EduTech
e-Journal of Education & Technology
ISSN 0975-5004

Ref # : 1102-00003
Issue Number: 1102
Published On: 01-August-2011

ARIMA Modeling on Google Search Results on Terminologies of Non-Traditional Modus Operandi of Education

Abstract

An array of inexhaustible range of labels, such as external studies, flexi education, correspondence education, are used to address the non-traditional modus operandi of education. The paper is based on first of its kind statistical analysis on various terms emanating from the realm of non-traditional modus operandi of education. The study uses `Search Results' of Google as an index of the strength/popularity of the terms. The terms `e-learning' and `distance learning' have yielded highest number of `search results' in the data collected fortnightly over the last one year. ARIMA modeling has been attempted for getting at the inferential dimensions of terminologies. Notwithstanding its constraints, the study also attempts to sensitize educators and its stakeholders to track and gauge trends on the shades of terms being used in the educational venture.

Keywords: e-learning, distanced education, flexi-education, part-time education, external studies, correspondence education, distance education lexicon.

1.0 Introduction

The non-traditional modus operandi of education backed by some systematic and scholarly studies is changing the landscape of education across the word. The non-traditional educational methods aimed at providing continuing education encompass various forms of education that do not follow from standard classroom based methods. While 'Eklavyaism' provides an unparallel instance of self-learning with a meditative mind without physical presence of a `Guru', Caleb Phillips and Issac Pitman are virtually credits for initiating the era of non-traditional modus operandi of education through correspondence education which got metamorphosed into what is arguably the most pervasive, most talked-about, and most globally recognized form of instruction at a distance (Jegede & Naidu, 1999). Of late efforts are fast being made for a formal world-wide recognition of non-traditional modus operandi of education as a discipline. Notably, unlike most of the other disciplines, it enjoys a unique character of being addressed by an inexhaustible range of labels, such as correspondence education, external studies, off-campus, non-traditional, independent learning, home studies, individualized learning, open learning, to name a few (Anand, 1979; Moore, 2001). Further the technological advancements, synchronized with emerging consciousness amongst the masses for information, knowledge and education, have led to the emergence of newer concepts and terms, such as `e-learning', `e-education' in different connotations. The enigma in these terms has indeed revolutionized the entire process of public instructions.

The present paper is probably a maiden attempt with a statistical analysis of the various terms which gained coinage in the realm of the non-traditional modus operandi of education.

2.0 Methodology

Based on the premise that in the eWorld `Search Results' could be argued as an index of importance or popularity of `terms', the scope of the paper has been confined to Google's search. The web engine `Google Search' is one of the most-used search engines on the web. "The main purpose of Google Search is to hunt for text in web-pages, as opposed to other data, such as with Google Image Search. The exact percentage of the total of web-pages that Google indexes are not known, as it is very hard to actually calculate. Google not only indexes and caches web pages but also takes "snapshots" of other file types, which include PDF, Word documents, Excel spreadsheets, Flash SWF, plain text files, and so on" (http://en.wikipedia.org/wiki/Google_Search)

While the process of data collection is still going on for a futuristic project, for the paper, the data collected at regular intervals of two weeks from April 15, 2010 till June 24, 2011 has been used. In all 27 observations collected over the period for about 60 terms were put to analysis.

For the exploratory study, two fold analyses were undertaken. At the first instance simple descriptive statistical tools, such as Arithmetic Mean (AM), Standard Deviation (SD), Coefficient of Variation (CV), Maximum Value (Max) and the Minimum Vale (Min), were used to describe the scenario.
Further for inferential contemplation SPSS based ARIMA (Auto-Regressive Integrated Moving Average) Modeling was attempted on the terms yielding maximum search results in various categories.

The Time Series Modeler procedure estimates exponential smoothing, univariate Autoregressive Integrated Moving Average (ARIMA), and multivariate ARIMA (or transfer function models) models for time series, and produces forecasts. The procedure includes an Expert Modeler that automatically identifies and estimates the best-fitting ARIMA or exponential smoothing model for one or more dependent variable series, thus eliminating the need to identify an appropriate model through trial and error.

ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging (Box, 1976, Pierse)
Precisely, $\{X_t\}$ process defined by:
$X_t=\alpha_1 X_{t-1}+ \alpha_2 X_{t-2}+\ldots+ \alpha_p X_{t-p+} Z_t$
(Where Z_t purely Random process with mean Zero and Variance $\sigma^{2.}$)
is called auto regressive process of order p. It is a regression of X_t on its own past value.
The process $\{X_t\}$ defined by :
$X_t = Z_t+\theta_1 Z_{t-1}+\ldots.+ \theta_q Z_{t-q}$
(Where Z_t purely Random process with mean Zero and Variance $\sigma^{2.}$) is called a moving average process of order q .
The combination of AR and MA processes is called ARMA(p,q) process .
If the time series is non stationary then to apply ARMA model, we need to difference(d) it to make it stationary .
If we difference it d times , then we will call this time series to be integrated of order d, and the process is called ARIMA(p, d ,q), where

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences, and
- q is the number of lagged forecast errors in the prediction equation.

For the study SPSS was used to get at the models.

3.0 Analysis and Results

The analysis has been undertaken on two counts. At the first instance, the data was scaled down to basic descriptive measures, and then inter-intra comparison were undertaken on selected terms with a view to get an insight into trends. While providing enough scope for the keen enthusiasts to read between the tables and graphs for drawing more inference of the topic, some of the key results are presented in the following sections.

3.1 Over-all descriptive measures

The following table (T.1) gives information on Arithmetic Mean (AM), Standard Deviation (SD), Coefficient of Variation (CV), Maximum Value (Max) and the Minimum Vale (Min) about 60 terms (arranged in alphabetic order).

T. I Search results on terms of educational modus operandi.

Terms	AM	Min	Max	SD	CV
"correspondence course"	120718.7	55600	162000	39148.46	32.43
"correspondence courses"	124555.8	55400	181000	42216.04	33.89
"correspondence education"	68779.2	11700	73100	24731.31	35.96
"correspondence studies"	4707.088	1770	10900	2303.97	48.95
"distance and open education"	2020.251	551	2090	559.45	27.69
"distance and open learning"	4092.704	2400	5390	862.95	21.09
"distance education"	1277488	289000	1990000	523569.40	40.98
"distance learning"	2635286	877000	3920000	916372.95	34.77
"e_education"	229.8978	88	247	61.35	26.69
"e_learning"	24437.78	13500	27700	5387.51	22.05
"education through Air"	16.0544	8	19	3.35	20.88

"education through correspondence"	670.5328	412	879	144.65	21.57
"education through distance"	5025.792	2450	12900	2184.14	43.46
"education through ICT"	1812.221	819	2220	523.02	28.86
"education through mobile"	246.7904	105	331	74.12	30.03
"education through Radio"	361.816	254	548	65.23	18.03
"e-education"	87371.56	47100	91200	19679.56	22.52
"eEducation"	9824.444	6200	10200	1815.83	18.48
"eLearning"	18020741	1860000	27200000	9827945.84	54.54
"e-learning"	8652000	3420000	17900000	3569755.23	41.26
"external studies"	10506.1	5470	15700	3060.82	29.13
"flexi education"	998.344	221	1500	439.02	43.97
"flexi learning"	520.2544	285	622	114.75	22.06
"independent education"	66065.92	23200	91900	24720.80	37.42
"independent learning"	89972.16	49600	110000	22862.27	25.41
"independent studies"	137938.1	92000	166000	22879.99	16.59
"instis of corr studies"	5.3776	3	6	1.08	20.14
"mobile education"	17276.24	6390	33000	6515.75	37.72
"mobile learning"	158283.2	73600	234000	42380.77	26.78
"non-contiguous education"	6.2608	4	8	0.96	15.38
"non-traditional education"	4895.008	2810	6500	1115.83	22.80
"non-traditional learning"	4510.368	1500	18400	3340.38	74.06
"non-traditional teaching"	1343.688	850	1660	255.02	18.98
"off-campus course"	1635.382	931	1960	361.89	22.13
"off-campus courses"	4240.256	2120	5340	1142.24	26.94
"off-campus education"	2716.864	1210	3890	790.59	29.10
"off-campus program"	3927.216	2350	5090	866.69	22.07
"off-campus studies"	4099.696	2000	5020	1144.74	27.92
"off-campus"	1529611	502000	2900000	605576.54	39.59
"open and distance education"	13151.01	6080	17300	3436.28	26.13
"open and distance learning"	35557.92	24600	42700	6033.08	16.97
"open education"	162264.8	83600	202000	45192.85	27.85
"open learning"	159472.2	84700	201000	36494.51	22.88
"part-time course"	48291.68	19900	66700	16351.98	33.86
"part-time courses"	197646.2	65600	335000	89384.56	45.22
"part-time education"	42508.16	7930	122000	23709.43	55.78
"part-time learning"	5002.112	1950	6960	1718.80	34.36
"part-time teaching"	32980.8	20500	40800	6467.54	19.61
"private education"	177965.1	87700	241000	43911.48	24.67
"private teaching"	17161.28	10500	21000	3789.14	22.08
e-education	3.57E+08	56000000	389000000	126304994.81	35.39
e-learning	96052444	3100000	100000000	35010648.72	36.45

At the outset the data reveal that the term e-education has resulted in highest arithmetic mean. The term has also yielded highest `minimum search results' and the `maximum search results' in comparison to other terms. However, in the "phrased search" the highest arithmetic mean on search results was obtained for the term 'elearning', while the term with the connotation 'e-learning' yielded highest minimum search results. On the other extreme was the term 'non-contiguous education' which yielded minimum arithmetic mean on search results. The term 'non-contiguous education' also conceded least `minimum search results' as also least `maximum search results' in comparison to other terms.

On the consistency level, the least `coefficient of variation' was obtained for the term `independent studies', which was otherwise ranked in 15 amongst terms with highest arithmetic mean. On this count the term `open and distance learning' got the next rank. On the other extreme were the terms `non-traditional learning', `elearning` and `part-time education' for which the `coefficient of variation' were computed to be amongst the maximum.

The analysis shows that the terms e-education and e-learning have not only got the highest average on search results amongst over 55 terms put under observation, but also registered a high gap with other terms throughout the search period. Interestingly, while with an average on search results of 356893333.33 the term e-education yielded the highest score, in the phrased search the term `e-education ' got ranked at 18 with an average on search results of 87371.55.

Notable is the variations in different connotation of the term elearning such as `elearning' or `e-learning', wherein the term `distance learning' has figured at distant fifth position.
Interestingly the terms prefixed with `e' yielded different results for `eLearning'ing' and `e-_ learning'.

3.2 ARIMA Modeling

In this section ARIMA modeling has been worked on for teams yielding maximum results in various categories.

3.2.1 Distance learning

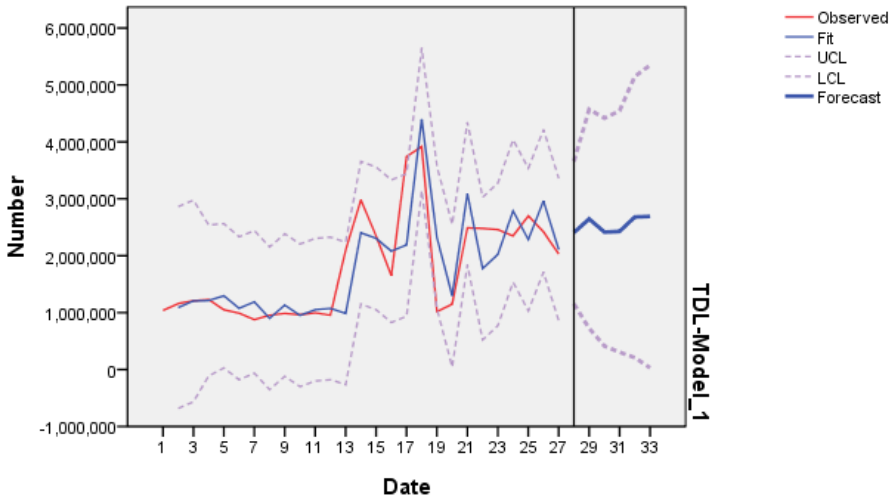
When focused was confined on the "phrased search results" (PSR), the terms other than those prefixed by `e', `distance learning' got highest (on an average 2635286.4) search results, The term `Distance education', matched quite favorably with the term `off campus' in the search results at the initial stage of data collection, but marked variations were observed in favor of the later in the latter half of the study. However, the terms `private education' and `part-time courses' have been competing with each other for a higher slot in the search results.

In the analysis, the terms distance learning got ARIMA(2,1,1) as the Model Type
The following table give forecast values, where forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Forecast

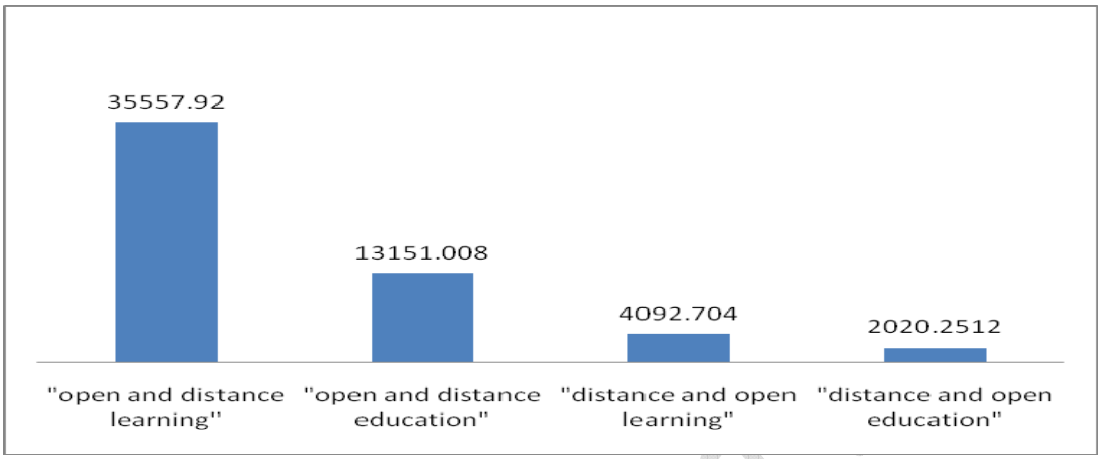
Model		28	29	30	31	32	33
TDL-Model_1	Forecast	2.41E6	2.65E6	2.42E6	2.43E6	2.68E6	2.69E6
	UCL	3.66E6	4.57E6	4.42E6	4.55E6	5.15E6	5.35E6
	LCL	1.15E6	7.24E5	4.15E5	3.07E5	2.09E5	3.13E4

The following graph gives plot on values as `observed' and `forecast', besides on Fit amidst Upper Control Limit (UCL) and Lower Control Limit (LCL).



3.2.2 Open and distance learning

In the search results on terms prefixed with `correspondence' or `open' and 'distance'; `education' and `learning', the term 'correspondence education' in particular and the terms prefixed by `correspondence' in general acquired lower rank, in comparison with the terms prefixed by 'distance' or 'open'. Further, the term 'Open and distance learning' stands apart at the highest pedestal in comparison to the terms 'Open and distance education' or 'distance and open learning' or 'distance and open education. (Following diagram speaks for it)
AM of PSR of terms with Prefixed with `open' and 'distance'; `education' and `learning'



In the analysis, the terms distance learning got ARIMA (0,1.0) as the Model Type
The following table gives ARIMA Model Parameters.

ARIMA Model Parameters

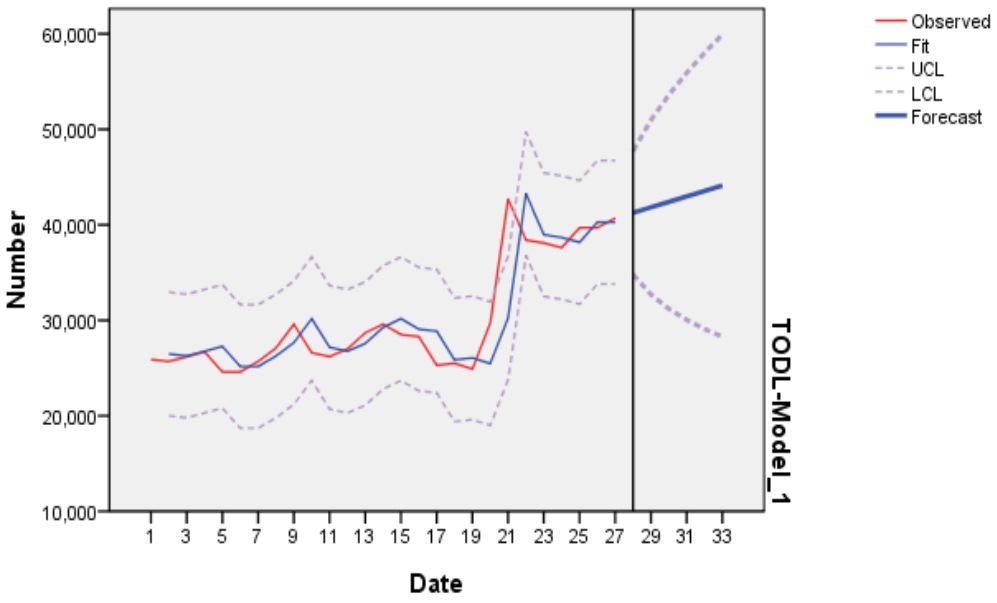
				Estimate	SE	t	Sig.
TODL-Model_1	TODL	No Transformation	Constant	569.231	615.661	.925	.364
			Difference	1			

The following table give forecast values, where forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Forecast

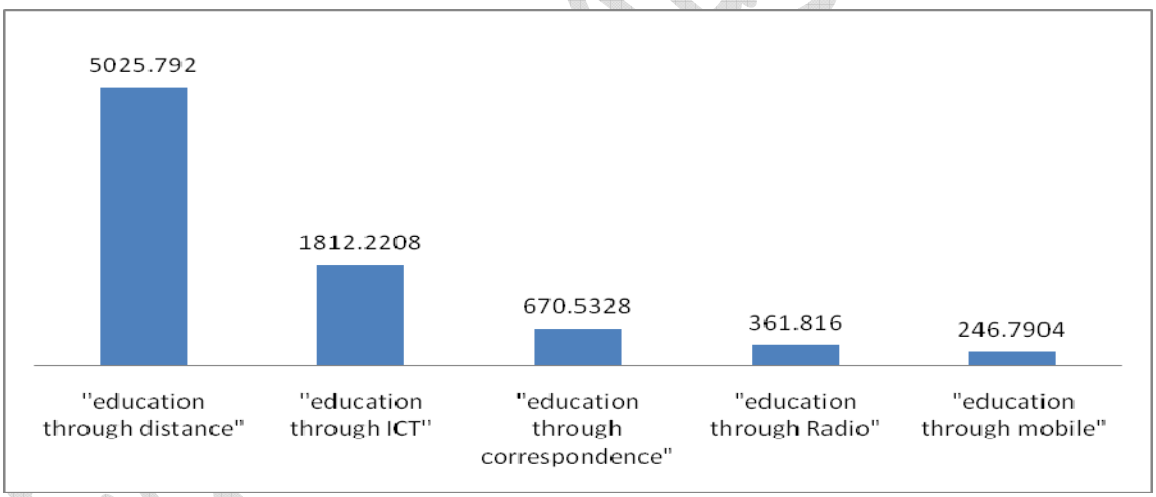
Model		28	29	30	31	32	33
TODL-Model_1	Forecast	4.13E4	4.18E4	4.24E4	4.30E4	4.35E4	4.41E4
	UCL	4.77E4	5.10E4	5.36E4	5.59E4	5.80E4	6.00E4
	LCL	3.48E4	3.27E4	3.12E4	3.00E4	2.91E4	2.83E4

The following diagram gives plot on values as `observed' and `forecast', besides on Fit amidst Upper Control Limit (UCL) and Lower Control Limit (LCL).



3.2.3 Education through distance

In the analysis of search results on some terms prefixed with 'Education through . . ', the term 'education through distance' has yielded the highest average search results and 'education through air' the least results. Following diagram gives the information on the average of search results on "education through distance", "education through ICT", "education through correspondence", "education through Radio" and "education through mobile".



In the analysis, the terms distance learning got ARIMA(0,1,1) as the Model Type

The following table gives ARIMA Model Parameters

ARIMA Model Parameters

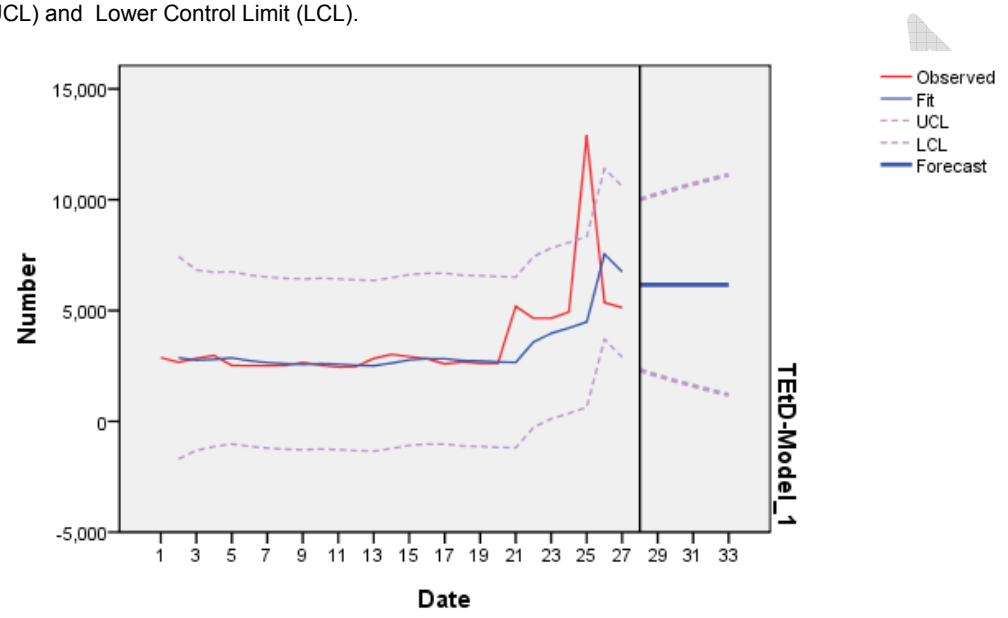
			Estimate	SE	t	Sig.
TEtD-Model_1	TEtD	No Transformation	Difference	1			
		MA	Lag 1	.636	.160	3.984	.001

The following table give forecast values, where forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Forecast

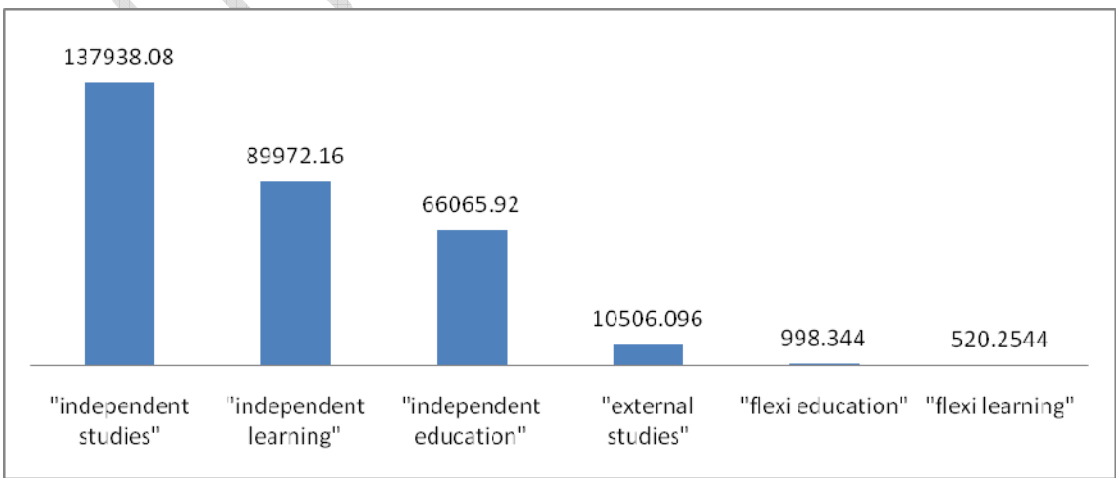
Model		28	29	30	31	32	33
TEtD-Model_1	Forecast	6160.17	6160.17	6160.17	6160.17	6160.17	6160.17
	UCL	1.00E4	1.03E4	1.05E4	1.07E4	1.09E4	1.11E4
	LCL	2307.95	2060.71	1827.55	1606.32	1395.34	1193.32

The following graph gives plot on values as `observed' and `forecast', besides on Fit amidst Upper Control Limit (UCL) and Lower Control Limit (LCL).



3.2.4 Comparison between the traditional terms

The results from Table T.1 suggest that in the array of over 55 terms under observation, the terms prefixed with `e' and 'Distance' have taken precedence over all other teams. Amongst the traditional terms, striking has been the reference to the traditional term 'Off Campus' which got the top slot with 'Distance Learning' and 'distance education' in the search of phrase. Else the terms 'external studies', 'flexi education', 'flexi learning" and 'non-contiguous education' culminated with least number of results on search.



In the analysis, the terms distance learning got ARIMA(0,1,0) as the Model Type
The following table gives ARIMA Model Parameters

ARIMA Model Parameters

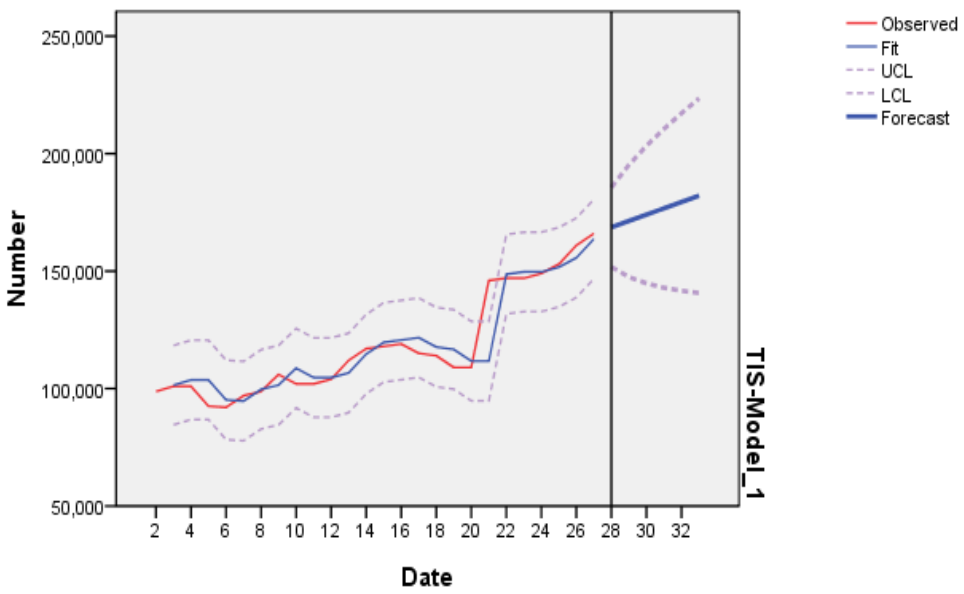
				Estimate	SE	t	Sig.
TIS-Model_1	TIS	No Transformation	Constant	2.688E3	1.637E3	1.642	.114
			Difference	1			

The following table give forecast values, where forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Forecast

Model		28	29	30	31	32	33
TIS-Model_1	Forecast	1.69E5	1.71E5	1.74E5	1.77E5	1.79E5	1.82E5
	UCL	1.86E5	1.95E5	2.03E5	2.11E5	2.17E5	2.24E5
	LCL	1.52E5	1.47E5	1.45E5	1.43E5	1.42E5	1.41E5

The following graph gives plot on values as `observed' and `forecast', besides on Fit amidst Upper Control Limit (UCL) and Lower Control Limit (LCL).



4.0 Conclusion

Notwithstanding constraints on count of limited time-scale data and the hidden intricacies of Google's algorithm, the study on various terms constituting 'inexhaustible range of labels' in the realm of non-traditional modus operandi of education, suggests that a number of tradition terms, such as 'external studies', `flexi-learning', 'flexi-education', 'correspondence education', `external studies' are paving the way for the terms 'elearning', 'distance learning' and 'distance education'. In the search results of eWorld, interestingly the term `Open and Distance' has yielded more results than `Distance and Open' in both cases when prefixed to the terms `learning' or 'education'.

Nonetheless, the study attempts to sensitize and invite the stakeholders in education to track and gauge trends on the shades of the terms emerging in the realm of education. May be that in future such a maiden pilot study will invoke theoreticians to further dwell on the growth of the Open and Distance education system.

5.0 References

Anand, S. (1979). *University Without Walls*, New Delhi: Vikas.

Box, G.E.P, & G.M. Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, (2nd ed.), San Fransisco: Holden-Day.

Brassard, M.L. (1982). Glossary, in learning at a distance: A world Perspective, Daniel et al., (Ed.), Edmonton: ICCE.
Introduction to ARIMA: nonseasonal models, http://www.duke.edu/~rnau/411arim.htm

Moore, M. G. Moore & Shattuck, K. (2001). Glossary of Distance Education Terms. The Pennsylvania State University

Jegede, O & Naidu, S. (1999). http://www.ascilite.org.au/ajet/e-jist/docs/vol3no1/ editorial.htm

Pierse, Richard G. "ARIMA Models",http://www.shs.surrey.ac.uk/economics/rpierse/files/arima.pdf
Young, D. M. (1982). Our Name, ICCE Newsletter on Distance Learning Vol. 12

Dr. Ravi K. Mahajan
University School of Open Learning
Panjab University, Chandigarh, INDIA

mahajan_krm@yahoo.co.in; rkmahajan@pu.ac.in