (2008), Relationships between oil price shocks and stock market: An empirical analysis from China, *Energy Policy*, 36(9), 3544-3553.
30. Aloui C. et al. (2008), Crude oil volatility shocks and stock market returns, *Journal of Energy Markets*, 1(3), 69-96.
31. Alvarez J. & Solis R. (2010), Crude oil market efficiency and modelling: Insights from the multi scaling autocorrelation pattern, *Energy Economics*, 32(5), 993-1000.
32. Inayat F. (2010), The Relationship between crude oil price and stock performance of European automobile Manufacturer, Master Thesis, School of Economics and Management, Universiteit Van Tilburg.
33. Maghyreh A. (2004), Oil price shocks and emerging stock markets: A generalized VAR approach, *International Journal of Applied Econometrics and Quantitative Studies*, 1(2), 27-40.
34. Jungwook P. & Ronald A. R. (2008) Oil price shocks and stock markets in the U.S. and 13 European Countries, *Energy Economics*, 30, 2587-2608.
35. Ono S. (2011), Oil price shocks and stock markets in BRICs, *The European Journal of Comparative Economics*, 8(1), 29-45.
36. Ready R.C. (2013), Oil prices and Long-run risk, Working Paper.
37. Akomolafe K.J., Jonathan & Danladi D. (2014), Oil price dynamics and the Nigerian stock market: An industry level analysis, *International Journal of Economics, Finance and Management*, 3(6), 308-316.
38. Subarna K. S. & Ali H. M. Z. (2012), Co-Movement of oil, gold, the US dollar and stocks, *Modern Economy*, 3, 111-117.
39. Chen S.T. et al. (2010), Modelling the relationship between the oil price and global food prices, *Applied Energy*, 87(8), 2517-2525.
40. Nandha M. & Faff R. (2008), Does oil move equity prices? A global view, *Energy Economics*, 30, 986-997.
41. Awerbuch et al. (2006), Exploiting the oil- GDP effect to support renewable deployment, *Energy Policy*, 34(17), 2805-2819.
42. Somoye R.O.C & Ilo B.M. (2008), Stock market performance and the Nigerian economic activity, *Lagos Journal of Banking, Finance and Economic Issues*, 2(1), 133-145.
43. Chaudhuri K. & Daniel B.C. (1998), Long-run equilibrium real exchange rates and oil prices, *Economics Letters*, 58, 231-238.
44. Bhunia A. (2013), Cointegration and causal relationship among crude price, domestic gold price and financial variables: An evidence of BSE and NSE, *Journal of Contemporary Issues in Business Research*, 2(1), 1-10.
45. Ojebiyi A. & Wilson D. O. (2011), Exchange rate volatility: An analysis of the relationship between the Nigerian Naira, oil prices and US dollar, Master Thesis in Business Administration 15 ECTS, Master of International Management, Gotland University, Spring 2011.
46. Hidhayathulla A. & Rafee M. (2012), Relationship between Crude oil price and Rupee, Dollar Exchange Rate: An Analysis of Preliminary Evidence. *IOSR Journal of Economics and Finance (IOSR-JEF)*, 3(2), 1-4.
47. Beaver, W.H. (1966), 'Financial ratios as predictors of failure', *Empirical research in Accounting. Selected Studies 1966. Supplement to Journal of Accounting Research 1966*, 71-111.
48. Altman E.I. (1968), Financial Ratios, Discriminant Analysis and the Predication of Corporate Bankruptcy, *Journal of Finance*, September 1978, 589-609.
49. Sinkey J.F. (1975), A multivariate statistical analysis of the characteristics of problem in banks, *Journal of Finance*, March 1975, 21-36.
50. Pettway R.H. & Sinkey J.F. (1980), Establishing on-site bank examination priority: An early warning system using accounting and market information, *Journal of Finance*, March 1980, 137-150.
51. Dince R.R. & Fortson C. (1972), The use of discriminant analysis to predict the capital adequacy of commercial banks, *Journal of Bank Research*, Spring 1972, 54-62.
52. Pinches C.E. & Mingo K.A. (1973), A multivariate analysis of industrial bond ratings, *Journal of Finance*, March 1973, 1-18.
53. Eisenbeis R.A. (1977), Pitfalls in the application of discriminant analysis in business finance, and economics, *Journal of Finance*, June 1977, 875-899.
54. Dombolena I. & Khoury S. (1980), Ratio stability and corporate failure, *Journal of Finance*, September 1980, 1017-1026.
55. Norton C. & Smith R.E. (1979), A comparison of general price level (GPL) and historical cost financial statements in the prediction of bankruptcy, *The Accounting Review*, January 1979, 72-87.
56. Solornon I. & Beck P.J. (1980), A comparison of general price level and historical cost financial statements in the prediction of bankruptcy: A Comment, *Accounting Review*, July 1980, 511-515.
57. Ketz J.E. (1978), The effect of general price level adjustments on the predictive ability of financial ratios, *Supplement to Journal of Accounting Research*, 273-284.
58. Mensah Y.M. (1983), The differential bankruptcy predictive ability of specific price level adjustments: Some empirical evidence, *The Accountancy Review*, April 1983, 228-246.
59. Bazley J.D. (1976), An examination of the ability of alternatives accounting measurement models to predict failure, *Review of Business and Economic Research*, Fall 1976, 32-46.
60. Short D.G. (1980), The impact of price-level adjustment on the meaning of accounting ratios, *Journal of Business, Finance and Accounting*, Autumn 1980, 377-391.
61. Falk H. & Heintz H.A. (1975), Assessing industry risk by ratio analysis, *The Accounting Review*, October 1975, 758-779.
62. Gupta M.C. & Huefner R.J. (1972), A cluster analysis study of financial ratios and industry characteristics, *Journal of Accounting Research*, Spring 1972, 77-95.
63. Thompson T.J. (1976), Sources of systematic risk in common stocks, *Journal of Business*, April 1976, 173-188.
64. Bildersec J.S. (1975), The association between a market-determined measure of risk and other measures of risk, *The Accounting Review*, January 1975, 81-98.
65. Foster F. (1986), *Financial Statement Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 2nd Edition, 1986.
66. O'Connor M.C. (1973), On the usefulness of financial ratios to investors in common stock, *The Accounting Review*, April 1973, 339-352.
67. Roenfeldt R.L. & Cooley P.L. (1978), Predicting corporate profitability for investment selection, *Journal of Business, Finance and Accounting*, Spring 1978, 57-65.

68. Lee S. (2004), Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS, *Environmental Management*, 34(2), 223-232.
69. Pardo et al. (2005), Minimum Θ -divergence estimator in logistic regression models, *Statistical Papers*, 47, 91-108.
70. Ohlson J. (1980), Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109-131.
71. Zmijewski M.E. (1984), Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22, 59-82.
72. Zavgren C. (1985), Assessing the vulnerability to failure of American industrial firms: A logistic analysis, *Journal of Business, Finance and Accounting*, 12(1), 19-45.
73. Abdel-Khalik A. R. (1974), On the efficiency of subject surrogation in accounting research, *Accounting Review*, 49(4), 743-750.
74. McConnell et al. (1986), The president's letter to stockholders: A new look, *Financial Analysis Journal*, 66-70.
75. Aminian F. et al. (2006), Forecasting economic data with Neural Networks, *Computational Economics*, 28, 71-88.
76. Yu L., Wang S. & Lai K. (2009), A neural-network-based nonlinear meta modelling approach to financial time series forecasting, *Applied Soft Computing*, 9, 563-574.
77. Aiken M.W. & Bsat M. (1999), Forecasting market trends with Neural Networks, *IS Management*, 16(4), 1-7.
78. Jaffe J. & Westerfield R. (1985), Patterns in Japanese common stock returns: Day of the week and turn of the year effects, *Journal of Financial and Quantitative Analysis*, 20, 261-272.
79. Kato K. et al. (1990), Day of the week effects in Japanese stocks, Chapter 13, *Calendar Anomalies and Arbitrage*, World Scientific Publishing Co., Singapore.
80. Lee S., Ryu J. & Kim L (2007), Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression and artificial neural network models: Case study of Youngin, Korea, *Landslides*, 4, 327-338.
81. Giancristofaro R.A. & Salmaso L. (2003), Model performance analysis and model validation in logistic regression. *Statistica*, 63(2), 375-396.
82. Harrell F.E., LEE K. L. & MARK D. B. (1996), Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy and measuring and reducing errors, *Statistics in Medicine*, 15, 361-387.
83. Huang Q., Cai Y. & Peng J. (2007), Modelling the spatial pattern of farmland using GIS and multiple logistic regressions: A case study of Maotiao River Basin, Guizhou Province, China *Environmental Modelling and Assessment*, 12(1), 55-61.
84. Hosmer D. W & Lemeshow S. (1989), *Applied Logistic Regression*. New York: Wiley.