

Trend and Regression Analysis of Stock Market Price of Select Industries in India

V. Richard Paul^{1*}, K. Somanathan²

¹Assistant Director, UGC-HRDC, and Assistant Professor, Department of Commerce, Bharathiar University, Coimbatore, Tamil Nadu, India

²Principal, IIKM Business School West Hill, Calicut, Kerala, & Ph.D. Part-time Research Scholar, Department of Commerce, Bharathiar University, Coimbatore, Tamil Nadu, India

***Corresponding author**

V. Richard Paul

Article History

Received: 07.01.2018

Accepted: 13.01.2018

Published: 30.01.2018

DOI:

10.36347/sjebm.2018.v05i01.003



Abstract: The base structure for every large scale industry is finance. The stock market remains a source of investment to raise financial status through an issue of share. Investors receive return from their investment. Stock prices are often very tempting in an economic downturn. Share price is the index of company's financial performance. The present study analyses the trend and regression of stock market price of select industries in India. There are number of reviews previously available related to stock market price analysis of companies. But there is limited study available on industry-wise stock market price analysis. Hence the paper has attempted the trend and regression analysis of stock price of select industries in Bombay Stock Exchange in India. The study is based on only secondary data and data collected for a period of 10 years from 2007-2016. Among the eighty six different industries the researcher has selected fifteen industries on the basis of convenient sampling method. Four companies in each industry which have market capitalization of Rs.200 crores and above have been selected for the analysis. The entire data is evaluated by using SPSS and Microsoft Excel software. This paper also finds forecast trend in the select Industries share price of growth by analyzing polynomial trend based on the forecasted trend pattern. The cubic trend method is applied in the first part of the paper. Among the 15 industries nine have positive growth and significant F value at 5% level except construction, fertilizer, hotel, steel and textiles industries. The second part of this paper estimates regression analysis of share price of select industries. It is to ascertain how collinearity in general, and the sign of correlations in particular, affect parameter inference, variable omission bias and their diagnostic indices in regression. The test is classified by included and excluded variables based on collinearity correlation relationship. The findings indicate a strong relationship between the dependent and independent variables. The trend and regression analysis have found that the pharmaceutical industries' shares prices have long lasting reciprocal of the financial investment foreseen.

Keywords: Share price, regression analysis, cubic trend, co-integration, linear regression method.

INTRODUCTION

The Indian stock market plays a significant role in investors' world not only in Asia but also at the global level. The Bombay Stock Exchange (BSE) is one of the oldest exchanges across the world. The present study pertains to trend and regression analysis of select industries in India. This paper analyses the comparison of aggregate share price from industry to industry. The present study is useful to find out the trend of best industries and indicate optimum cloud for future investment market. Generally the share price is interdependent as well as dependent on a wide multitude of external stimuli like announcement of

government policies, change in interest rates, changes in political scenario, announcement of quarterly results by the listed companies and many others factors. The variables are normally historical data of select industries which included 60 companies of the annual market prices. The forecasting assumes the future trend of industries. The paper has attempted trend analysis, regression analysis and multi-collinearity analysis. Trend analysis is a legitimate destiny of any research analysis' statistical tools. Regression analysis tries a bit of modeling that gets a bit deeper into predicting something by looking at other variables, and what that other ought to be. In statistics, multi-collinearity (also

collinearity) is a phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy. One of the unique features of this paper is to find out the impact of both internal and external factors on share price. Linear multiple regression analysis is used to determine whether the selected independent variables have influence on share prices or not.

LITERATURE REVIEW

Ebru Yuksela and Ozlem Turker Bayrak [1] furnished the relation between the cyclical behaviours of stock market indices of industry, service, finance and technology sectors at Istanbul Stock Exchange and gross domestic product of Turkey between 1998 January and 2011 September. Zeeshan Arshad *et al.* [2] identified the determinants of share prices for the listed commercial banks in Karachi stock exchange over the period 2007-2013. Mohammad Abdel Karim Almumani [3], in his paper traced out the quantitative factors that influence share prices for the listed banks in Amman Stock Exchange over the period 2005-2011 using empirical analysis of a set of independent and dependant variables. V. Richard Paul *et al.* [4] in their article "Role of Construction Sector in Economic Growth in India" pointed out that on the basis of fitted trend, future value has been predicted. The economic growth was analysed the direction of causality between GDP and construction in public-private corporate and household sectors. Adeboye N. O. *et al.* [5] found the effect of Multi-collinearity on the standard error for regression coefficients when it is present in a Classical Linear Regression model (CLRM). A classical linear regression model was fitted into the GDP of Nigeria, and the model was examined for the presence of Multi-collinearity using various techniques such as Farrar-Glauber test, Tolerance level, Variance inflation factor, Eigen values etc. and the result obtained shows that Multi-collinearity has contributed to the increase of the standard error for regression coefficients, thereby rendering the estimated parameters less efficient and less significant in the class of Ordinary Least Squares estimators. Thus, in this study, the ratio analysis, correlation and a linear multiple regression models have been used to measure the individual as well as combined effects of explanatory variables on the dependant variables.

RESEARCH METHODOLOGY

The study analyses the forecasted trend in the select Industries across India, with the company's share prices growth by polynomial trend equation to study the forecasted trend pattern. The cubic trend equation of the form is:

$$Y = b_0 + b^1 t + b^2 t^2 + b^3 t^3,$$

Where,

b_i 's ($i = 1, 2 \& 3$)

are trend coefficients and

b_0 = constant,

is fitted,

$t_i = i^{th}$ year ($i = 1, 2, \dots, 10$) and the results are presented below.

The F-values indicate the overall significance of the trend equation fitted. The coefficients of determination or R^2 are points out a measure to an extent that the trend coefficients are able to explain the variations of the dependent variables under study.

The general form of equation of multiple linear regression is:

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_k \cdot X_{ik} + \varepsilon$$
 and $i = 1, 2, \dots, n$ are the observations from the sample;

Y_i = observation i of the dependent variable;

X_1, X_2, \dots, X_k = independent variables;

B_0 = constant (free term of equation);

B_1, \dots, β_k = coefficients of independent variables;

ε = error term of equation.

Tolerance Level

In multiple regressions, tolerance is used as an indicator of multi-collinearity. Tolerance is estimated by $1 - R^2$, where R^2 is calculated by regressing the independent variable of interest unto the remaining independent variables included in the multiple regression analyses. Researchers desire higher levels of tolerance, as low levels are known to affect adversely the result associated with a multiple regression analyses. The tolerance level is the $1 - R^2$ value when each of the independent variables is regressed on the other independent variables. Low tolerance levels indicate high levels of multi-collinearity. When a tolerance levels get somewhere below 0.40, then multi-collinearity exist.

Variance Inflation Factor (VIF)

In multiple regressions, the VIF is used as an indicator of multi-collinearity. Computationally, it is defined as the reciprocal of tolerance: $1 / (1 - R^2)$. Researchers desire lower levels of VIF, as higher levels of VIF are known to affect adversely the result associated with a multiple regression analyses. In fact, the utility of VIF, as distinct from tolerance, is that VIF specifically indicates the magnitude of the inflation in the standard errors associated with a particular beta weight that is due to multicollinearity. VIF of over 2.50 is starting to indicate relatively high levels of multicollinearity.

The research method applied here is the enter method of linear regression into SPSS. The variables are select industries and time variables. The analysis is classified under nine included variables and six excluded variables based on the correlation and regression models.

Trend in Annual Share Price of Select Industries in India

The following table shows the trend in annual share price of select industries. Each industry consists of four companies.

Table-1: Trend in Annual Share Price of Select Industries 2007-2016

Sl. No.	Industries	Trend Coefficient				R ²	D.F	F
		b ₀	b ₁	b ₂	b ₃			
1.	Banking	329.69	17.37	40.24	-3.82	0.60	3	3.03*
2.	Cement	-423.56	1427.41	-259.54	22.11	0.85	3	11.43*
3.	Chemical	463.55	148.16	-10.38	.652	0.72	3	5.23*
4.	Software	1735.53	-630.76	236.60	-14.24	0.90	3	19.17*
5.	computer hardware	1035.54	-377.96	56.96	-2.27	0.63	3	3.49*
6.	Construction	3.74.71	-1139.49	205.50	-10.26	0.42	3	1.47
7.	Engineering	698.84	-106.39	19.20	0.104	0.84	3	10.96*
8.	Electric engineering	1586.52	-603.54	100.86	-3.87	0.70	3	4.67*
9.	Entertainment	1713.42	-887.99	162.76	-7.14	0.96	3	52.18*
10.	Fertilizer	221.39	108.22	-14.51	0.716	0.29	3	0.833
11.	Health care	381.11	-118.56	16.22	1.06	0.93	3	28.13*
12.	Hotel	572.18	-140.06	19.68	-.586	0.45	3	1.66
13.	Pharmaceutical	3377.35	-1205.94	461.62	-24.25	0.92	3	23.65*
14.	Steel Industry	1847.35	-145.40	16.73	-1.495	0.36	3	1.14
15.	Textile	8797.56	-1935.37	234.03	-6.33	0.28	3	0.810

Source: Calculated by the researcher

* Significance at 5 % level.

Y _t =	329.69	+17.37	+40.24	-3.82	for Banking
Y _t =	-423.56	+1427.41	-259.54	+22.11	for Cement
Y _t =	463.55	+148.16	-10.38	+0.652	for Chemical
Y _t =	1735.53	-630.76	+236.60	-14.24	for Software
Y _t =	1035.54	-377.96	+56.96	-2.27	for computer hardware
Y _t =	3.74.71	-1139.49	+205.50	-10.26	for Construction
Y _t =	698.84	-106.39	+19.20	0.104	for Engineering
Y _t =	1586.52	-603.54	+100.86	-3.87	for Electric engineering
Y _t =	1713.42	-887.99	+162.76	-7.14	for Entertainment
Y _t =	221.39	+108.22	-14.51	+0.716	for Fertilizer
Y _t =	381.11	-118.56	+16.22	+1.06	for Health care
Y _t =	572.18	-140.06	+19.68	-0.586	for Hotel
Y _t =	3377.35	-1205.94	+461.62	-24.25	for Pharmaceutical
Y _t =	1847.35	-145.40	+16.73	-1.495	for Steel Industry
Y _t =	8797.56	-1935.37	+234.03	-6.33	for textile

The result of liner model regression and the estimated cubic trend values of select industries are shown in table 1. The results reveal that entertainment industry's R² is at 5 percent level and the F value is 52.18 which are high among the other industries and likewise the textile industry's F value is 0.81. Among the fifteen industries the results show that the positive growth and significant for F value at 5% level except construction, fertilizer, hotel, steel and textile industries resulted as not significant.

Correlation Coefficient and the Significance of the Dependent and Independent Variables

The Pearson's correlation coefficients are between -1 and 1, the positive values indicate a direct correlation, while a negative values indicate an inverse correlation. The correlation coefficient (Pearson) indicates a stronger correlation as its value is approaching the 1 value. Furthermore, the significance has to be lower than 0.05 to express a good accuracy. Analyzing the results from the Table 2, for all 10 observations, the correlation coefficients are found maximum positive values. Thus there are positive correlations between the dependent and independent variables and some variables are found having negative correlation.

Table-2: The Correlation Coefficient and the Significance for the Dependent and Independent Variables

		T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pearson Correlation	T	1.00	.52	.87	.85	.93	.13	.21	.83	.59	.76	.51	.86	.33	.94	-.59	-.046
	1	.52	1.00	.44	.40	.58	-.23	-.02	.24	-.01	.19	.27	.16	-.12	.47	-.10	-.207
	2	.87	.44	1.00	.75	.82	.38	.35	.84	.70	.82	.37	.91	.51	.90	-.48	.210
	3	.85	.40	.75	1.00	.89	.38	.55	.91	.74	.70	.84	.80	.65	.86	-.18	.349
	4	.93	.58	.82	.89	1.00	.29	.41	.86	.71	.80	.57	.83	.49	.96	-.42	.123
	5	.13	-.23	.38	.38	.29	1.00	.90	.61	.79	.68	.30	.51	.92	.32	.26	.907
	6	.21	-.02	.35	.55	.41	.90	1.00	.66	.80	.65	.52	.48	.92	.38	.36	.892
	7	.83	.24	.84	.91	.86	.61	.66	1.00	.87	.91	.67	.92	.77	.87	-.25	.474
	8	.59	-.01	.70	.74	.71	.79	.80	.87	1.00	.87	.48	.87	.90	.76	-.21	.607
	9	.76	.19	.82	.70	.80	.68	.65	.91	.87	1.00	.34	.91	.71	.83	-.37	.416
	10	.51	.27	.37	.84	.57	.30	.52	.67	.48	.34	1.00	.44	.55	.50	.20	.444
	11	.86	.16	.91	.80	.83	.51	.48	.92	.87	.91	.44	1.00	.65	.92	-.55	.274
	12	.33	-.12	.51	.65	.49	.92	.92	.77	.90	.71	.55	.65	1.00	.51	.19	.878
	13	.94	.47	.90	.86	.96	.32	.38	.87	.76	.83	.50	.92	.51	1.00	-.55	.113
	14	-.59	-.10	-.48	-.18	-.42	.26	.36	-.25	-.21	-.37	.20	-.55	.19	-.55	1.00	.596
15	-.04	-.20	.21	.34	.12	.90	.89	.47	.60	.41	.44	.27	.87	.11	.59	1.000	
Sig. (1-tailed)	T	.	.06	.00	.00	.00	.35	.27	.00	.03	.00	.06	.00	.17	.00	.03	.450
	1	.06	.	.09	.12	.03	.25	.47	.25	.48	.29	.22	.32	.36	.08	.38	.284
	2	.00	.09	.	.00	.00	.13	.15	.00	.01	.00	.14	.00	.06	.00	.07	.280
	3	.00	.12	.00	.	.00	.13	.04	.00	.00	.01	.00	.00	.02	.00	.30	.162
	4	.00	.03	.00	.00	.	.20	.11	.00	.01	.00	.04	.00	.07	.00	.11	.368
	5	.35	.25	.13	.13	.20	.	.00	.02	.00	.01	.19	.06	.00	.17	.23	.000
	6	.27	.47	.15	.04	.11	.00	.	.01	.00	.02	.06	.07	.00	.13	.14	.000
	7	.00	.25	.00	.00	.00	.02	.01	.	.00	.00	.01	.00	.00	.00	.23	.083
	8	.03	.48	.01	.00	.01	.00	.00	.00	.	.00	.07	.00	.00	.00	.28	.031
	9	.00	.29	.00	.01	.00	.01	.02	.00	.00	.	.16	.00	.01	.00	.14	.116
	10	.06	.22	.14	.00	.04	.19	.06	.01	.07	.16	.	.09	.04	.07	.28	.099
	11	.00	.32	.00	.00	.00	.06	.07	.00	.00	.00	.09		.02	.00	.04	.222
	12	.17	.36	.06	.02	.07	.00	.00	.00	.00	.01	.04	.02	.	.064	.29	.000
	13	.00	.08	.00	.00	.00	.17	.13	.00	.00	.00	.07	.00	.06	.	.04	.378
	14	.03	.38	.07	.30	.11	.23	.14	.23	.28	.14	.28	.04	.29	.047	.	.035
15	.45	.28	.28	.16	.36	.00	.00	.08	.03	.11	.09	.22	.00	.378	.03	.	
N	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Source: Calculated by researcher

T: Time.Nos .1-15¹

¹ The names of the Industries: 1. Banking, 2. Cement, 3.Chemical, 4.Software, 5.Computer Hardware, 6.Construction, 7. Engineering, 8. Electric Engineering, 9. Entertainment, 10. Fertilizer, 11.Health Care, 12.Hotel, 13. Pharmaceutical, 14. Steel, 15. Textile

The Linear Regression: A Relationship between the Economic Growth Rate and the Others Indicators

The linear regression is based on the calculation of the correlation coefficient for the all the variable group. The correlation between a dependent variable and the other independent variables are being analysed. If the correlation coefficient has a value approaching 1, this means that the correlation is strong. The aim of using the linear regression is to determine what effect the time value growth has on the independent variables to different industries. The optimal method used for the linear regression model is the enter method, which is based on the elimination, at every step of iteration, of the independent variable which has the weakest influence on the dependent variable. Six industries were (variables) removed, as it

is shown in Table 3. From the Table 4, it can be observed that among the variables, there is a good correlation, but not very strong, because the correlation coefficient is 0.95. In addition to this value, six industries are (independent variables) removed, and so nine industries (variables) are found significant for regression analysis. The significance is below 0.05, which means that there are small errors determined by chance. As a remark, the total credit influence on the time value growth is very good and strong (sig =0.000), and the tolerance is 0.257, greater than 1-Adjusted R square (1-0.2=0.98), which eliminates the uncollinearity risk. VIF (Variance Inflation Factor = 1/Tolerance) also helps for the collinearity analysis, being able to warn about an uncollinearity situation if its value has a greater value than 6.

Table-3: Variables Entered / Removed^a

Model	Variables Entered	Variables Removed	Method
1	1. Textile 2. Pharmaceutical 3. Banking 4. Fertilizer 5. Cement 6. Entertainment 7. Electric 8. Steel 9. Construction ^b	1. Chemical 2. Software 3. Hardware 4. Engineering 5. Healthcare 6. Hotel	Enter

a. Dependent Variable: Time; b. Tolerance = .000 limits reached.

Table-4: The Correlation Coefficient Model Summary b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics		Durbin-Watson
					R Square Change	F Change	
1.	0.95 ^a	0.98	.02	.	1.000	.	2.486

a. Predictors: (Constant), Textile, Pharmaceutical, Banking, Fertilizer, Cement, Entertainment, Electric, Steel, Construction; b. Dependent Variable: Time

Table-5: The Linear Regression Coefficient Model Parameters

Sl.No	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
	(Constant)	2.390	.000		.	.		
1.	Banking	-.003	.000	-.365	.	.	.082	12.122
2.	Cement	.001	.000	.539	.	.	.052	19.353
3.	Construction	.001	.000	.218	.	.	.019	53.336
4.	Electric	-.003	.000	-.440	.	.	.017	57.794
5.	Entertainment	.002	.000	.470	.	.	.063	15.791
6.	Fertilizer	.005	.000	.232	.	.	.254	3.938
7.	Pharmaceutical	.001	.000	.781	.	.	.022	44.837
8.	Steel	.003	.000	.544	.	.	.024	41.014
9.	Textile	-.001	.000	-.872	.	.	.023	43.847

a. Dependent Variable: VAR00016

In this case, VIF is 43.84, which also eliminates the uncollinearity risk. Thus, using the coefficients calculation (column B – Table 5), the linear regression equation obtained are for construction and electric (-0.003), for Cement, construction,

pharmaceutical and textile (0.001), for entertainment (0.002), for fertilizer (0.005), for steel (0.003) and time 2.390. The interpretation of coefficients from regression equation points out that, considering data for the period 2006–2017, on a short period of time, the following

correlations are expressed: if construction and electric industries increase with one point, then time growth rate decreases with -0.003 percent; if Cement, construction, pharmaceutical and textile industries increase with one percent, then time growth rate decreases with 0.001 percent; if entertainment industry increases with one

percent, then time growth rate decreases with 0.002 percent; if fertilizer industry increases with one percent, then time growth rate decreases with 0.005 percent; if steel industry increases with one point, then time growth rate increases with -0.001 percent.

Table-6: The Linear Regression Coefficient Excluded Variables ^a

Sl. No	Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1.	Chemical	. ^b000	.	.000
2.	Software	. ^b000	.	.000
3.	Hardware	. ^b000	.	.000
4.	Engineering	. ^b000	.	.000
5.	Healthcare	. ^b000	.	.000
6.	Hotel	. ^b000	.	.000

a. Dependent Variable: Time; b. Predictors in the Model: (Constant), Textile, Pharmaceutical, Banking, Fertilizer, Cement, Entertainment, Electric, Steel, Construction.

In this case, VIF is 0.0, which also eliminates the collinearity risk. Thus, using the coefficients calculated (column B – Table 6), the linear regression equation obtained zero. The interpretation of coefficients from regression equation points out that,

considering data for the period 2006–2017, the period of time, it is expressed the following correlations the six variable namely chemical, software, hardware, engineering, healthcare and hotel industries exclude and non-collinearity.

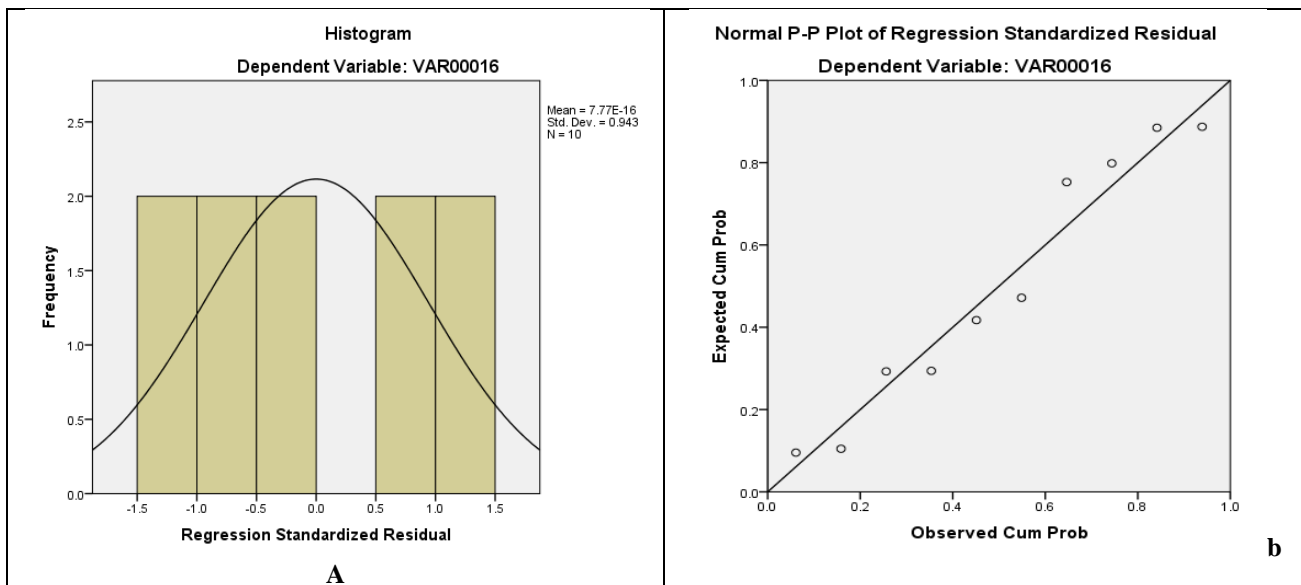


Fig -1

Figure 1 presents the residuals by comparing them with the normal distribution law. A residual is an observable estimate of the unobservable statistical error. The residuals generally comply with the normal distribution law (an empirical analysis based on the comparison of the curve – Figure 1(a) and of the points

near the line – Figure 1(b), thus the linear regression model can be applied for the data analyzed. As an observation, for the 0.6-0.9 interval the residuals do not comply with the normal distribution law, and so on this interval the errors can be higher.

Table-7: Collinearity Diagnostics^a

Model	Dimension	Eigen value	Condition Index	Variance Proportions										
				Constant	Banking	Cement	Construction	Electric	Entertainment	Fertilizer	Pharmaceutical	Steel	Textile	
1	1	8.984	1.000	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	2	.641	3.743	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00
	3	.217	6.428	.01	.02	.00	.00	.00	.00	.01	.00	.00	.00	.00
	4	.058	12.489	.04	.01	.09	.00	.00	.00	.00	.08	.01	.01	.00
	5	.046	14.005	.12	.01	.06	.00	.00	.00	.08	.12	.00	.00	.00
	6	.032	16.848	.46	.04	.08	.00	.00	.00	.03	.01	.01	.00	.00
	7	.016	23.853	.01	.03	.00	.00	.00	.05	.39	.27	.02	.00	.00
	8	.004	49.450	.01	.14	.17	.06	.06	.06	.02	.39	.71	.12	.03
	9	.001	77.768	.12	.20	.26	.01	.22	.08	.08	.05	.00	.86	.82
	10	.001	92.625	.24	.56	.33	.92	.67	.37	.37	.08	.26	.00	.14

a. Dependent Variable: Time

The most important information transmitted by table 8 is represented by the values of the condition indexes. Theoretically, an index higher than 15 shows that there is a collinearity problem, while a value higher than 30 indicates serious collinearity problems. The study comes across values of the condition index, above 15, for models representing the linear combination of independent variables which explain best the evolution of economic efficiency.

CONCLUSION

The present study derives that the cubic trend found among the 15 industries has positive growth and significant for F value at 5% level except construction, fertilizer, hotel, steel and textiles industries as not significant. The multi-collinearity analysis has analysed dependent and independent Pearson correlation analysis and it is tested at the level of significant. In this paper the effects of collinearity on omitted variables, bias and parameter variance, estimates are examined. It is found that, consistent with prior results, negative and positive correlations can yield less precise estimates, and can induce parameters to switch signs. The trend and regression analysis resulted found that the pharmaceutical industries share prices having long lasting reciprocate of the financial investment foreseen.

REFERENCES

1. Yuksel E, Bayrak OT. Cyclical behavior of stock exchange index by sectors: a case from Turkey. *Procedia-Social and Behavioral Sciences*. 2012 Oct 24;62:947-51.
2. Arshad Z, Arshaad AR, Yousaf S, Jamil S. Determinants of share prices of listed commercial banks in Pakistan. *Journal of Economics and Finance*, Mar-Apr. 2015:56-64.
3. Almumani MA. Determinants of equity share prices of the listed banks in Amman stock exchange: Quantitative approach. *International*

Journal of Business and Social Science. 2014 Jan 1;5(1).

4. Paul, Richard V. and Savitha P. "Role of Construction Sector in Economic Growth in India." *Web*. 28 Oct. 2017. <http://www.shanlaxjournals.in/pdf/COM/V5N2/COM_V5_N2_008.pdf>.
5. Adeboye N, Fagoyinbo I, Olatayo T. Estimation of the effect of multicollinearity on the standard error for regression coefficients. *IOSR Journal of Mathematics*. 2014;10(4):16-20.